Terminologie & Ontologie : Théories et Applications

Actes de la conférence

TOTh 2014

Le Bourget du Lac – 12 & 13 juin 2014

Les ouvrages TOTh précédents sont disponibles sur le site du Comptoir des Presses d'Universités (www.lcdpu.fr) ou auprès de : contact@toth.condillac.org

Éditeur : Presses Universitaires Savoie Mont Blanc
27 rue Marcoz
BP 1104
73011 CHAMBÉRY CEDEX
www.univ-smb.fr

Réalisation : C. Brun, C. Roche
Collection « Terminologica »
ISBN : 978-2-37741-050-7
ISSN : 2607-5008
Dépôt légal : novembre 2020

Terminologie & Ontologie : Théories et Applications

Actes de la conférence

TOTh 2014

Bourget du Lac – 12 & 13 juin 2014

http://toth.condillac.org

avec le soutien de :

Université Savoie Mont Blanc

École d'ingénieurs Polytech Annecy Chambéry

Ministère de la Culture - Direction Générale à la Langue Française et aux Langues de France

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

MINISTÈRE
DE LA CULTURE

Presses Universitaires Savoie Mont Blanc
Collection « Terminologica »

Comité scientifique

Président du Comité scientifique : Christophe Roche

Comité de pilotage

Rute Costa	Universidade Nova de Lisboa
Humbley John	Université Paris 7
Kockaert Hendrik	University of Leuven
Christophe Roche	Université Savoie Mont Blanc

Comité de programme 2019

Le comité de programme est constitué chaque année à partir du comité scientifique de TOTh en fonction des soumissions reçues. La composition du comité scientifique est accessible à l'adresse suivante : http://toth.condillac.org/committees

Guadelupe Aguado	Universidad Politécnica de Madrid – Spain
Amparo Alcina	Universitat Jaume I – Spain
Bruno Bachimont	Université Technologie de Compiègne – France
Jean-Paul Barthès	Université Technologie de Compiègne – France
Christopher Brewster	TNO – The Netherlands
Danielle Candel	CNRS, Université Paris Diderot – France
Sylviane Cardey	Université de Franche-Comté – France
Stéphane Chaudiron	Université de Lille 3 – France
Manuel Célio Conceição	Universidade do Algarve – Portugal
Rute Costa	Universidade NOVA de Lisboa – Portugal
Bruno Courbon	Université Laval – Canada
Lyne Da Sylva	Université de Montréal – Canada
Luc Damas	Université Savoie Mont-Blanc – France
Éric De La Clergery	INRIA – France
Dardo De Vecchi	Kedge Business School – France
Valérie Delavigne	Université Paris 3 – France
Sylvie Després	Université Paris 13 – France
Juan Carlos Diaz Vasquez	EAFIT University – Colombia
Hanne Erdman Thomsen	Copenhagen Business School – Denmark
Pamela Faber	Universidad de Granada – Spain
Christiane Fellbaum	Princeton University – USA
Iolanda Galanes	Universidade de Vigo – Spain
Christian Galinski	INFOTERM – Austria
François Gaudin	Université de Rouen – France
Teodora Ghiviriga	Alexandru Ioan Cuza University – Romania
Jean-Yves Gresser	ancien Directeur à la Banque de France – France
Ollivier Haemmerlé	Université de Toulouse – France
Gernot Hebenstreit	University of Graz – Austria
Amanda Hicks	University of Florida – USA
John Humbley	Université Paris 7 – France

Kyo Kageura	University of Tokyo – Japan
Heba Lecocq	INALCO – France
Hélène Ledouble	Université de Toulon – France
Patrick Leroyer	Aarhus University – Denmark
Georg Löckinger	University of Applied Sciences Upper Austria – Austria
António Lucas Soares	University of Porto, INESC – Portugal
Bénédicte Madinier	Dispositif d'enrichissement de la langue française – France
Candida Jaci de Sousa Melo	Universidade Federal do Rio Grande do Norte – Brazil
Jean-Guy Meunier	Université de Montréal – Canada
Christine Michaux	Université de Mons – Belgium
Fidelma Ní Ghallchobhair	Foras na Gaeilge, Irish-Language Body – Ireland
Henrik Nilsson	TNC – Sweden
Silvia Piccini	Italian National Research Council – Italy
Suzanne Pinson	Université Paris Dauphine - France
Marina Platonova	Riga Technical University – Latvia
Thierry Poibeau	CNRS Lattice – France
Maria Pozzi	el colegio de méxico – Mexico
Michele Prandi	Università degli Studi di Genova – Italy
Jean Quirion	Université d'Ottawa – Canada
Renato Reinau	Suva – Switzerland
Christophe Roche	Université Savoie Mont Blanc – France
Mathieu Roche	CIRAD – France
Laurent Romary	INRIA & HUB-ISDL – Germany
Micaela Rossi	Università degli studi di Genova – Italy
Bernadette Sharp	Staffordshire University – Great Britain
Marcus Spies	Universität München - Germany
Anne Theissen	Université de Strasbourg – France
Philippe Thoiron	Université Lyon 2 – France
Marc Van Campenhoudt	Université libre de Bruxelles – Belgium
Kara Warburton	City University of Hong Kong – China
Maria Teresa Zanola	Università Cattolica del Sacro Cuore – Italy
Fabio Massimo Zanzotto	University of Roma – Italy

Avant-propos

La Terminologie est une discipline scientifique à part entière qui puise à de nombreux domaines dont la linguistique, la théorie de la connaissance et la logique. Pour que cette diversité soit une richesse, il faut lui offrir un cadre approprié au sein duquel elle puisse s'exprimer et s'épanouir : c'est une des raisons d'être des *Conférences TOTh*.

Dans ce contexte, la formation et la transmission des connaissances jouent un rôle essentiel. La *Formation TOTh* précédant la Conférence se déroule sur deux années consécutives dédiées pour l'une à la dimension linguistique et pour l'autre à la dimension conceptuelle de la terminologie, deux dimensions étroitement liées.

A la présentation de travaux sélectionnés par un Comité de programme international, la *Conférence TOTh* inclut une *Conférence invitée* et une *Disputatio*. La première, donnée par une personnalité reconnue dans son domaine vise l'ouverture à d'autres approches de la langue et de la connaissance. La seconde, à travers une lecture commentée effectuée par un membre du comité scientifique, renoue avec une forme d'enseignement et de recherche héritée de la scolastique.

Cette année notre collègue Christiane Fellbaum, Professeur à l'Université de Princeton, a ouvert la *Conférence TOTh* par un exposé sur le thème de « *Ontology meets natural language : What can linguists learn ?* ». Pierre Lerat, Professeur à l'Université Paris XIII, nous a parlé, dans le cadre de la *Disputatio*, de « *Harris et les immunologistes : un exemple à suivre ?* ».

La conférence s'est terminée par le compte-rendu de la Journée d'Étude TOTh 2013 « Verbal and non verbal representation in terminology » du 8 novembre 2013 (Copenhague, Danemark) et par la présentation de la prochaine Journée d'Étude TOTh 2014 « Multilingual thesaurus and terminology » qui se déroulera le 5 décembre prochain (Bruxelles, Belgique).

Les 16 communications, auxquelles se sont ajoutés 9 posters, ont permis d'aborder de nombreux sujets tant théoriques que pratiques rappelant qu'il ne peut y avoir de Terminologie sans langue ni savoir de spécialité. 9 articles longs et 2 articles courts ont été retenus pour publication. Je vous invite à les découvrir à travers ces actes.

Avant de vous souhaiter bonne lecture, j'aimerais terminer en remerciant tous les participants pour la richesse des débats et des moments partagés.

Christophe Roche
Président du comité scientifique

Sommaire

Conférence d'ouverture 13

Ontology meets natural language :
Does the lexicon have a regular structure ?
Christiane Fellbaum 15

Disputatio 21

Harris et les immunologistes : un exemple à suivre ?
Pierre Lerat 23

Articles 45

Systèmes conceptuels et isotopies métaphoriques
dans les langues de spécialité : analyse contrastive de
quelques domaines en français et en italien
Micaela Rossi 47

Méthodologie d'ingénierie des connaissances pour la représentation
des définitions lexicographiques dans le cadre de la théorie Sens-Texte
Maxime Lefrançois, Fabien Gandon, Alain Giboin 69

Les concordanciers bi-multilingues en ligne à l'épreuve de
l'approche d'« archive » : le cas des termes de l'égalité H/F
Rachele Raus 85

Étude des termes diagnostiques des troubles du langage en orthophonie
Frédérique Brin-Henry 111

La terminologie *in statu nascendi*
et ses implications au niveau sémantique.
Exemple de la terminologie française et polonaise du génie biomédical
Kaja Gostkowska 133

Objet, concept, terme, discours…
Repenser le domaine dans le cadre de la terminologie comparée
Héba Medhat-Lecocq 149

Du classement alphabétique à la représentation ontologique : la terminologie des systèmes d'enseignement supérieur
Maria Teresa Zanola, Mario Corveddu, Carol Rolla 173

Application de ressources linguistiques à grande échelle pour compléter une ontologie de domaine
Olena Orobinska , Jean-Hugues Chauchat, Natalya Charonova 177

Dynamicité et formalisation. Une structuration à facettes de la Terminologie
Natascia Leonardi 193

Are definitions syntactic structures or lexical frames ? A formal analysis based on Spanish evidence
César Aguilar 215

ARTICLES COURTS 235

Modelage d'un domaine économique conçu pour servir de feuille de route à la traduction spécialisée, offrant traçabilité des notions
Jan Roald 237

Multilingual and Information Retrieval Systems

A proposal to manage multi-word translation and ontology alignment
Johanna Monti, Mario Monteleone, Maria Pia di Buono, Federica Marano 241

Conférence d'ouverture

Ontology meets natural language : Does the lexicon have a regular structure ?

Christiane Fellbaum

Princeton University (USA)

1. Introduction : Lexicon, ontology

It has been estimated that speakers of American English with a high school education have a passive knowledge of approximately 60,000 root words (Miller 1991). The fact that speakers effortlessly and quickly access words in their lexicon when encoding and decoding messages challenges cognitive scientists to provide an explanation. It seems safe to assume efficient, economic mechanisms for storing and retrieving lexical items and the associated knowledge, although a complete, comprehensive theory of the organization of the lexicon is still outstanding. It is highly likely that the lexicon is well-structured rather than a random collection of words, and that lexicalization--the labeling of concepts with words--follows patterns that are accessible to speakers (though they are largely are unaware of them and unable to articulate them).

While questions about the structure of the mental lexicon is of concern to linguists and psychologists, philosophers and information scientists ask different, though very related questions. Their focus is on ontology, the attempt to formally represent knowledge as encoded in concepts. Both linguists and philosophers distinguish different types of words and concepts as well as their properties. Both have proposed similar organizational structures, in particular hierarchies of words and concepts constructed around hyponymy as well as additional relations. While linguists readily admit the idiosyncrasies of the lexicon, ontologists seek regularity and systematicity and focus on formal properties of concepts and conceptual organization that can make it possible to query knowledge and enable automatic reasoning and inferencing.

A major distinction between lexicons and ontologies is that the former, even when structured along ontological principles, typically restrict their coverage to the actual words and phrases of a language, whereas ontologies

are rich in concepts that are posited simply by virtue of providing desired distinctions and structure. An obvious question that arises is, how closely can the lexicon of a language be mapped onto an ontology, that is, to what extent are a lexicon and a formal ontology compatible ?

In this contribution, we ask whether an examination of the lexicon yield insight into categorization and the conceptual structure of its speakers and whether non-lexicalized concepts found in ontologies are merely accidental gaps in specific languages.

2. Categories in the lexicon

Studies focusing on the acquisition of categories and words (Keil, 1979, inter alia) show that children early on recognize the salient properties of entities and use this knowledge to identify new category members, initially overextending (e.g., by calling a goat a dog) but gradually learning the boundaries of a category. By analyzing speakers' judgments concerning the goodness of category exemplars, Rosch (1978) demonstrated that categories are formed around a central prototype with members that share more or fewer properties. These properties can form the basis of distinguishing similarity relations among category members and entire categories. Rosch's experiments, while pathbreaking, extended only to a few, concrete categories such as vegetables. The challenge is to show that the lexicon of language makes systematic, regular and productive use of categorization.

3. WordNet : A lexical ontology ?

To examine the structure of the English lexicon on a large scale, we created WordNet (Miller 1995, Fellbaum 1998 ; 2005). WordNet was manually constructed following principles borrowed from ontology, though without aiming to provide for automated reasoning and inferencing. Nonetheless, WordNet is often called a "lexical ontology" and has been mapped onto formal ontologies (Niles and Pease 2003 ; Gangemi et al. 2002, inter alia) and applied to reasoning tasks (Clark et al. 2008).

The atomic units of language are words, while those of ontologies are concepts. Human languages all exhibit synonymy, whereby a concept has multiple linguistic expressions, a fact of no interest to ontologists. To represent a concept level, WordNet's creators grouped words into roughly synonymous groups, called synsets. Synsets were then organized into categories,

with "sibling" members sharing a common superordinate synset. Each such tree constitutes a category, and smaller trees are arranged into a very large hierarchical structure.

The properties distinguishing category members from one another as well as from the common superordinate (genus) term are captured by definitions in their "classical" format. For example, a *mango* (defined as a *large oval tropical fruit having smooth skin*) was placed into a WordNet hierarchy below *fruit* as a sibling entry of *apple (fruit with red or yellow or green skin*)*, pear, banana*, etc. Following ontological principles, WordNet thus defines the class of fruit in terms of its members (mango, apple, etc.) and, conversely, defines the category members in terms of their shared superordinates.

WordNet currently includes over 155 000 unique word forms grouped into 116 000 synsets. Besides hyponymy (the super-/subordinate relation), WordNet encodes meronymy (the part-whole relation, as between *mango* and *pit*), antonymy (as between *hot* and *cold*) and a variety of entailment relations among event-denoting verbs (Fellbaum 1998). The resultant network of interrelated words and concepts represents a directed graph, which can be navigated by humans as well as automatic systems and which allows the measurement of semantic similarity in terms of distance among the words and synsets. Supporting WordNet's claim as a possible model of human semantic organization are word association norms, which show that speakers strongly associate antonymic words and words related by hyponymy and meronymy (Moss and Older 1996, inter alia) as well as corpus-based analyses of semantic similarity based on shared co-occurrence patterns (Justeson and Katz 1991).

4. Types, Instances and Roles

The diagnostic formula for hyponymy, *X is a Y*, hides in fact at least three different kinds of relations. One is the type or kind relation, which can be paraphrased with the formula *X is a type/kind of Y*, as exemplified by the *fruit-mango/apple/pear* example. The Instance relation in WordNet holds among concepts expressed by common nouns and specific instances such as persons, countries and geographic entities, expressed by proper nouns (or "named entities"). Thus, while an *armchair* is a type or kind of *chair*, *Barack Obama* is not a type or kind, but an instance of a president ; similarly, *Zambia* is an instance of a country and the *Rhine* is an instance of a river. Instances are always leaf (terminal) nodes of a hierarchy--no instance can have a subordinate. Miller and Hristea (2006) report on their manual classification of instances

in WordNet, showing that the type/instance distinction is not always clear-cut. For example, *Bible* can be seen as both an instance of *book* (the Christian Bible) but also as a type of book, with hyponyms like *King James Bible* and *Vulgate Bible*. This ontological distinction, while intuitively clear-cut, is in fact not always obvious when applied to the lexicon.

Another distinct relation encoded by the formula *X is a Y*, is that of role. Nouns like *president* and *chairman* encode roles rather than types. A hallmark of roles is that they are not time-independent: a person may bear a role such as chairman or president for only a limited time; by contrast, a person will always be a *type* of a biological creature. Types, but not roles, encode "rigid," or necessary, properties of an entity (Gangemi et al. 2002). Roles can be lexically encoded in word pairs that denote the related nouns. Such relational noun pairs include *parent-son/daughter* and *teacher-student*. (Biological relations such as *father-child* are rigid.) *Friend* implies at least two people in a friend-of relation to each other (a friendship-giving and a friendship-receiving person), though there is only one relational noun, *friend,* for both. WordNet currently does not encode roles distinct from types, but work is underway to detect role-bearing nouns in corpora semi-automatically by means of lexical-semantic patterns like *X used to be a Y*, where the former is a type and the latter a role noun (*this dog used to be a pet*) (Hicks and Herold 2009).

5. Accidental and systematic gaps

Though the WordNet experiment has shown that the English lexicon can be structured with classic ontological relations, the lexicon does not show the degree of regularity that is desirable in an ontology. In particular, the English lexicon shows "gaps" in regular patterns, where words appear to be missing. For example, nouns denoting tools can readily be turned into verbs that refer to the usage of that tool (*to hammer, to mop, to knife*, *to pencil*, etc.). But this seemingly productive pattern has not yielded English verbs like **to broom*, **to scissor*, **to sword*. Possibly, such verbs could be added; children produce them spontaneously, following the logic of the pattern. A related observation is that Dutch lacks a word equivalent to English container, though it has words for kinds of containers like basket, boxes and buckets. Such gaps are often language-specific and their presence in a particular language can be considered accidental rather then reflecting a general constraint on lexicalization.

Accidental gaps are unsystematic and randomly distributed across the lexicon. But there appear to be categorical constraints on encoding concepts.

Carter (1976) notes that many languages systematically encode causative verbs following the template in (1):

(1) x CAUSE (y BE Adjective)

For example, English has verbs like *kill* and *blind* (e.g., x CAUSE y BE *dead/blind = kill/blind).* However, Carter observes that the pattern in (2), which encodes a causation embedded inside another, is not found:

(2) y CAUSE (x Cause (z BE Adjective)

Possibly, there exists a limit on the complexity of concepts that humans can encode and process imposed by a more general cognitive constraint.

6. Covert categories

In striving to follow ontological practice, WordNet has allowed some "artificial nodes" into its network to fill in gaps that are arguably accidental rather than systematic and motivated. The implicit claim is that speakers have a concept but their language happens to lack a lexicalized form to refer to this concept. For example, to distinguish among the category of *vehicles* that run on wheels (*car, truck, bus*) and those that run on rails (*streetcar, tramway, train*), WordNet introduced the superordinates *wheeled vehicle and vehicle on rails.* This intuitively convincing solution avoids that *truck* is represented as similar to *tramway* as its sibling *car.* Perhaps more importantly, a system using WordNet in challenging deep language understanding tasks might need to recognize or infer critical differences between the two classes of vehicles.

7. Conclusion

Wikipedia defines ontology as *the philosophical study of the nature of being, becoming, existence, or reality, as well as the basic categories of being and their relations—fundamental notions that one would expect to be captured by* the lexicon of a natural language. A large-scale model of the English lexicon reveals that while the encoding of concepts and their interrelations follow regular patterns, there are both systematic and apparently unmotivated, accidental gaps. Nevertheless, humans effortlessly categorize and reason successfully, even over concepts that are not linguistically encoded. This suggests the existence of cognitive structures of the kind proposed by ontology.

References

Carter, R. (1976). Some Constraints on Possible Words. Semantikos l : 27-66.

Clark, P., Fellbaum, C., Hobbs, J., Harrison, P., Murray, W. and J. Thompson. (2008). Augmenting WordNet for Deep Understanding of Text. Proceedings of ACL-SigSem, Venice, Italy.

Fellbaum, C. Ed. (1998). *WordNet : An Electronic Lexical Database*. Cambridge, MA : MIT Press.

Fellbaum, C. (2005). WordNet and wordnets. In : Brown, Keith et al. (eds.), Encyclopedia of

Language and Linguistics, Second Edition, Oxford : Elsevier, 665-670.

Gangemi, A., Guarino, N., Masolo, C., Oltramari, A. and Schneider, L. (2002). Sweeting ontologies with DOLCE. In : Knowledge engineering and knowledge management :Ontologies and the semantic Web. Heidelberg : Springer, 166-181.

Hicks, A., and Herold, A. (2009). Evaluating Ontologies with Rudify. In Proceedings of *KEOD*, 5-12.

Justeson, J. S. and Katz, S. M. (1991). Co-occurrences of antonymous adjectives and their contexts. *Computational linguistics*, *17*(1), 1-19.

Keil, F. C. (1979). *Semantic and conceptual development : An ontological perspective.* Cambridge, MA : Harvard University Press.

Miller, G. A. (1991). *The Science of Words*. New York : Scientific American Library.

Miller, G. A. (1995). WordNet : A lexical database for English. Communications of the ACM 38 : 39-41.

Miller, G.A. and Histrea, F. (2006). WordNet nouns : Classes and instances. *Computational Linguistics*, *32*(1), 1-3.

Moss, H. and Older, L. (1996). *Birkbeck Word Association Norms*. Hove : Psychology Press.

Niles, I., and Pease, A. (2003). Linking Lexicons and Ontologies : Mapping WordNet to the Suggested Upper Merged Ontology. Proceedings of the IEEE International Conference on Information and Knowledge Engineering, pp 412-416.

Rosch, E. (1978). Principles of Categorization. In : Rosch, E. & Lloyd, B.B. (eds), *Cognition and Categorization*, Hillsdale, NJ : Lawrence Erlbaum Associates, 27-48.

Disputatio

Harris et les immunologistes : un exemple à suivre ?

Pierre Lerat

Université Paris XIII
pierre.lerat@wanadoo.fr

1. Introduction

« Le dernier Harris » (Habert et Zweigenbaum 2002 : 83), celui de l'Université Columbia, est celui que retiennent surtout beaucoup de contributions à l'ouvrage collectif *The Legacy of Zellig Harris*. Il tient en trois ouvrages, dont un en collaboration, celui qui nous intéresse. Le premier est un petit livre plus lisible que d'autres (probablement parce qu'il s'agit de conférences réunies par l'éditeur) qui s'intitule *Language and Information* (1988) et qui a été traduit en français (voir Ibrahim et Martinot 2007). Le dernier (1991) comporte aussi dans son titre *langage* et *information*. Entre les deux se situe un livre de près de 600 pages mais dont les deux tiers sont faits de tables à base d'extraits de rapports. Il s'appelle *The Form of Information in Science. Analysis of an Immunology Sublanguage*. Seules les 84 premières pages sont de Zellig Harris, mais les apports de ses collaborateurs aident à comprendre l'intérêt et les enjeux de l'entreprise.

Pour présenter en deux tomes sensiblement égaux l'hommage posthume de 2002, Bruce Nevin et Stephen Johnson ont distingué deux ensembles de textes : dans le tome I ce qui relève de la linguistique, dans le tome II ce qui concerne les mathématiques et l'informatique. En fait, la diversité des contributeurs, tous redevables à Harris d'une façon ou d'une autre, donne une idée du rayonnement de ce linguiste autour des trois disciplines (notamment en philosophie des sciences, en philosophie du langage et en didactique des sciences). Harris avait raison de dire qu'il était plus soucieux de méthode que de linguistique, même si c'est en linguistique qu'il a travaillé professionnellement.

Le livre collectif de 1989 utilise dans son sous-titre le mot *sublanguage* (en français *sous-langue* ? *sous-langage* ?). C'est cette notion qui est précisée ici à titre principal. Auparavant, et pas seulement pour tenir compte du titre, il paraît important de mentionner brièvement les rapports que Harris a entre-

tenus avec la science, avec la théorie de l'information et avec les approches formelles héritées principalement de Bloomfield. Ensuite, quand le concept de « sublanguage » aura été examiné, le moment sera venu de se demander en quoi ce « dernier Harris » peut intéresser les spécialistes de terminologies et d'ontologies.

Les travaux de Harris sont en partie accessibles en français, notamment grâce aux équipes de Maurice Gross et de Jean Dubois, en particulier dans la revue *Langages*. On ne s'étonnera donc pas de voir coexister ci-dessous des citations en anglais et des citations en français.

2. Harris et la science

2.1. Son entourage familial

Assistante de Von Neumann, puis d'Einstein, sa femme est une physicienne qui a cosigné avec le créateur de la théorie de la relativité généralisée plusieurs articles. Ce qui est plus important pour notre propos, c'est que de 1955 à 1960 elle a dirigé un programme de recherche intitulé « Mathematical Linguistics », en épouse parfaite.

De même, que T.N. et S. Harris, respectivement son frère et sa belle-sœur, aient écrit dans les meilleures revues biomédicales n'est pas ce qui importe le plus ici ; en revanche, la confiance faite au linguiste de la famille pour analyser des textes correspondant à divers états des connaissances en immunologie reflète une solidarité durable.

2.2. Parmi ses amis

Le souvenir qu'a laissé Harris à l'Université de Philadelphie, où il a été étudiant puis enseignant jusqu'à sa retraite (en 1969), est visiblement bon puisque beaucoup de ses collègues ont contribué à ses mélanges posthumes (il ne voulait pas entendre parler de mélanges, « affront à son sens des valeurs », aux dires d'un disciple téméraire). Il n'est pas question de dresser une liste qui se voudrait exhaustive. Un peu arbitrairement, mais dans l'ordre alphabétique, voici quelques admirateurs fidèles :

- Maurice Gross, selon qui dans son article de 1969 « Harris demonstrated an exceptional intellectual courage » (Gross 2002 : 60). Qu'est-ce qui justifie cet éloge ? C'est la décision de « considérer les phrases élémentaires comme les résultantes d'un opérateur (un segment verbal) agissant sur l'argument

primitif (les noms sujet et objets) » (Harris (1969) 1973 : 61)
- Stephen Johnson, éditeur du tome II de *The Legacy*
- Bruce Nevin, éditeur du tome I
- Hilary Putnam, ancien auditeur de Harris (1947), philosophe, auteur de la préface du livre collectif de 1969
- Thomas Ryckman, auteur d'une thèse sur l'information
- Naomi Sager, auteur du Linguistic String Project (1967-)
- Marcel-Paul Schutzenberger, membre du jury de thèse de M. Gross aux côtés de Harris, Jean Dubois et Robert-Léon Wagner

2.3. Un linguiste plus

Harris est ouvert à bien d'autres disciplines que la linguistique. En créant le premier département de linguistique des États-Unis à Philadelphie après la guerre, il s'entoure en particulier de collègues et futurs collègues qui, comme lui à ses débuts et comme Sapir, sont aussi nécessairement anthropologues parce qu'ils enquêtent auprès de locuteurs des langues amérindiennes. Nevin est l'un d'entre eux.

2.3.1. Un linguiste à la mode (1946-1960)

Le succès de Harris n'est pas un succès populaire : dans son séminaire de 1947 il n'a que deux étudiants, mais ce sont Chomsky (en majeure) et Putnam (en mineure). Ce succès n'en est pas moins éclatant : il devient président de la Linguistic Society of Linguistics en 1955. C'est un honneur qu'ont connu Boas, Sapir et Bloomfield, et que Jakobson aura en 1956. Chomsky, dont le père est professeur d'hébreu, prépare sous sa direction une thèse sur la grammaire de l'hébreu moderne. Son ingratitude prochaine sera donc doublement œdipienne. « His protégé Chomsky » (Matthews 1999)…

2.3.2. Un linguiste démodé (1960-1992)

En 2002, sollicité pour l'hommage posthume *The legacy*, « Noam Chomsky declined to contribute to the volume » (Nevin 2002 : xxvi). La phonologie générativiste fait oublier la phonologie structurale, la sémantique componentielle est à l'honneur aux USA comme en Europe, la phrase de base est devenue structure profonde. Harris continue à penser, comme Valéry, que ce qu'il y a de plus profond c'est la peau, en l'occurrence les corpus tels quels. Ce sont finalement surtout les informaticiens, notamment ceux qui travaillent dans les sciences médicales, qui rendent justice à

Harris, notamment Naomi Sager et Carol Friedman. En France, Maurice Gross, qui a travaillé avec lui à l'Université de Pennsylvanie, l'invite à faire à Vincennes en 1973 un séminaire qui sera publié en 1976 sous le titre austère de *Notes du cours de syntaxe*. Gross enseigne une linguistique plus lexicalisée sous le nom de *lexique-grammaire*, qui influence la contribution d'Anne Daladier au livre collectif de 1989, mais non Harris. Celui-ci reste fidèle à l'analyse de discours, de 1946 à sa mort en 1992. Cette fidélité ne déplaît pas à l'Université Columbia, qui l'invite en 1986 à faire les 4 conférences de Bampton qui paraîtront en 1988 sous le titre *Language and Information*, et qui sont une synthèse de son approche des langues et sous-langues.

2.3.3. Le statut des langues

La conclusion de *Language and Information* aurait pu se trouver dans les premiers écrits de Harris : une langue est « a self-organizing system growing out of real life conditions in combining sound sequences » (1988 : 113). L'intuition du « native speaker » est socialement conditionnée, comme le savent les dialectologues et les spécialistes des langues menacées. La grammaticalité, pour Harris, est donc « a circular criterion in any case » (1988 : 16). Quant au psychologisme, c'est pour lui du temps perdu ; dès 1941, dans son compte rendu des *Grundzüge der Phonologie* de Trubetzkoy, il déclarait que « particularly undesirable are psychological explanations » (Harris 1941 : 346).

2.4. L'inventeur des sous-langues

La définition d'une sous-langue est empirique quand elle désigne un corpus : il y a sous-langue « when what is said is limited to what seems relevant to the field » (Harris 1988 : 37). Cette définition est proche de celle de « technolecte »[1]. Elle est compatible avec une autre, plus abstraite, qui correspond à l'idée de « langue spécialisée »[2] : « discourse uses directly observable events whereas a sublangue is a construct » (Harris 1982b : 231). La première vaut pour les énoncés répertoriés dans le livre de 1989, la seconde pour les « formules » qui en sont tirées et qui en sont des

1 1 *Technolecte* : « tout ensemble langagier appartenant à un domaine de spécialité » (Leila Messaoudi, « Langue spécialisée et technolecte. Quelles relations ? », *Meta* 55-1, 2010 : 135).

2 *Langue spécialisée* : « une langue naturelle considérée en tant que vecteur de connaissances spécialisées » (Pierre Lerat, *Les langues spécialisées*, Paris, PUF, 1995 : 20).

généralisations.

L'auteur du système TAUM qui permet la traduction automatique des bulletins météorologiques à Radio Canada, Richard Kittredge, date de 1968 « the introduction of the notion of sublanguage by Zellig Harris » (Kittredge 1982 : 107). À partir de cette date, on peut observer chez Harris une prédilection pour une application concrète de sa méthodologie linguistique : « un système construit sur les sous-langues de la science afin que la corrélation de la structure et du système prédicatif devienne plus raffinée et que les relations particulières de la science soient mises en lumière de façon précise » (Harris (1969) 1973 : 79).

La notion de sous-langue n'est pas réservée aux sciences. C'est à celles-ci que s'intéresse Harris, mais l'expérience collective de 1989 ouvre des perspectives plus larges : « the sublanguage properties are not unique to science, but are shared by other uses of language in (relatively) closed domains, such as law, technical manuals etc. » (Mattick 2002 :34).

3. Harris et l'information

Divers linguistes, et non des moindres, ont utilisé les termes *information* et *code* de façon métaphorique par rapport à la théorie de l'information, à commencer par Saussure, Jakobson et Martinet (voir Ryckman 1990 : 23-26). Harris, au contraire, ne voit jamais dans la langue un code, et quand il parle d'information, ce qui est constant dans son œuvre, c'est pour appliquer aux langues naturelles des propriétés mathématiques reconnues aux systèmes de signaux et qui peuvent donc, jusqu'à un certain point, donner un fondement scientifique (statistique) au traitement d'une langue « as an information-bearing system » (Harris 1982a : 10-11).

3.1. Harris et Shannon

Il ne semble pas que Harris ait rencontré Shannon, mais il a eu des relations de travail avec Weaver. En gros, la théorie de l'information de Shannon et Weaver est une « théorie statistique du signal et des canaux de communication » (Fuchs et Habert 2004). Il n'est pas facile de calculer la probabilité de rencontrer un mot, mais c'est possible si l'on dispose d'un corpus écrit ou oral ; simplement, on disposera d'un état de langue à un certain moment et dans des conditions données. C'est mieux que rien, l'enquête des années 50 sur le « français fondamental » en témoigne. Les mots très fréquents sont des outils grammaticaux : « highest probability, lowest information » (Harris 2002a : 13). Dans le cas du « whole language » à grande échelle, il est pra-

tiquement impossible de calculer la « non-équiprobalité des combinaisons » (Harris 1982a : 9). Dans le cas d'un sous-langage, la fréquence des attestations est pertinente, surtout si elle caractérise non pas des mots mais des paires de mots : « in respect of each operator, many different arguments have different likelihoods » (Harris 1988 : 50). Par exemple, *administration de morphine* est une indication courante (normale) dans un service hospitalier de soins intensifs, *administration de bordeaux* est dépourvu de probabilité pour des raisons non linguistiques.

L'autre concept majeur en théorie de l'information, après celui de probabilié, est celui de redondance : « in mathematical information theory, developped within statistics, what was studied was the amount of information, as limited by the total amount of redundancy in the system » (Harris 1988 : 83). Qu'est-ce que la redondance dans un système grammatical ? C'est ce qui n'ajoute pas d'information mais qui la balise et la confirme.

Ex. : « vous ne croyez pas ? » (double marque de 2e personne du pluriel + double morphème négatif)
Il y a là un démarquage de l'aléatoire, « a departure from randomness, a redundancy » (Harris 1988 : 82). C'est une facilitation de la communication, et non pas une redondance au sens de la rhétorique.

3.2. Discours et information

L'information s'inscrit dans un processus de communication, et Harris la voit comme un phénomène combinatoire. L'information est grammaticale en tant qu'interprétation des structures syntaxiques ; elle est lexicale comme sélection de l'acception des mots au fil du texte. L'important est donc que l'information est un processus résultant du discours lui-même : « the information is neither independent of the structure nor additional to it, but is an interpretation of it » (Harris 1988 : 76). C'est ce qui dissipe l'ambiguïté dans la communication : *polysemy* est étranger à la métalangue linguistique de Harris. La polysémie fait partie des connaissances du locuteur et du récepteur, mais elle n'a pas sa place dans le discours, du fait de la clarification apportée par ce qui suit. Non seulement il y a une « step-by-step connection of information with structure » (Harris 1988 : 57), mais la communication inclut un feedback : « the required words are identified by what they in turn require » (Harris 1988 :14).

Cette conception exclut des corpus appropriés les textes comportant un jeu sur les mots. Ce qui est corrélé, ce sont « la forme syntaxique et l'infor-

mation objective » (Harris (1969) 1973 : 60). Ce n'est donc pas seulement un sentiment fraternel qui conduit Harris à dépouiller des textes scientifiques, c'est cette exigence d'objectivité. Nevin a bien résumé les conditions requises : « the linguistic information is conventional and it is objective to the degree that it is public, transmitted and shared by means of language » (Nevin 1994). L'information est présumée objective. « Dans cette conception, le langage procède moins d'un génome humain que d'une pratique sociale partagée » (Ryckman 1990 : 37).

3.3. Information et sens

Le fait que Harris ait délibérément, et depuis toujours, mis le sens entre parenthèses comme quelque chose de trop obscur pour une approche scientifique n'empêche pas que *sense* et *meaning*, qui chez lui sont synonymes (voir Ibrahim et Martino 2007 : 64, n. 45), soient clairement distincts d'*information* (voir Leeman 1996) : « words carry meaning and sentences carry information » (Harris 1988 : 57).

Au demeurant, sa conception du sens se rapproche plus de celle des terminologues que de la sémasiologie des générativistes dite « sémantique componentielle ». Les « propriétés » constitutives du sens correspondent à ce que l'on appelle en terminologie des caractères : « the meaning of a word might be characterized by the properties common to its possible set of referents » » (Harris 1988 : 64). Ce fondement onomasiologique vaut pour les entités référentielles, mais pour le vocabulaire exprimant des relations c'est la distribution qui constitue le critère.

On peut parler en pareil cas de « differential semantics » (Gottfried & Ryckman 1989 : 112). Plus précisément, le sens est indiqué par la structure prédicative : « the particular operator or arguments of the word indicate a particular meaning of it to the exclusion of the other meanings » (Harris 1988 : 90).

Ex : déclaration des droits de l'Homme / déclaration des droits d'auteur

Au niveau de l'énoncé global, les transformations ne changent pas l'« information objective ». Quant au jugement de paraphrase, il est le fait du récepteur compétent (« native speaker » et/ou expert du domaine). Par exemple, *Antibody appears in lymphocytes* et *Lymphocytes contain antibody* sont des « paraphrastic transforms » (Harris 1988 : 74). Ces énoncés en miroir tiennent compte de la construction des deux verbes et du sens des entités référentielles

désignées par *antibody* et *lymphocyte*. Ainsi, « content follows upon form and form follows upon content » (Harris 1991 : 354).

3.4. Information et référence

L'information n'est possible que sur fond de connaissances préexistantes. Ce ne sont pas les textes qui contiennent toutes les connaissances utiles à qui veut les comprendre, mais les lecteurs compétents. La référence est un problème d'énonciation, y compris dans les discours spécialisés : « resolution of referentials requires that the sublanguage be augmented by various implicit sentences » (Gottfried & Ryckman 1989 : 94).

Même l'anaphore lexicale peut dépendre de l'univers du discours considéré. L'exemple de Hiz « *steel... that metal* » (1969), comme ceux de Harris dans son livre de 1976, est gagé sur la relation entre le « classifier » et le « classificand », donc il est conforme à une hiérarchie lexicale classique : le passage de l'hyponyme à l'hyperonyme. On peut toutefois aisément imaginer des discours professionnels, au sein même des industries métallurgiques, où la propriété pertinente de l'acier serait d'être un matériau.

L'anaphore grammaticale est entièrement à la charge de l'auditeur ou du lecteur. Voici deux exemples qui se passent de commentaires.

Ex. 1 : « the injection has referential to the single occurrence of an antigen injection sentence » (Gottfried & Ryckman 1989 : 129)
Ex. 2 : « she […] the hearer has to infer the antecedent of such a pronoun as best as he can » (Harris 1988 : 78)

4. Harris et la forme

4.1. Harris et les distributions

La notion de distribution est liée à celle de contrainte. Un phonème s'identifie à partir de son comportement par rapport à un phonème qui le précède ou le suit immédiatement. Il en va de même pour les morphèmes, les mots et même les fragments de discours, d'après les articles des années 50. C'est dans l'un d'entre eux que l'on trouve la définition classique : « la distribution d'un élément sera définie comme la somme de tous les environnements de cet élément » (Harris (1954) 1970 : 14). L'idée de contrainte est exprimée de façon plus explicite par l'un de ses disciples : il s'agit de « restrictions on how pieces can be joint together » (Goldsmith 2005 : 721). Cette idée a été fondatrice en

phonologie : c'est elle qui, par exemple, à la fin du XIX[e] siècle, sous-tend le concept de semi-consonne (ou semi-voyelle) : *y* et *w* en indo-européen sont voyelles devant consonnes et consonnes devant voyelles.

Chez Harris, l'idée de distribution est associée à celle de discours : « les règles de distribution sont toujours faites à partir d'un corpus d'énoncés » (Harris (1954) 1970 : 32). Elle est appliquée à la structure des énoncés, à commencer par les phrases de base, qui correspondent à des propositions logiques élémentaires. Elle constitue même la première contrainte retenue par Harris : « there is a linear order among the coarguments » (1988 : 13). Ainsi, en anglais, le sujet précède le verbe, qui précède les compléments. Il en va de même en français.

La méthode distributionnelle est particulièrement féconde quand elle est appliquée aux structures prédicatives, où l'opérateur et l'opérande s'enchaînent. Plus généralement, elle rend compte de l'unité à partir des couples où celle-ci figure, « making meaning partly a word-pair property rather than a word property » (Harris 1988 : 108).

4.2. Harris et les transformations

La notion de transformation vient des mathématiques, par analogie avec les déplacements et les retournements en géométrie. Elle intéresse Harris depuis la fin de l'entre-deux-guerres d'après le témoignage de Nevin : « he was talking about transformations in his seminars as early as the end of 1939 » (2002 : xvii). Elle vise à rendre compte des phrases complexes où l' « information objective » est la même et où le vocabulaire est constant mais qui résultent de déplacements en tête (dans l'interrogation, en anglais), d'insertions (dans la proposition relative), d'ajouts (dans les complétives) etc.

Anne Daladier a bien vu que pour Harris la transformation est une « méthode de régularisation » (1990 : 59). Ce n'est pas pour rien qu'il privilégie les réductions : l'important est de rendre compte par une transformation de ce qui va contre l' « ordre linéaire » dans le discours. L'expérience dont les résultats sont publiés en 1989 instrumentalise complètement la méthode. Pour régulariser les énoncés analysés afin de dégager des « formules » faites de propositions élémentaires, les linguistes de l'équipe procèdent à des ajustements transformationnels entre le complexe et le simple. Ce sont des linéarisations et délinéarisations (Ryckman & Gottfried 1989 : 89), des nominalisations et dénominalisations (Gottfried et Ryckman 1989 : 116) ou encore une « depassivization » (*ibid.* : 147).

4.3. Les réductions

La densité des textes scientifiques et techniques est une caractéristique « stylistique » bien connue. Elle s'obtient par des ellipses de mots-outils (ex. : *bec Bunsen*) ou d'arguments évidents (ex. : l'observateur dans une expérience). Plus généralement, la réduction pour Harris est une question d'information : « high-likelihood words may be reduced, even to zero » (1988 : 18).

Ex. : « I knocked and entered » (Harris 1988 : 8) : celui qui frappe est celui qui entre, la mise en facteur commun du pronom limite la redondance.
La réduction d'une relative est une limitation courante de la redondance dans les comptes rendus d'immunologie étudiés par l'équipe de Harris.
Ex. : *pathogenic bacteria which are carried* peut sans perte d'information être réduit à *pathogenic bacteria carried* : « *which are* is zeroable as bearing low information » (Gottfried & Ryckman 1989 : 133).
Dans des comptes rendus cliniques Friedman et son équipe observent un type de réduction qui concerne cette fois la structure prédicative : *joints were painful* (et aussi *pain in joints* et *joint pain*)
suffit tout à fait, le malheureux patient étant une évidence du discours.

4.4. Une linguistique minimaliste

Le titre d'un article de Nevin en hommage à Harris, « a minimalist program » (1993) est aussi celui d'un livre de Chomsky, mais il y a minimalisme et minimalisme. Chez Harris, la démarche est « from the bottom up » (Nevin 1984 : 205), chez Chomsky elle est « top down ». En partant d'observables (des discours réels), Harris dégage des régularités par une « méthode constructiviste », ainsi que le disent à la fois Daladier et Ryckman dans leurs articles de 1990.

Le minimalisme de Harris est une économie de moyens. La prédication est à la fois la formulation et le sens de la phrase élémentaire, et la phrase complexe est faite de « prédications de prédications » (Harris 1990 : 18). La prédication est donc un moteur récursif de la communication verbale ; le mot *récursivité* ne fait pas partie du vocabulaire de Harris, mais sa méthode est récursive.

En s'intéressant de façon de plus en plus privilégiée aux sous-langues, Harris découvre dans les phrases spécialisées une prévisibilité qui est très caractéristique des sciences et des techniques : celle des arguments pertinents. C'est peut-être Naomi Sager qui a le mieux dit en quoi un discours spécialisé est un discours fortement contraint : « in the sublanguage many of the science

specific verbs have only one or two objects possibilities, fewer than in their use in english as a whole » (Sager 2002 : 12).

L'emprunt du concept de prédication à la logique (mutatis mutandis) autorise le passage de la phrase simple à la notation polonaise et réciproquement ; « *His smoking continues* = continue (smoke (he)) » (Harris (1969) 1973 : 62). C'est principalement en mettant au centre de sa méthode la prédication, à partir de 1969, que Harris s'approche le mieux de son objectif : « a mathematical object, a system defined by its relations alone » (Harris 1988 : 89).

4.5. Les opérateurs

Selon les textes (et selon les traductions), on trouve chez Harris le couple « opérateur-opérande », « prédicat-argument » ou (le plus souvent) « opérateur-argument ». L'important est que « operators are predicates in a broad sense and therefore are sentence-forming » (Hiż 1994). Comme *prédicat*, *opérateur* est à prendre « au sens large » : « chaque opérateur opère sur un argument pour produire une phrase » (Harris (1969) 1973 : 62).

Comme les prédicats de la logique, les opérateurs de Harris peuvent relever du premier ordre (phrase verbale simple) ou être d'ordre supérieur.

Ex. : (premier ordre) : *the man sleeps* / (ordre supérieur) : *He asserts that* (Harris 1988 : 13)

De ce fait la diversité des opérateurs fait que les verbes sont loin d'être les seuls opérateurs. Voici quelques exemples :

- « three operators : *eat*, *close to* and *father of* » (Harris 1988 : 104)
- « the operator on the nouns may appear as adjective » (N. Sager 1982 : 15)
- « les noms à deux places sont des termes relationnels (*synonymy*, *father of*) » (Harris (1968) 1970 : 229)
- « and (left, he), (left, she) » (Harris (1969) 1973 : 72)
- « an operator like *after*, *following* etc. » (Mattick 2002 : 47)

5. Un livre pluridisciplinaire

Le projet de ce livre est ancien. Il n'est pas possible de le dater, mais le témoignage de l'un des proches collègues de Harris indique que « for many years he and his cowriters studied the sublanguage of immunology » (Hiż 1994). Ce type de discours l'intéresse depuis longtemps ; ainsi, comme « dis-

cours scientifique stylisé », les « rapports d'expériences » sont un bon exemple (Harris (1954) 1970 : 21). En outre, en 1988 (p. 38), il reprend textuellement un exemple déjà utilisé dans un ouvrage sur les sous-langues (1982b : 232) : « we can find *we washed the polypeptides in HCl* but not *we washed the HCl in polypeptides* ».

5.1. Une entreprise coopérative

L'association de linguistes, d'immunologistes et de philosophes des sciences permet des éclairages différents sur un corpus significatif : « un nombre raisonnable d'articles (environ 15) pour l'anglais et le français » (Ryckman 1990 : 34).

- Une équipe de linguistes : Harris, Gottfried, Ryckman, Daladier
- La participation d'une Française présente un double intérêt : apport d'un corpus dans une langue autre que l'anglais et validation de la méthode du point de vue d'une francophone.
- Un couple d'immunologistes : « S. Harris and T.N. Harris served as immunology informants » (Gottfreid 2002 : 96, n. 11).
- Des philosophes des sciences : Putnam (auteur de la préface) et Mattick (ami et voisin)
  Mattick valide les résultats en y reconnaissant des phrases porteuses de vérité : « science sentences have the truth-bearer characteristics » (Mattick 2002 : 52).

5.2. Des apports complémentaires

Cette expérience est la rencontre d'un vocabulaire et d'un univers de discours. Du point de vue du vocabulaire, le plus important est la relation entre hyperonyme et hyponyme. En 1976, les « classifieurs » subsument des classes en extension. En 1989, ce sont clairement ce que les lexicologues appellent des hyperonymes : « classifiers, i.e. words whose meaning is that of all members of the class or of the type » (Harris 1989 : 57). Un intérêt des classifieurs est qu'ils fournissent de bons définisseurs : « the semantic relation of a classifier to its classificand is largely definitional » (Ryckman & Gottfried 1989 : 123), par exemple dans le cas suivant : « the classifier-classificand relation stated in *lymphocytes are substances* » (Mattick 1989 : 158). Un autre intérêt est qu'ils se situent à un niveau de généralité où la communauté mondiale des spécia-

listes est d'accord sur les équivalents : « The words for *antigen*, marked *G*, were either synonyms of it or names of various antigens » (Harris 1989 : 42).

L'autre relation majeure est la « synonymie locale », classe d'équivalence lexicale en discours, donc de validité limitée à une structure prédicative donnée.

> Ex. 1 : « the antibodies present (or contained) in the lymph nodes » (Gottfried & Ryckman 1989 : 126). Il est clair que l'adjectif *present* et le participe *contained* ne sont pas interchangeables dans beaucoup d'énoncés, mais seulement entre les mêmes arguments.
>
> Ex. 2 : « *normal* was used to mean *not injected* » (Harris 1988 : 43).
>
> L'univers du discours est celui des immunologistes, et il inclut les faits aussi bien que les entités.
>
> Ex. 1 : « what relations *can* be stated among sublanguages » (Harris 1989 : xvi)
>
> Ex. 2 : « for a cell, to divide *is* to multiply » (Harris 1988 : 42).

On est renvoyé à une ontologie de domaine, et l'exemple 2 montre à quel point celle de l'immunologie n'est pas celle des mathématiques.

6. Intérêt pour les terminologies

6.1. Complémentarité

Par rapport aux terminologies, qui restent dépendantes du modèle triangulaire incluant l'entité nommée, les sous-langues au sens de Harris sont englobantes et ne constituent pas un simple étiquetage documentaire. Les sous-langues sont dans les langues : « the sentences of the sublanguages are also sentences of the language » (Harris 1988 : 39). En particulier, les verbes sont à l'honneur parce qu'ils rendent compte d'états (et de changements d'états), de processus et d'interventions : « a world of facts », déclare Mattick (1989 : 165).

La prise en compte de la dimension syntagmatique est également une avancée importante. Il ne s'agit pas de simples cooccurrences, collocations ou figements, mais d'une réalité transversale : une syntagmatique des structures prédicatives propres à un domaine. Si l'on regarde par exemple les « infor-

mational categories of the clinical domain », on trouve surtout des états, des processus et des actions « diseases, procedure, medication » (Friedman *et al*. 2002 : 225), à côté d'un seul nom d'entité référentielle (*body location*).

Un autre acquis important est que le corpus ne contient pas toutes les connaissances utiles à sa compréhension : « that may require knowledge from outside its research articles, and even outside the science itself » (Harris 1989 : xvi).

Parmi les autres raisons de ne pas compter exagérément sur des dépouillements automatiques, la plus importante est que l'importance des structures prédicatives rend insuffisants les logiciels pour lesquels l'unité de compte est le mot. Il est certain en effet que « les phrases d'une langue naturelle n'indiquent pas directement la relation argument-opérateur » (Harris (1969) 1973 : 60). En outre, contrairement à une idée reçue, passé un certain seuil un corpus devient inutilement encombrant : « as the corpus of material being investigated grows, the set of sentence-types grows little at all » (Harris 2002 : 10).

Dans ces conditions, les analyseurs syntaxiques à base de parties du discours se heurtent à une grave difficulté : dans les sous-langues, « verbs frequently occur in the nominal form » (Friedman *et al*. 2002 : 225). Par exemple, en français, une location peut être soit un bien loué soit la location d'un bien. L'immunologie n'échappant pas à un fait de langue aussi massif, Harris a dû admettre que « the work is to make by hand » (1989 : 12). Habert et Zweigenbaum (2002 ; 93) voient là une curiosité de son travail, mais c'en est une contrainte d'origine linguistique.

Carol Friedman a étudié des rapports cliniques à la façon de Harris. Elle a bien vu l'intérêt de sa méthode : « the sublanguage operators reflect the salient relations and arguments that are meaningful in the specialized domain » (2002 :224) et c'est important. Comme Kittredge, elle a néanmoins cru bon de compléter cette approche par le recours à un vocabulaire contrôlé, « which consists of well defined or standard concepts corresponding to words and phrases in the domain », convaincue que « a controlled vocabulary is important for computerized applications » (2002 : 226).

6.2. Vers une lexicographie spécialisée

Un vocabulaire contrôlé est un dictionnaire bien fait. Anne Daladier, alors membre du laboratoire de Maurice Gross et adepte de l'option « lexique-grammaire », évoque dans sa contribution « our dictionary » (1989 : 186). Le contexte ne permet pas de savoir si c'est un projet, une esquisse pour le fran-

çais ou une question débattue avec Harris. Quoi qu'il en soit, celui-ci a sur les dictionnaires des idées difficiles à mettre en œuvre au niveau d'une langue mais réalisables avec profit au niveau d'une sous-langue : « a dictionary that goes into detail on the meaning of words has to give examples of the use of the given word in respect to the others. This relation among meanings is partly circular, but it shows the role of word combination in specific meaning » (Harris 1988 : 63).

Parmi les relations opérateur-opérande, la terminologie privilégie celles qui chez les informaticiens s'appellent *is_a* et *part-of*. Elles sont de celles qui se prêtent à cette circularité restreinte (du classifieur au subordonné et réciproquement, de la partie au tout et réciproquement). De même, la relation dite *causale* par Ryckman et Gottfried (1989 : 98) et évoquée par l'ISO comme une simple relation « associative » parmi d'autres (« séquentielle » et « temporelle ») est une relation opérateur-opérande car « *consequence of* is a relational noun » (Hiż 1969 : 34).

Une autre idée propre à privilégier les concepts spécialisés est proposée par Daladier : « list the morphologically related words, such as *produire*, *production*, *producteur* » (Daladier 1989 : 186). Il est clair que la polysémie du mot produirait des effets pervers dans un dictionnaire général, comme l'a montré une expérience lexicographique malheureuse chez Larousse. En revanche, en lexicographie spécialisée, le risque est faible. Ainsi, dans un dictionnaire juridique, *produire une preuve* ou *production d'une preuve* dit le même contenu, et *producteur* n'a pas sa place.

Enfin, une idée de Johnson, informaticien de Penn et co-éditeur de l'hommage de 2002, est que, si des ressources plus ambitieuses sont difficiles à mettre en œuvre, il est toujours possible d'utiliser des outils bureautiques : il parle, à propos de l'expérience réalisée avec des immunologistes, d' « interesting possibilities for the design of flexible and expressive databases » (Johnson 1989).

7. Intérêt pour les ontologies

7.1. Formules et relations entre formules

Les formules « représentent des phrases canoniques » (Ryckman 1990 : 35). « Each fact-sentence in the science can be written as a formula » (Harris 1989 : 1). Ce sont des structures prédicatives utilisant un vocabulaire générique. On pourrait se contenter de phrases-types en langue naturelle, avec

des classifieurs. L'usage dans les sciences étant plutôt de recourir à des symboles, les auteurs ont choisi de remplacer les classifieurs par des majuscules. Il y a là une convention maison qui présente plusieurs inconvénients. Au sein même de l'anglais, remplacer *antigène* par *G* est arbitraire ; hors de l'anglais, *B* pour *body* est mal venu. Mieux vaut, semble-t-il, « a paraphrastic pattern » (Friedman 2002 : 225), fait de classifieurs.

Les formules ont l'avantage de la brièveté. Ainsi, *GJB* = un antigène est injecté dans une partie du corps. Il vaut sans doute mieux utiliser en clair des mots qui ont une traduction internationale parce qu'il s'agit de concepts partagés par la communauté mondiale en la matière.

> Ex. 1 : « PERSON have SYMPTOM in BODY LOCATION » (Friedman *et al*. 2002 : 225)

> Ex. 2 : « Dichlofenac has_side_effect Gastrointestinal Bleeding » (Schultz *et al*. 2009 : 33)

Une relation intéressante entre les formules est leur enchaînement temporel et/ou causal : « a relation between facts acquires the status of fact itself » (Mattick 2002 ! 50). L'enchaînement est noté par « : ».

> Ex. 1 : *GJB : AVC* « permet d'exprimer qu'une injection d'antigène est suivie d'une réponse immunitaire » (Ryckman 1990 : 35)

> Ex 2 : preuve (crime contre l'humanité, personne physique i) : comparution devant (personne physique i, Tribunal pénal international)

L'implémentation de formules est encore plus problématique que celle des structures prédicatives. Harris reste prudent à cet égard : « the work can in principle be carried out by a computer program, even through considerable complexity in so doing » (1989 : 7). Johnson (1989) évoque plusieurs pistes, sans en préconiser une : « Formulas can be implemented using Prolog, relational data tables, a semantic set, un object-oriented system or a frame-based representation. » Aux spécialistes de juger.

7.2. Le dit et le vrai

On peut tirer de textes une ontologie, moyennant l'expertise du domaine et une mise en forme logique[3]. L'une des difficultés majeures provient d'une différence importante entre les langues naturelles et les langages symboliques : « natural language is a much richer system than logical calculus » (Mattick 1989 : 160). Une autre différence, plus radicale encore, est que l'opérateur grammatical n'est pas une prédication de vérité : « a sentence is said, it is not asserted as true » (Mattick 1989 : 165).

Catherine Fuchs observe avec raison que chez Harris « l'énonciation n'est pas distincte de la prédication » (1986 : 228). Il importe donc que seuls soient pris en compte des discours bénéficiant d'une présomption de vérité, ce qui conduit par une nouvelle force des choses vers les textes scientifiques, où ce qui est dit est présumé vrai, du moins pour une communauté de travail donnée à un moment donné. Harris, pour qui la véritable élégance est l'explicitation totale, souligne que « the structure (« grammar) of the formulas accords with the fact-structures of the science » (1989 : xv). C'est un point important, effectivement.

Reste un travail qui incombe aux spécialistes des ontologies : remplacer les « : » par des formalismes plus rigoureux. En particulier, les enchaînements ne sont pas tous des formulations d'inférences : « *so*, *thus* distinct from logical *if... then* » (Mattick 1989 : 166).

Pour finir sur une note optimiste, une justification importante du travail que demande l'exploitation des résultats de l'équipe réunie par Harris est que les concepts spécialisés sont fortement partagés internationalement, ce qui réduit les vocabulaires concernés à de simples étiquettes[4] par langue. « The sublanguage of a given science in one language is similar to the sublanguage for that science in another language » (Harris 1989 : 80).

8. Conclusion

Le titre d'un article du fidèle Ryckman (1990) résume bien l'évolution de Harris : « De la structure d'une langue aux structures de l'information dans

3 Une ontologie, rappelons-le, est une « spécification formelle des concepts et de leurs relations » (selon la définition de Christophe Roche dans son article « Faut-il revisiter la terminologie ? » (in *TOTh* 2008, Annecy, Institut Porphyre, p. 65).

4 Voir Guadalupe Aguado de Cea « Terminología, ontologías y multilingualidad », *Puntoycoma* n° 115 S, p. 5 ; http://ec.europa.eu/translation/bulletins/puntoycoma/115/pyc115_su.

les discours et dans les sous-langues scientifiques ». Évolution, et non pas révolution, car pour accéder à la structure des langues il est toujours parti de discours réels, y compris dans sa thèse sur la grammaire du phénicien, langue morte accessible seulement dans des corpus – au sens philologique – d'inscriptions.

Dans quelle mesure peut-on parler d'un « dernier Harris », dans ces conditions ? Célèbre comme promoteur du structuralisme, puis pourfendu à ce titre, il est resté fidèle à un programme qui s'applique d'abord témérairement à tout discours suivi, avec pour principes une mathématisation des descriptions et un respect scrupuleux des faits de discours observés. En voulant donner à l'étude des discours une rigueur comparable à la fois à celle des phonologues et à celle de sa femme physicienne et de son frère immunologiste, il a manqué longtemps des moyens de sa politique. Ceux-ci sont apparus en 1969, par le recours systématique à un schéma prédicatif au sens de Frege. Le jeu des relations entre opérateurs et opérandes circonscrit la sémantique dans un implicite partagé socialement (dans les mots) et dans une information discursive (interprétation des structures prédicatives par le récepteur). Mais c'est seulement dans les sous-langages que les contraintes de sélection permettent un rendement satisfaisant de la méthode. Il est important que le livre de 1989 ait fini par trouver un éditeur.

C'est en effet ce travail coopératif qui valide par des contraintes fortes une façon nouvelle de travailler. Ce sont d'abord des contraintes externes : comme les textes ne sont pas forcément porteurs de vérités, les conditions optimales sont celles où la vérité de leurs affirmations est validée par une expertise reconnue. La faiblesse de l'analyse de discours des années 50 résulte du manque de fiabilité du corpus : un slogan publicitaire selon lequel « des millions d'Américains ne peuvent pas se tromper » est sujet à caution, bien avant la crise des subprimes ; il en va tout autrement dans les textes scientifiques ou techniques, jusqu'à l'émergence d'un nouvel « état de l'art ». C'est également en pareil cas que les contraintes internes sont très fortes. Nevin l'a bien vu : « technical sublanguages of science have more well-defined semantic and syntactic restrictions ».

Il fallait aussi un travail à la fois pluridisciplinaire et coopératif pour que l'on dispose de résultats significatifs. Un exemple à suivre, évidemment, et même à suivre de près.

Bibliographie

Daladier, Anne (1989) « Information Units in a French Corpus » in Harris *et al*. 1989 p. 170-191

Daladier, Anne (1990) "Aspect constructif des grammaires de Zellig Harris", *Langages*, n° 99, p. 57-84

Friedman, Carol, Kra, Paulina & Rzhetsky, Audrey (2002) "Two biomedical sublanguages : a description based on the theories of Zellig Harris", *Journal of Biomedical Informatics*, 35, p. 222-235, www.sciencedirect.com

Fuchs, Catherine (1986) « Harris, ou l'énonciation esquivée », *Histoire-Épistémologie-Langage* VIII-2, p. 221-231

Fuchs, Catherine & Habert, Benoît (2004) "Le traitement automatique des langues : des modèles aux ressources", *Le français moderne*, LXXII-1, *halshs archives-ouvertes.fr*

Goldsmith, John A. (2005) review of *The Legacy of Zellig Harris*, vol. 1, *Language*, 81-3, p. 719-736, http://hum.uchicago.edu/~jagoldsm/Papers/ZelligHarris.pdf

Gottfried, Michael (2002)"Grammatical specification of scientific sublanguages" in *The Legacy of Zellig Harris*, vol I, p. 89-102

Gottfried, Michael & Ryckman, Thomas (1968) "The Apparatus of Sublanguage Transformations" in Harris *et al.* 1989, p. 112-150

Gross, Maurice (2002) 'Consequences of the Metalanguage beeing included in the Language' in *The Legacy of Zellig Harris*, vol. II, p. 57-67

Habert, Benoît & Zweigenbaum, Pierre (2002) « Régler les règles », *TAL*, 43 (3), p. 83-105, http://net.u-paris10.fr/habert_benoit

Harris, Zellig (1941) review of Trubetzkoy's *Grundzüge der Phonologie*, *Language* 17-4, p. 345-349

Harris, Zellig (1946) « From Morpheme to Utterance », *Language*, 22-3, p. 161-183 ; trad. fr. "Du morphème à l'expression", *Langages*, 9, 1968, p. 23-50

Harris Zellig (1951) *Methods in Structural Linguistics*, Chicago, Chicago University Press

Harris, Zellig (1952) « Discourse Analysis », *Language* 28-1, p. 1-30 ; trad. fr. « Analyse du discours », *Langages* 13, 1969, p. 8-45

Harris, Zellig (1954) « Distributional Structure », *Word*, 10, n° 2-3, p. 146-162 ; trad. fr. *Langages*, 20, 1970, p. 14-34

Harris, Zellig (1968) *Mathematical Structures of Language*, New York, Interscience Publishers ; trad. fr. C. Fuchs (1970) *Structures mathématiques du langage*, Paris, Dunod

Harris, Zellig (1969) "The Two Systems of Grammar : Report and Paraphrase" in *Papers in Structural and Transformational Linguistics*, Dordrecht, Reidel, 1970, p. 612-692 ; trad. fr. « Les deux systèmes de la grammaire : prédicat et paraphrase », *Langages*, 29, 1973, p. 55-81

Harris, Zellig (1976) *Notes du cours de syntaxe*, trad. et introd. M. Gross, Paris, Larousse

Harris, Zellig (1982a) *A Grammar of English on Mathematical Principles*, New York, John Wiley

Harris, Zellig (1982b) "Discourse and Sublanguage" in *Sublanguage. Studies of Language in Restricted Semantic Domains,* p. 231-245

Harris, Zellig (1988) *Language and Information*, New York, Columbia University Press

Harris, Zellig (1990) "La genèse de l'analyse des transformations et de la métalangue", *Langages* 99, p. 9-20 (trad. angl. Harris 2002)

Harris, Zellig (1991) *A Theory of Language and Information* : *A Mathematical Approach,* Oxford, Clarendon Press

Harris, Zellig (2002) "The background of transformational and metalanguage analysis" in *The Legacy of Zellig Harris*, vol. I, p. 1-15

Harris, Zellig, Gottfried, Michael, Ryckman, Thomas, Mattick Jr, Paul, Daladier, Anne, T.N. & S. Harris (1989) *The Form of Information in Science. Analysis of an Immunology Sublanguage*, Dordrecht, Kluwer Academic Publisher

Hiż, Henry (1969) « Referentials in Transformations and Discourse Analysis Papers » (1968), *Semiotica* 2, p. 136-166

Hiż, Henry (1994) « Zellig Harris », *Proceedings of the American Philosophical Society*, 138-4, p. 519-527, www.ircs.upenn.edu

Ibrahim, Amr-Helmy & Martinot, Claire (2007) *La langue et l'information*, Avant-propos, traduction et index du livre de Harris *Language and Information*, Paris, Cellule de Recherche en Linguistique, http://crl.exen.fr

Johnson Stephen B. (1989) review of *The Form of Information in Science*, http://www.ircs.upenn.edu/SJ-FISrev.htm

Kittredge, Richard (1982) "Variation and Homogeneity of Sublanguage" in *Sublanguage. Studies of Language in Restricted Semantic Domains*, p. 107-137

Kittredge, Richard & Lehrberger, John eds. (2002) *Sublanguage. Studies of Language in Restricted Semantic Domains*, Berlin-New York, Walter De Gruyter

Leeman, Danielle (1996) « Le « sens » et l'« information » chez Harris », *LINX*, 8, p. 209-220

Matthews, Peter (1999) « Zellig Sabbetai Harris », *Language* 75-1 p. 112-119, http://www.ircs.upenn.edu/zellig/ZSH-Matthews.htm

Mattick, Paul (1989) « Extending the Analysis : the Informational Environment of the Science Sentences » in Harris *et al* 1989, p. 151-168

Mattick, Paul (2002) "Some implications of Zellig Harris's work for the philosophy of science" in *The Legacy of Zellig Harris*, vol. I, p. 39-55

Nevin, Bruce (1984) compte rendu de *A Grammar of English on Mathematical Principles*, *Computational Linguistics*, vol. 10, 3-4, p. 203-211, http://www.ircs.upenn.edu/zellig/gemp.html

Nevin, Bruce (1988) compte rendu de *Language and Information*, *Computational Linguistics*, 14-4, p. 87-90

Nevin, Bruce (1993) "A minimalist program for linguistics. A perspective on the work of Zellig Harris" in *Historiographia Linguistica*, vol. 20, 2-3, p. 355-398, www.ircs.upenn.edu/zellig/Minimalist.html

Nevin, Bruce & Stephen Johnson eds. (2002) *The Legacy of Zellig Harris*, 2 vol., Amsterdam, John Benjamins

Ryckman, Thomas (1990) « De la structure d'une langue aux structures de l'information dans le discours et dans les sous-langages scientifiques », *Langages*, 99, p. 21-38

Ryckman, Thomas (2002) "Method and Theory in Harris's Grammar of Information" in *The Legacy of Zellig Harris*, vol. I, p. 19-38

Ryckman, Thomas & Gottfried, Michael (1989) "Sublanguage Formulas as Information Units" in Harris *et al*. 1989 p. 85-111

Sager, Naomi (1982) "Syntactic formatting of science of information" in *Sublanguage. Studies of Language in Restricted Semantic Domains*, p. 9-26

Schultz, S., Stenzhorn, H., Boecker, M. & Smith, B. (2009) "Strenghts and limitations of formal ontologies in the biomedical domain", *Electronic Journal of Communication, Information and Innovation in Health*, vol. 3, 1, p. 31-45

Articles

Systèmes conceptuels et isotopies métaphoriques dans les langues de spécialité : analyse contrastive de quelques domaines en français et en italien

Micaela Rossi*

*Università di Genova Dipartimento di Lingue e Culture moderne
Piazza Santa Sabina, 2 16124 Genova
micaela.rossi@unige.it

Résumé. Notre étude se propose de mettre en évidence la fonction modélisatrice des isotopies métaphoriques qui sous-tendent de nombreuses terminologies de spécialité ; en particulier, nous nous intéresserons aux mécanismes sémiotiques et aux dynamiques cognitives qui sont à la base de métaphores constitutives contribuant à la formalisation d'ontologies des domaines spécialisés. Notre recherche s'appuiera sur des corpus parallèles en français, en italien et en anglais, appartenant à différentes langues de spécialité.

1. Introduction

Qu'entend-on lorsqu'on parle d'un *nain* dans le langage technique des chemins de fer ? Et quel est l'objet du discours quand notre médecin utilise le terme *langue géographique* ? Un vin *maigre* est-il plus prisé qu'un vin *gras* ? Ces questions mettent l'accent sur l'omniprésence des termes métaphoriques dans les langages techniques et scientifiques.

Pendant longtemps, sur la base de l'approche wüstérienne ou Théorie Générale de la Terminologie, les terminologues ont réduit ces termes métaphoriques à de simples étiquettes apposées sur les concepts spécialisées, dépourvues de tout rapport avec l'analogie qui les avait générées, réduites au rang de simples dénominations catachrétiques. Ce n'est que pendant les dernières décennies que les linguistes et terminologues ont commencé à s'intéresser à la nature sémiotique et aux dynamiques textuelles impliquant la présence de métaphores dans les langues de spécialité. À partir des études d'Assal (1994), les métaphores terminologiques ont fait l'objet d'études approfondies dans le

domaine francophone (entre autres, voir par exemple Temmerman 2000 et 2007, Gaudin 2003 Cortès 2003 et Oliveira 2005 et 2009, Rossi 2015), mais également et surtout dans le domaine de la linguistique cognitive d'empreinte anglo- saxonne (nous renvoyons par exemple à Bowdle et Gentner, 2005, mais également à Gibbs, 2008 pour un aperçu général). Les études des terminologues et des linguistes rejoignent dans les années 2000 les analyses des épistémologues (Montuschi, 1993 ; Schlanger 1995 ; Brown, 2003), qui, eux, s'intéressent plutôt au pouvoir de modélisation de la métaphore dans les théories scientifiques.

Les fonctions de la métaphore dans les langues de spécialité, loin de se limiter à la formation de catachrèses terminologiques, sont donc complexes et multiformes ; Oliveira (2009) dresse un premier inventaire des emplois de la métaphore dans les sciences, énumérant – outre la fonction dénominative, la fonction didactique, la fonction herméneutique et la fonction de vulgarisation – une fonction heuristique, de modélisation du discours scientifique, qui se développerait autour d'isotopies analogiques fondatrices : c'est le cas par exemple de la conceptualisation moderne du système de la circulation du sang, qui reproduit la circulation de l'eau dans un système d'irrigation (Oliveira, 2009), ou bien des isotopies cinématographiques dans le domaine du génie génétique (Temmerman 2007).

Prandi (Prandi, Rossi 2012) perfectionne ensuite plus récemment la typologie des métaphores terminologiques sur la base de leur nature sémiotique, différenciant alors les simples catachrèses basées sur des analogies formelles (le *nain* des signaux ferroviaires, la *manche tulipe* d'une robe, l'*épingle à cheveux* dans notre ADN) des concepts métaphoriques partagés, qui peuvent générer des « essaims métaphoriques » (Prandi, 2012), tels que la métaphore organiciste dans la langue du droit ou la métaphore mécaniciste dans le domaine de la médecine. Un dernier cas de figure est enfin celui des métaphores créatrices de théories, qui ont la fonction de défier notre logique commune, projetant notre pensée vers de nouveaux paradigmes cognitifs – c'est par exemple le cas, cité par Montuschi (1993), de la sélection *naturelle* darwinienne, de la *main invisible* d'Adam Smith ou de la *spaghettification* évoquée par Hawking (Rossi, 2015).

Dans ces derniers exemples, le pouvoir heuristique de la métaphore révèle avec évidence que les termes métaphoriques sont bien plus puissants que de simples étiquettes catachrétiques sur base d'analogie référentielle, et que, par le biais de leur nature de « prédications allotopiques » (Cortès 2003), ils

peuvent contribuer à la formation, voire à la systématisation, de nouveaux domaines du savoir.

La démarche analogique peut alors jouer un rôle dans la création d'ontologies pour la structuration des concepts dans des domaines fortement thématisés : dans le cas de la circulation du sang (voir encore une fois Oliveira 2009), la métaphore sousjacente de la circulation de l'eau dans les canaux d'irrigation fournit une modélisation permettant ensuite de structurer tout le domaine ; de même, l'isotopie filmique (Temmerman, 2007) est à la base de la conceptualisation du génie génétique, ou la métaphore anthropomorphique (Rossi, 2009) est au cœur des ontologies dans les terminologies de dégustation du vin.

On ne saurait enfin oublier, dans l'étude des métaphores terminologiques et de leurs multiples réalisations, le poids de la culture d'appartenance dans la dynamique analogique, notamment dans le transfert des isotopies métaphoriques d'une langue/culture à une autre. Humbley (2005, 2012, 2018) met à plusieurs reprises en évidence l'impact des facteurs culturels dans la formation de « terminologies secondaires » par la voie de la traduction, impact qui se manifeste dans toute typologie d'expression métaphorique. Pour ne citer que quelques exemples, les catachrèses, quoique fondées sur des analogies formelles, changent sensiblement d'une langue à l'autre, le *nain* des chemins de fer français devenant une *marmotte* en italien et l'*unghia* du vin étant plutôt perçue dans la culture des dégustateurs français comme un *ménisque*. Cependant, l'empreinte culturelle affecte également les isotopies partagées : la métaphore cinématographique citée par Temmerman (2007) fut transposée très difficilement dans d'autre langues, étant profondément ancrée dans l'imaginaire collectif nord-américain ; de même, la métaphore zoomorphe dans le domaine financier anglo-américain des *bears and bulls* est intraduisible dans d'autres langues comme l'italien (qui a recours à l'emprunt intégral) ou le français (qui a recours au calque ou à la néologie, voir Rossi, 2018). Enfin, on peut se demander si les métaphores constitutives de nouvelles théories sont également sujettes à l'hypothèque de la culture d'appartenance (voir par exemple la *crêpe stellaire* dans le domaine de l'astrophysique, citée par Giaufret et Rossi 2010), et dans quelle mesure les systèmes conceptuels peuvent alors diverger ou converger.

2. Ontologies, terminologies et métaphores

Les trois termes qui forment le titre de ce paragraphe sont caractérisés par des relations épistémologiques complexes, parfois controversées, qui

témoignent en tout cas du lien étroit et durable qui les relie. Le terme *ontologie*, appartenant initialement au domaine de la philosophie, désigne au début l'étude de la nature des êtres et des choses au niveau de la pensée abstraite, universelle, au-delà de toute représentation linguistique et culturelle concrète. La notion philosophique d'ontologie présuppose une vision partagée, qui dépasse les manifestations visibles des entités et des concepts dans la réalité de la communication naturelle, visant l'attestation de la nature de l'existence des êtres et des processus, ainsi que les modalités de représentation de ces données (nous renvoyons pour un aperçu complet à Øhrstrøm, 2005, cité par Leonardi, 2012 : 2). C'est cette notion philosophique d'ontologie la pierre angulaire de toute tentative de classification taxonomique scientifique, à partir d'Aristote pour arriver à Linné : on y décèle l'idéal d'une science universelle, organisée selon une ontologie partagée.

L'objectif d'une ontologie commune, déclinée ensuite dans diverses terminologies standardisées parallèles en plusieurs langues, est également le noyau fondamental du « rêve » wustérien, dans la version originale de Théorie Générale de la Terminologie (TGT). Bien que l'ontologie décrite par Wüster ne recoupe pas parfaitement le concept philosophique d'ontologie (pour une histoire des rapports entre ontologies et terminologies dans le domaine français, nous renvoyons à Aulitto, 2014), le but fondamental est bien celui d'une analyse détaillée des êtres et des concepts formant l'ensemble des connaissances dans un domaine technique et scientifique, ainsi que l'étude de leurs relations, afin d'en élaborer une représentation globale univoque et partagée. Le principe de l'uninotionnalité énoncé par Wüster repose sur l'idée qu'une ontologie du domaine est bien possible, et que la terminologie doit être abordée selon une approche conceptuelle ou onomasiologique.

Plus récemment, avec l'avènement des approches terminologiques textuelles et culturelles – la terminologie textuelle de Bourigault et Slodzian (1999), la socioterminologie de Gaudin (2003), la terminologie culturelle de Diki-Kidiri (2000), la terminologie communicative de Cabré (1999) – le principe onomasiologique est progressivement mis en discussion par les terminologues et linguistes, ainsi que l'idéal d'une représentation ontologique formelle universelle ; les courants récents mettent plutôt l'accent sur la diversité linguistique et culturelle qui sous-tend non seulement les terminologies, mais également les concepts et notions que les terminologies désignent : chaque communauté de locuteurs ne partagerait que partiellement les concepts fondateurs d'un domaine, le reste étant découpé sur la base de facteurs liés au contexte de communication, tels que les variables pragmatiques, discursives,

culturelles. Le courant variationniste focalise l'attention sur la dimension sémasiologique, privilégiant l'étude des réalisations textuelles des concepts.

Ce n'est qu'à la fin du XX^{e} siècle, dans le sillage du tournant cognitif en linguistique, que les terminologues et linguistes redécouvrent le concept d'ontologie, concept qui avait d'ailleurs entre-temps subi une évolution sensible du domaine philosophique au domaine connexe et nouveau de l'ingénierie de la connaissance et des sciences de la communication. Loin d'être des conceptualisations abstraites, appartenant à la sphère philosophique, les nouvelles ontologies sont plutôt des outils d'organisation des connaissances, en mesure d'interagir avec des usagers humains aussi bien qu'avec des machines, permettant un découpage fonctionnel des domaines de spécialité professionnels. La terminologie sociocognitive de Temmerman (2000) remet au premier plan le concept comme « unité de connaissance », focalisant l'attention sur les procédés cognitifs qui sous-tendent les terminologies dans les discours professionnels ; quelques années après, l'équipe Condillac (entre autres, voir Roche 2007) conjugue les objectifs des terminologues avec les sciences des ingénieurs, les ontologies représentant dans ces approches une forme de gestion des connaissances, en mesure d'interagir avec des banques de données ou des *knowledge-bases* gérées aussi bien par des humains que par des machines, les concepts étant alors exprimés dans un langage formel et universellement compréhensible. Dans les deux approches, des mots-valises tels que *termontographie* (Temmerman, Kerremans, Vandervoort, 2005) ou *ontoterminologie* (Roche, 2007) représentent de façon emblématique le rapprochement des deux disciplines.

Les relations entre les noyaux conceptuels exprimés par les termes *métaphore* et *terminologie* sont bien plus difficiles à définir, en premier lieu pour des raisons épistémologiques et historiques. S'il est vrai, d'un côté, que les termes à base métaphorique sont omniprésents dans les vocabulaires techniques et scientifiques, de l'autre côté la TGT, par l'importance qu'elle accorde à la monosémie des termes, à leur nature de dénominations objectives et universellement partagées, ne considère les termes métaphoriques que dans leur fonction d'étiquettes catachrétiques apposées sur des concepts précis. La métaphore dans ce contexte n'a aucun pouvoir modélisateur, elle n'est qu'une stratégie néologique parmi les autres. Kocourek (1991) dans son essai consacré aux langues spécialisées, reconnaît à la métaphore un pouvoir d'évocation et d'immédiateté pour les spécialistes, mais il met en évidence le risque d'am-

biguïté généré par la figure, qui produit une terminologie imparfaite, à risque de polysémie et d'imprécision.

Cette hypothèque théorique est destinée à peser sur les termes métaphoriques jusqu'à l'avènement du paradigme cognitiviste en terminologie : ce n'est qu'en 1994 que l'article d'Assal sur la métaphore en terminologie remet au premier plan la valeur heuristique de la figure dans la création de nouvelles dénominations techniques :

> La métaphore terminologique est loin d'être une simple façon de parler, elle est essentiellement une façon de penser. Certes elle est un emprunt imagé, mais une fois que cet emprunt est réinvesti dans une pratique sociale, une fois que sa signification est réglée par les acteurs agissant dans le cadre de cette pratique, elle devient l'expression d'un nouveau concept (Assal, 1994, p. 26)

Les dernières décennies sont marquées par un foisonnement d'études sur la métaphore en terminologie, le plus souvent sous l'égide de l'approche des ICM (*idealised cognitive models*) élaborée en 1980 par Lakoff et Johnson. Les linguistes s'intéressent alors plutôt aux interactions conceptuelles (*mappings*) en mesure de façonner les domaines notionnels techniques et scientifiques : Temmerman (2000, 2007), Vandaele et Lubin (2004) proposent dans ce sillage des analyses approfondies des métaphores cognitives dans les sciences de la vie. D'autres études (Gaudin, 2003, Cortès, 2003, Oliveira, 2007 et 2009), dans la tradition de l'analyse du discours d'empreinte francophone, proposent plutôt une lecture des dynamiques textuelles et discursives concernant la métaphore dans les discours de spécialité, prenant en compte surtout les dimensions de la vulgarisation, de la communication et de la pragmatique. Parmi les approches les plus récentes, deux synthèses sont offertes par Rossi, 2015 et Humbley, 2018 (chapitre 7).

Enfin, les termes *métaphore* et *ontologie* sont indissolublement liés par un lien heuristique. La métaphore, en tant qu'interaction conceptuelle en mesure de générer de nouvelles visions du monde, peut jouer un rôle dans la définition des ontologies, des domaines du savoir ; elle reprend dans ce contexte philosophique tout son pouvoir de modélisation, d'orientation de la compréhension :

> [L]e rôle de la métaphore n'est pas de permettre une compréhension partielle d'un type d'expériences dans les termes d'un autre type d'expériences, c'est d'exprimer la perception de corrélations entre divers domaines d'expériences (Détrie, 2001 : 172)

C'est essentiellement au courant cognitiviste que revient le mérite d'avoir reconnu le pouvoir modélisateur des métaphores au niveau ontologique ; lorsque Temmerman par exemple (2007) analyse le domaine du génie génétique, elle révèle jusqu'à quel point la métaphore cognitive L'ADN EST UNE CARTE GEOGRAPHIQUE influence la définition des concepts dans le domaine, constituant le squelette de l'ontologie domaniale :

Domaine source : carte géographique	Domaine cible : carte génétique
continents	chromosomes
repérer des emplacements	repérer des gènes
marqueurs ou points de repère	RFLP en tant que marqueurs
position relative d'emplacements par rapport à d'autres (en km ou *miles*)	*carte de liaison* : position de chaque gène par rapport à un autre gène (distance génétique en centimorgan) *carte physique* : nombre de paires de bases entre deux gènes (distance entre deux repères exprimée en paires de bases)
région	région
site	site

Figure 1 - Temmerman 2007

Selon ces études, la projection conceptuelle de l'ontologie du domaine source sur celle du domaine cible contribuerait de façon conséquente à modéliser le domaine cible et les concepts qui le composent, comme dans une sorte de « photocopie » ontologique. Sur cette dernière question, justement, s'enchaîne notre problématique et notre questionnement : dans quelle mesure les isotopies métaphoriques qui sous-tendent un domaine spécialisé peuvent-elles façonner les concepts qui le constituent ? Existe-t-il une prévisibilité des ontologies fondées sur ce genre d'interaction conceptuelle ? Et enfin, est-ce que ces isotopies génèrent des ontologies comparables dans des langues/cultures différentes ?

3. Quelques exemples dans des domaines divers

3.1. La métaphore anthropomorphique dans le domaine de la dégustation du vin

Le domaine de la dégustation du vin représente un exemple intéressant pour l'analyse de la fonction modélisatrice des métaphores dans les termi-

nologies. En dépit de sa nature fortement imagée et lyrique, qui a pendant longtemps représenté un obstacle à la légitimation de ce langage comme terminologie à tous les effets (Amoraritei, 2002), le vocabulaire des dégustateurs est un vocabulaire technique, fondé sur des concepts partagés, dont les origines remontent à l'époque romaine et dont la normalisation scientifique date du XVIII siècle (Peynaud, Blouin, 2013). À la base de ce vocabulaire, on peut reconnaitre en filigrane une métaphore cognitive sous-jacente, que l'on pourrait classifier dans la catégorie des métaphores « ontologiques » selon la typologie de Lakoff et Johnson (1980), à savoir la métaphore anthropomorphique. En d'autres mots, de nombreux concepts appartenant à ce domaine reposent sur l'interaction conceptuelle exprimée par l'ICM suivant : LE VIN EST UNE PERSONNE. On parlera donc d'un vin *franc, puissant, charnu, costaud*... ou bien d'un vin *élégant, fin, féminin, rond*. Au-delà du pouvoir évocateur de cette métaphore cognitive, il nous semble intéressant de vérifier l'hypothèse suivante : est-ce que l'ontologie du domaine source (dans ce cas, l'être humain) peut être projetée sur le domaine cible (dans ce cas, le vin) afin de donner une structure ontologique précise au domaine cible ? et en deuxième lieu, est-ce que l'ontologie du domaine source (l'être humain) est transférée de façon homogène dans le domaine cible (le vin) dans des langues et cultures différentes ?

Commençons par la première question. L'ontologie du domaine [être humain] peut prendre des formes diverses ; nous en proposerons une représentation extrêmement simplifiée dans le schéma suivant :

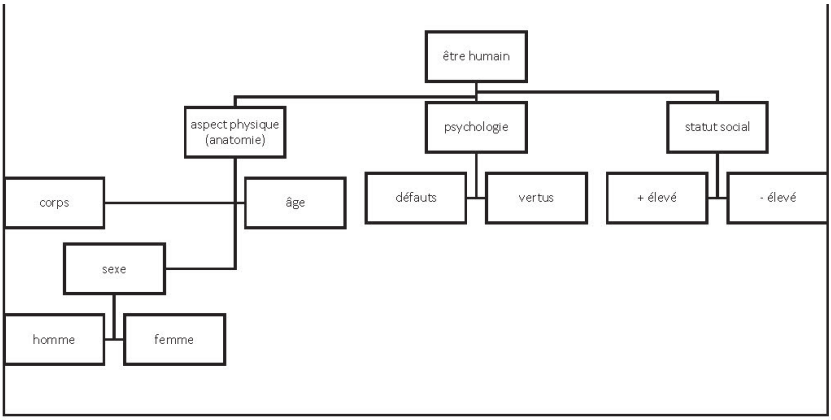


Figure 2 - ontologie du domaine source

Sur la base de cette représentation, un [être humain] se définit sur la base des critères physiques (sa taille, la couleur des yeux ou des cheveux, des particularités liées aux parties du corps, l'âge, le sexe…), psychologiques (ses vertus, telles que l'honnêteté, la générosité, ou bien ses défauts comme l'hypocrisie, l'égoïsme…) et sociales. Que se passe-t-il dans le processus d'interaction conceptuelle entre le domaine source et le domaine cible (le vin) ?

Le vocabulaire du vin semble avoir hérité la structure ontologique du domaine source, du moins dans ses grandes lignes : le langage des dégustateurs prévoit en effet des expressions métaphoriques désignant à la fois l'aspect physique, les caractéristiques psychologiques et même le statut social, comme il ressort des schémas suivant, où nous avons répertorié les expressions utilisées dans la pratique des dégustateurs professionnels (Coutier, 1994, Rossi, 2009).

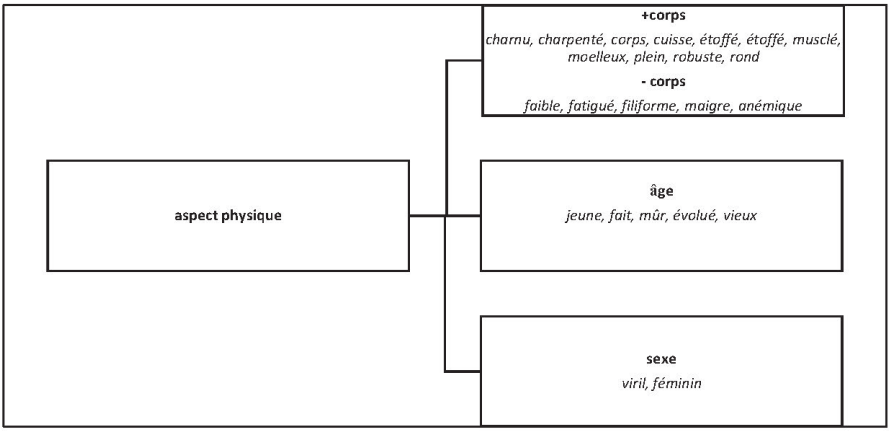


Figure 3 - ontologie du domaine cible - aspect physique

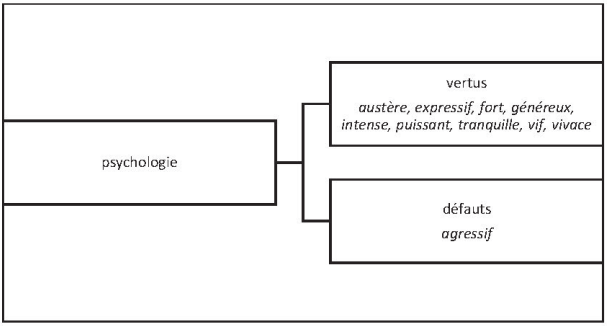


Figure 4 - ontologie du domaine cible - aspect psychologique

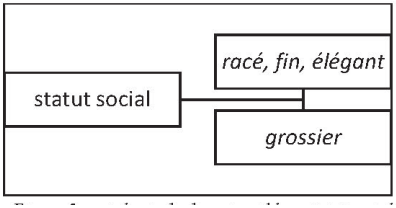


Figure 5 - ontologie du domaine cible – statut social

L'extrême richesse du langage du vin, ainsi que la stratification et la segmentation des publics utilisant le langage de la dégustation (Peynaud, Blouin, 2013, Gilardoni, 2009), débordent largement l'ensemble des termes cités ; toutefois, il nous semble évident que dans ce cas particulier, l'ICM LE VIN EST UNE PERSONNE s'avère du moins partiellement utile à la structuration de l'ontologie conceptuelle du vocabulaire de l'analyse organoleptique ; dans le cas d'éventuels néonymes, on peut imaginer qu'ils seraient créés à partir de la même isotopie conceptuelle.

De même, la productivité de l'ICM est attestée par sa récurrence – du moins partielle – dans diverses langues et cultures : en français, anglais, italien, les réalisations partielles de l'interaction métaphorique de fond permettent de structurer le domaine de l'analyse organoleptique de façon homogène, tout en manifestant parfois des variations dans les réalisations terminologiques ; ainsi, en italien le vin est le plus souvent qualifié par des adjectifs normalement référés au sexe masculin (*forte, potente, robusto,* etc.) alors qu'en français les termes de la dégustation renvoient également à un imaginaire féminin (un vin a de la *cuisse*, un vin est *rond* et sa couleur est appelée *robe*...) :

Corps +	Corps -
Carnoso IT / Charnu FR / Meaty-Fleshy ENG / **Corpo** IT / Corps /chair FR / Body ENG / **Pieno** IT / Plein FR / Full ENG / **Pingue** IT / Cuisse FR / Cuisse ENG / **Rotondo** IT / Rond FR / Round ENG / **Robusto** IT / Etoffé/musclé/robuste/charpenté/musclé FR / Robust ENG /	**Anemico** IT / Anémique FR / Anaemic ENG / **Debole** IT / Faible-chétif FR /Weak-feeble EN / **Fiacco** IT / Flasque/fatigué FR / Flaccid/weak EN / **Magro** IT / Maigre FR / Thin ENG / **Molle** IT / Mou FR /Flabby ENG /

Figure 6 – comparaison interlinguistique – domaine du corps du vin (Rossi, 2009)

Excès de caractère	Equilibre	Manque de caractère
Aggressivo IT / Agressif FR / Aggressive ENG / **Grossolano** IT / Grossier FR / Coarse ENG /	**Amabile** IT / Aimable FR / Agreeable/pleasant ENG / **Elegante**/fine IT / Elégant FR / Elegant ENG /	**Debole** IT / Faible-chétif FR /Weak-feeble EN / **Fiacco** IT / Flasque/fatigué FR / Flaccid/weak EN /

Figure 7 – comparaison interlinguistique – domaine du caractère du vin (Rossi, 2009)

3.2. La métaphore animale dans le domaine de la finance

S'il est vrai, comme le démontre le cas que nous avons analysé, que parfois la métaphore cognitive sous-jacente à un domaine peut s'avérer une clé d'accès efficace pour en élaborer une ontologie partagée, dans d'autres cas cette hypothèse est démentie par l'analyse de l'ontologie domaniale. Un exemple emblématique est constitué par le domaine de la finance (sur ce sujet, voir aussi Rossi, 2018). Dans ce domaine, les ICM dominants sont liés à la métaphore animale, dépendant à son tour de manière très étroite de l'imaginaire de la culture dominante dans le secteur financier, à savoir la culture nord-américaine. Pour cette raison, les termes à base métaphorique sont fondés sur des ICM du genre LES ACTEURS DE LA FINANCE SONT DES ANIMAUX, comme il ressort très clairement du schéma suivant, qui enregistre les termes anglais désignant les professionnels de la finance :

INVESTORS ARE ANIMALS	*lemming* […] an investor that follows the crowd into an investment that will inevitably end unsuccessfully. *ostrich* […] an investor who ignores important pieces of information, which have the ability to impact them or the market in which they operate. *pig* […] an investor who is often seen as greedy, having forgotten [his] original investment strategy to focus on securing unrealistic future gains *sheep* […] an investor who lacks a focused trading strategy and trades on the suggestions of others, including friends, family and financial advisors *shark* […] an investor that is hostile to the target firm's management and that is interested in taking over the firm

MARKETS ARE ANIMALS	*deer market* […] a flat market, characterised by low activity and investors' uncertaintly and unwilling-ness to buy or sell, as they are waiting for a sign of which way the market is going to end up moving *bull market* […] characterised by an upward trend in the price of securities *bear market* […]characterised by a downward trend – a market condition in which the prices of securities are fall- ing
COMPANIES ARE ANIMALS	*gorilla* […]a large company that dominates an industry *gazelle* a company [able to] to grow at an unusually high annual rate *lame duck* a company who has defaulted on its debts or has gone bankrupt due to the stock market *turkey* […] a start-up company that may subsequently go bankrupt. *elephant* […] an elephant is a large institution that has the funds to make high volume trades, thus having decisive influence on the price of the underlying financial asset

Figure 8 – les acteurs de la finance (adapté de Silaški, 2009)

Tout en représentant une métaphore fondatrice, l'interaction conceptuelle LE MONDE DE LA FINANCE EST UNE JUNGLE n'est pas pourtant une métaphore permettant d'élaborer par projection une ontologie efficace du domaine concerné : le transfert du domaine source sur le domaine cible s'avère trop faible et limité à certains aspects, tels que le comportement de l'animal (c'est le cas de *bear and bulls*, dont la métaphore est fondée sur les stratégies de lutte des deux animaux), sa taille (c'est le cas de *gorilla* ou d'*elephant*) ou d'autres traits saillants (comme dans le cas de *sheep* ou de *shark*). Cette limite dans l'interaction métaphorique entre domaine source et domaine cible se manifeste dans l'impossibilité de transférer les termes issus de la métaphore dans d'autres langues/cultures. Dans le passage de l'anglais vers le français et l'italien, par exemple, le terme *bears and bulls* perd son pouvoir d'évocation métaphorique : ancrée dans la culture nord-américaine, l'analogie devient complètement opaque pour un locuteur francophone ou italophone.

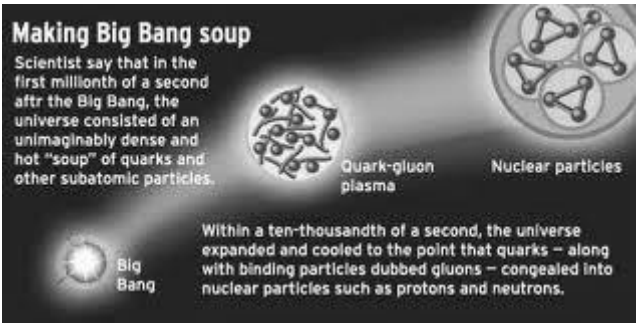
En français, la métaphore est éliminée au profit de termes explicatifs comme *marché haussier* et *marché baissier* (présents dans les banques de données Franceterme, Belterme, Termdat et GDT, voir Rossi, 2015), alors qu'en italien, l'emprunt intégral ne permet aux destinataires du message qu'un accès partiel à l'interaction conceptuelle déclenchée par la métaphore d'origine.

3.3. La métaphore gastronomique dans le domaine de l'astrophysique

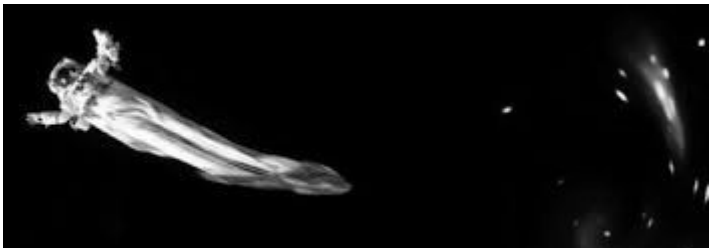
Un troisième cas de figure se manifeste enfin lorsque le *mapping* métaphorique, tout en étant constant dans un domaine donné, n'est pas exploité jusqu'au point de pouvoir constituer une isotopie métaphorique ontologique générant un essaim métaphorique à la base d'une ontologie spéculaire dans le domaine cible. Dans ce cas, l'isotopie reste limitée à des analogies ponctuelles, isolées, qui ne sont aucunement partie d'une ontologie de fond.

Un bon exemple de cette catégorie est fourni par l'interaction entre le domaine source de la gastronomie et le domaine cible de l'astrophysique. Au cours de son histoire, l'astrophysique a eu abondamment recours au domaine bien distant de la gastronomie afin de donner une forme plus accessible par voie d'analogie à des concepts abstraits et difficilement compréhensibles (Merakchi, Rogers, 2013). Ce *mapping* a généré de nombreux termes fondateurs de théories et de paradigmes scientifiques, parmi lesquels nous citerons à titre d'exemple les termes suivants :

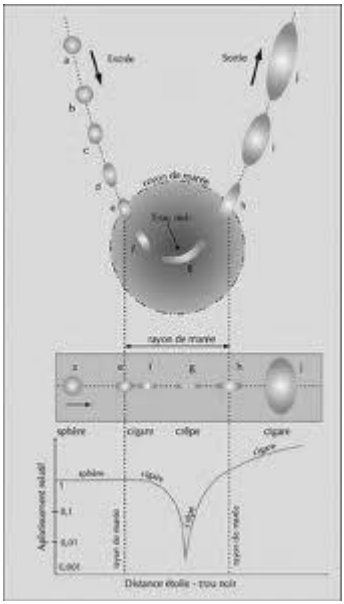
1. La *soupe de quarks (quark-gluon soupe)* : plasma de quarks et de gluons (QGP, de l'anglais quark-gluon plasma) ou *quagma* est un état de la chromodynamique quantique (QCD) qui existe à des températures et/ou des densités extrêmement élevées. Ce terme, créé par Bohr et Nielsen en 1977, est normalement utilisé comme variante de vulgarisation de *quark-gloun plasma* ;



2. La *spaghettification* : phénomène d'étirement vertical et de compression horizontale subi par les objets dans un fort champ gravitationnel, résultant de l'effet extrême des forces des marées d'un trou noir. Ce terme, proposé par le physicien anglais Calder en 1977 et popularisé par Hawking en 1998, est un terme technique à tous les effets, désignant un concept précis et univoque :



3. La *crêpe stellaire* : l'aplatissement d'une étoile comprimée par des forces de marée lors du passage près de l'horizon des événements d'un trou noir. Ce terme, élaboré par Luminet et Carter en 1982, n'est pas encore un terme pleinement attesté dans la communauté des spécialistes, mais son emploi est en train de s'imposer progressivement depuis 2004 :



Apparemment, un fil rouge relie ces termes métaphoriques ; ils sont tous issus d'un *mapping* conceptuel entre le domaine source de la gastronomie et le domaine cible de l'astrophysique.

Pourtant, cette métaphore cognitive diffère des cas analysés dans les pages précédentes : dans ce cas, l'interaction métaphorique reste cantonnée à des analogies formelles sous forme de dénominations catachrétiques épisodiques, et elle ne génère pas un essaim métaphorique qui pourrait être interprété comme une sorte d'ontologie spéculaire dans le domaine cible.

Dans les trois cas (la soupe de quarks, la spaghettification, la crêpe stellaire) la dénomination métaphorique joue sur un trait, sur un aspect du concept dans le domaine source, et profite de ce trait pour élaborer un terme simple, accessible. Aucune ontologie du domaine source n'est présente en filigrane dans le *mapping*, ce qui fait de ces termes des métaphores isolées, qui ne sont pas normalement productives. Dans ce cas, donc, la présence d'une série de termes métaphoriques produits d'un même *mapping* ne suffit pas pour activer une ontologie métaphorique qui serait l'image spéculaire de l'ontologie du domaine source. N'oublions pas, néanmoins, que l'interaction conceptuelle

pourrait être à tout moment réactivée et élargie pour formaliser un rapport ontologique entre le domaine source et le domaine cible.

4. En guise de conclusion

Sur la base des exemples qui précèdent, qui peuvent être considérés comme paradigmatiques du cas de figure qui nous intéresse dans ces pages, quelques considérations finales s'imposent. Nous essaierons de les exposer sous forme de réponses brèves aux questions posées au début de cet article :

1. **dans quelle mesure les isotopies métaphoriques qui sous-tendent un domaine spécialisé peuvent-elles façonner les concepts qui le constituent ?**

Les isotopies métaphoriques peuvent contribuer à façonner les ontologies des domaines spécialisés lorsque le *mapping* à la base de la métaphore inclut un transfert de l'ontologie du domaine source sur le domaine cible. C'est le cas de la métaphore anthropomorphique dans le langage de l'œnologie : l'ontologie de l'être humain interagit à fond avec le domaine cible, générant une charpente pour la définition des concepts, et par conséquent des termes, de spécialité.

2. **existe-t-il une prévisibilité des ontologies fondées sur ce genre d'interaction conceptuelle ?**

Dans les cas de transfert d'une ontologie du domaine source sur le domaine cible par voie métaphorique, l'ICM reste productif, en mesure d'engendrer de nouveaux termes, ce qui témoigne d'une certaine prévisibilité : si le domaine source est disponible, on peut imaginer que d'éventuels néonymes dans le domaine cible suivront les liens notionnels de l'ontologie du domaine source ;

3. **et enfin, est-ce que ces isotopies génèrent des ontologies comparables dans des langues/cultures différentes ?**

Lorsque l'ontologie du domaine source est partagée (dans le cas de l'interaction conceptuelle sur base anthropomorphique ou zoomorphique, par exemple), les ontologies et par conséquent les notions et les termes générés dans le domaine cible ne diffèrent pas sensiblement d'une langue/culture à l'autre. C'est le cas de la terminologie de la dégustation du vin, dont bien des termes métaphoriques sont parallèles dans diverses langues. Dans d'autres cas, si l'interaction entre l'ontologie du domaine source et celle du domaine cible est moins cohérente, et surtout plus étroitement liée à des concepts ou à des notions culturellement connotées, le transfert d'une langue à l'autre s'avère plus délicat et difficile. C'est le cas de la métaphore animale dans la

finance et, à plus forte raison, de la métaphore gastronomique dans le langage de l'astrophysique, dont les images, parfois brillantes et fortement évocatrices, perdent souvent leur pouvoir heuristique et tout leur… goût, dans le processus de transfert interlinguistique.

Références

Amoraritei, L. (2002), *La métaphore en Œnologie,* « Metaphorik.de » 2002, p. 1-16.

Assal A. (1994), *La métaphorisation terminologique*, « Terminologie et traduction » 2, 1994, p. 235-242.

Aulitto, S. (2014), *Terminologie e ontologie : studi francesi*, in Collesi, P., Serpente, A., Zanola, M.T., *Terminologie e ontologie. Definizioni e comunicazione tra norma e uso*, Milano, Educatt, p. 65-78.

Bachelard G. (1980), *La formation de l'esprit scientifique : contribution à une psychanalyse de la connaissance objective*, Paris.

Beccaria, G. L. (2008), *Per difesa e per amore. La lingua italiana oggi*, Garzanti, Milano. Black M. (1962), *Models and metaphors : Studies in language and philosophy*, Cornell University Press, Ithaca.

Bowdle B.F, Gentner, D. (2005), *The carreer of metaphor*, « Psychological Review » 112, p. 193-216.

Boyd R. (1979), *Metaphor ad Theory Change : What is "Metaphor" a Metaphor for ?*, in A. Ortony (éd.), *Metaphor and Thought,* Cambridge University Press, Cambridge, p. 356-408.

Cabré M. T. (1999), *La terminología. Representación y comunicación. Una teoría de base comunicativa y otros artículos*, Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra (Sèrie Monografies, 3), Barcelona.

Collombat I. (2003), *Le discours imagé en vulgarisation scientifique : étude comparée du français et de l'anglais*, "Metaphorik.de" 05/2003 (en ligne, www.metaphorik.de/05/collombat.pdf).

Collesi, P., Serpente, A., Zanola, M.T. (2014), *Terminologie e ontologie. Definizioni e comunicazione tra norma e uso*, Milano, Educatt.

Cortès C. (éd.) (2003), *La métaphore. Du discours général aux discours spécialisés*, Cahier du C.I.E.L. 2000-2003.

Desmet I. (2007), *Terminologie, culture et société. Éléments pour une théorie variationniste de la terminologie et des langues de spécialité*, « Les Cahiers du Rifal », 26, décembre 2007, en ligne, p. 3-13.

Détrie C. (2001), *Du Sens dans le processus métaphorique*, Champion, Paris.

Diki-Kidiri M. (éd.) (2000), *Terminologie et diversité culturelle*, « Les Cahiers du Rifal », 21, juin 2000, en ligne (http://www.rifal.org/cahiers/).
Diki-Kidiri M. (2007), *Éléments de terminologie culturelle*, « Les Cahiers du Rifal », 26, décembre 2007, en ligne, p. 14-25.
Dury P., Maniez F., Arlin N., Rougemont C. (éds.) (2009), *La métaphore dans les langues de spécialité*, Presses Universitaires de Grenoble, Grenoble.
Gardes Tamine J. (2007), *Les métaphores lexicalisées dans la langue et dans les langues de spécialité : un obstacle à la compréhension*, in M.-B. Vittoz (éd.), « Synergies Italie » 3/2007, p. 13-22.
Gaudin F. (2003), *Socioterminologie : une approche sociolinguistique de la terminologie,* De Boeck, Bruxelles.
Gaudin F. (2007), *Quelques mots sur la socioterminologie*, « Les Cahiers du Rifal », 26, décembre 2007, en ligne, p. 26-35.
Giaufret A, Rossi M. (2010), *L'année mondiale de l'astronomie : la diversité terminologique et culturelle de l'espace métaphorique*, Actes du Colloque GLAT 2010, Ed. du GLAT, Brest, p. 131-144.
Giaufret A, Rossi M. (2013), *Métaphores terminologiques, circulation des savoirs et contact entre langues,* « Signes, discours, sociétés » [en ligne], SDS_10, La métaphore dans les discours spécialisés, 2013, http://www.revue-signes.info/document.php?id=2928.
Gibbs R. W. Jr. (2008), *The Cambridge Handbook of Metaphor and Thought*, Berlin-New York, De Gruyter.
Gilardoni, S. (2009), *Strategie di marketing e strategie testuali : il caso del vino*, "mediAzioni" 7, http://mediazioni.sitlec.unibo.it
Humbley J. (2005), *La traduction des métaphores dans les langues de spécialité : le cas des virus informatiques*, « Linx » 52, 2005 (en ligne).
Humbley John (2018), *La néologie terminologique*, Limoges, Lambert-Lucas.
Jacobi D. (1986), *De la recherche à la vulgarisation ; itinéraires du texte scientifique*, Annales de l'Université de Besançon, les Belles Lettres, Paris.
Jacobi D. (1999), *La communication scientifique ; discours, figures, modèles*, Presses Universitaires de Grenoble (coll. Médias & sociétés), Grenoble.
Kocourek R. (1991), *La langue française de la technique et de la science. Vers une linguistique de la langue savante*, Oscar Brandstetter, Wiesbaden, 2e édition augmentée, refondue et mise à jour avec une nouvelle bibliographie, 1991.
Kuhn T. S. (1979), *Metaphor in science* in A. Ortony (éd.), *Metaphor and Thought.* Cambridge University Press, Cabridge, 1979, p. 409-419.
Lakoff G., Johnson M. (1980), *Metaphors we live by*, Chicago University Press, Chicago. Leonardi, N. (2012) 'Ontology' and Terminological

Frameworks : an Overview of Issues and Term(s), in « HERMES », 48 ; Aarhus, Department of Business Communication at Aarhus University, School of Business and Social Sciences, p. 19-33.

Merakchi K., Rogers, M. (2013), *The translation of 7culturally bound metaphors in the genre of popular science articles : A corpus-based case study from Scientific American translated into Arabic,* "Intercultural Pragmatics" 10(2) : 341-372.

Montuschi E. (1993), *Le metafore scientifiche*, Milano, Franco Angeli, 1993.

Musacchio M. T., Corbolante L. (2012), *When are science and technology (in) accessible ? A diachronic study of the poularisation of new terms in physics and information technology*, in AAVV., Terminologie : textes, discours et accès aux savoirs spécialisés, GLAT GENOVA 2012, éd. GLAT, Brest, p. 225-238.

Normand C. (1978), *Métaphore et concept*, Complexe, Bruxelles.

Oliveira I. (2005), *La Métaphore terminologique sous un angle cognitif*, « Meta, Journal des traducteurs », vol. 50, nº 4, 2005, p. 83-104.

Oliveira I. (2009), *Nature et fonctions de la métaphore en science. L'exemple de la cardiologie,* L'Harmattan, Paris, 2009.

Øhrstrøm, P., Andersen, J., Schärfe, H., *What Has Happened to Ontology*, "Conceptual Structures : Common Semantics for Sharing Knowledge ; Lecture Notes in Computer Science" Volume 3596, 2005, p. 425-438.

Peynaud, E., Blouin, J. (2013), *Le goût du vin,* Paris, Dunod.

Prandi M. (2012), *A Plea for Living Metaphors : Conflictual Metaphors and Metaphorical Swarms,* « Metaphor and Symbol » 27, 2, pp.148-170.

Prandi M., Rossi M. (2012), *Les métaphores dans la création de terminologie,* dans Terminologie : textes, discours et accès aux savoirs spécialisés, 2012, Ed. du GLAT, Brest, p. 7-19.

Resche C. (2002), *La métaphore en langue spécialisée, entre médiation et contradiction : étude d'une mutation métaphorique en anglais économique*, « ASp » 35-36, 2002, p. 103-119.

Richards, I. A. (1936), *The Philosophy of Rethoric*, Oxford University Press,

Roche, C. (2007) *Terme et concept : fondements pour une ontoterminologie*, Actes du Colloque TOTh 2007, p. 1-22.

Roche, C., Calberg-Challot ; M., Damas, L., Rouard, Ph., (2009), *Ontoterminology A New Paradigm for Terminology*, "KEOD", INSTICC Press, p. 321-326.

Rossi M. (2009), *L'emploi de la métaphore comme ressource pour la néologie terminologique : le cas du langage de la dégustation du vin*, in P. Dury

et al (éds.), « La métaphore en langues de spécialité », Lyon, PUL, 2009, p. 199-227.

Rossi M. (2014), *Métaphores terminologiques : fonctions et statut dans les langues de spécialité,* SHS Web of Conferences, Volume 8, 4e Congrès Mondial de Linguistique Française, p. 713-724.

Rossi M. (2019), « Métaphores et discours experts : conflit et cohérence à l'épreuve du transfert interlinguistique », *Langue française*¸ 204, 2019, pp. 71-86.

Rossi M. (2018), « Métaphores et néologie : quelques réflexions sur la francisation des termes de l'économie et de la finance », *Analele Universitatii din Craiova*, Seria Stiinte filologice Langues et littératures romanes, 1, 2018, pp. 93-110.

Rossi M. (2015), *In rure alieno. Métaphores et termes nomades dans les langues de spécialité*, Berne, Peter Lang.

Stengers I., Schlanger J. (1991), *Les concepts scientifiques,* Gallimard, Paris, 1991. Schlanger J. (1995), *Les métaphores de l'organisme*, L'Harmattan, Paris, 1995.

Silaški, N. A. (2009), *Animal metaphors in some business-related terms in English*, Paper presented at the International Conference *Across Languages and Cultures*, Herceg Novi, Montenegro, 4-6 June 2009.

Temmerman R. (2000), *Toward New Ways of Terminology Description. The Sociocognitive Approach*, John Benjamin, Amsterdam/Philiadelphie.

Temmerman, R., Kerremans, K., Vandervoort, V. (2005). *La termontographie en contexte(s)*, in : Blampain, D., Thoiron, P., Van Campenhoudt, M. (eds.), *Mots, Termes et Contextes. Actes des septièmes Journées scientifiques du réseau de chercheurs du réseau Lexicologie, Terminologie et Traduction,* Editions des Archives Contemporaines, p. 429-439.

Temmerman R. (2007), *Les métaphores dans les sciences de la vie et le situé socioculturel*,

« Les Cahiers du Rifal », 26, décembre 2007, en ligne, p. 72-81.

Vandaele S. (2004), *Approche cognitive de la traduction dans les langues de spécialité : vers une systématisation de la description de la conceptualisation métaphorique,* « Meta : journal des traducteurs / Meta : Translators' Journal », Volume 50, numéro 2, p. 415-431.

Vandaele S., Lubin L. (2005), *Approche cognitive de la traduction dans les langues de spécialité : vers une systématisation de la description de la conceptualisation métaphorique*, « Meta, Journal des traducteurs » vol. 50, n. 2, 2005, p. 415-431.

Summary

This paper aims at describing the various and different types of terminological metaphors, their semiotic and pragmatic aspects, from a comparative point of view. We will take into account different domains (from oenology to astrophysics) and different languages (French and Italian), in order to highlight the similarities and differences among different discourse communities and cultures.

Méthodologie d'ingénierie des connaissances pour la représentation des définitions lexicographiques dans le cadre de la théorie Sens-Texte

Maxime Lefrançois[*], Fabien Gandon[**], Alain Giboin[**]

[*] Univ. Nice Sophia Antipolis, CNRS, I3S, UMR 7271, 06900 Sophia Antipolis, France
maxime.lefrancois.86@gmail.com
[**] Inria
[fabien.gandon, alain.giboin]@inria.fr

Résumé. La Théorie Sens-Texte (TST) est l'une des théories linguistiques les plus précises d'un point de vue descriptif. En représenter les connaissances est un défi pour des applications de lexicographie, mais également pour injecter des connaissances linguistiques fines dans les systèmes de TALN. Cet article décrit la méthodologie et les principaux résultats de notre recherche qui consiste à mener une étude d'Ingénierie des Connaissances pour modéliser les connaissances lexicales sémantiques de la TST et en formaliser la sémantique. Nous nous intéressons plus particulièrement aux prédicats linguistiques, aux représentations linguistiques, et aux définitions lexicographiques. Nous nous sommes inspirés de la méthodologie de Bachimont, et avons opéré en trois étapes principales : (i) extension de la conceptualisation du domaine, (ii) recherche d'un formalisme de représentation des connaissances adapté, (iii) opérationalisation du formalisme pour le raisonnement logique et pour le partage et l'interrogation des connaissances sur le web des données.

1. Introduction

En tant que théorie linguistique hautement précise du point de vue descriptif, la Théorie Sens-Texte (TST) tient une place centrale dans le champ de la sémantique lexicale (Mel'čuk *et al.*, 1995 ; Kahane, 2003). La richesse de description des fonctions lexicales et des définitions lexicographiques du

Dictionnaire Explicatif et Combinatoire (DEC) fait sa force, mais elle a aussi des limites : (i) l'absence de représentation formelle et d'axiomatisation ; (ii) la faible couverture de la langue ; et (iii) le peu d'applications en TALN.

Lorsque Mel'čuk *et al.* (1995, p.208) promeut l'informatisation du DEC dans le but de « *"faire tourner" un appareillage logico-déductif sur cette description* ». Il met notamment en évidence l'émergence de besoins spécifiques :

- vérifier la validité du lexique ;
- découvrir de nouvelles connaissances sur la langue ;
- raisonner sur les textes.

L'émergence de tels besoins existe en réalité lors de la production de connaissances par tout domaine de spécialité. L'Ingénierie des Connaissances (IC) vise justement à répondre à ces besoins. Il s'agit de permettre par exemple de représenter, de manipuler, d'échanger, d'interroger, ou de raisonner avec les connaissances. Les besoins envisagés par Mel'čuk seraient donc couverts.

En parallèle aux travaux de formalisation déjà menés sur le DEC, nous nous sommes donc proposés de représenter les prédicats linguistiques, les représentations linguistiques, et les définitions lexicographiques au sens de l'IC, en restant au plus proche des standards de la Représentation des Connaissances (RC).

Dans cet article nous rapportons la méthodologie et les principaux résultats de notre recherche en ingénierie des connaissances appliquée aux connaissances lexicales sémantiques du DEC. La section 2 introduit la notion de définition lexicographique, à laquelle nous nous intéressons plus particulièrement. Nous détaillons ensuite en section 3 notre méthodologie, qui s'inspire de celle proposée par Bachimont (2000).

2. Les Définitions Lexicographiques dans le cadre de la TST

La définition lexicographique d'une unité lexicale L présente de façon formelle le sens dénotationnel de L. Dans le DEC, les définitions lexicographiques sont analytiques, c'est-à-dire de type genre prochain et différences spécifiques. Par exemple, nous pouvons trouver à l'entrée $\text{PEIGNE}_{\mathbf{B.2.D}}$ (le peigne du tisserand) la définition :

$$peigne_{B.2.d}\ de\ personne\ X\ pour\ objet\ Y =$$

(Outil de tissage qu'une personne X utilise pour démêler$_2$ les fibres d'un objet Y)

Le membre gauche de l'égalité représente le défini, $\text{PEIGNE}_{\text{B.2.D}}$, sous sa forme propositionnelle : une expression élémentaire dans laquelle apparaissent $\text{PEIGNE}_{\text{B.2.D}}$ et sa *structure actancielle sémantique*, définie selon la théorie des Actants Sémantiques (Mel'čuk, 2004). Le membre droit de l'égalité représente le définissant. Le genre prochain de ‹$\text{peigne}_{\text{B.2.D}}$› est ‹outil›, et le reste de la définition précise les différences spécifiques de ‹$\text{peigne}_{\text{B.2.D}}$› par rapport à ‹outil›. La structuration des composantes sémantiques de la définition est encore discutée par les chercheurs en lexicologie explicative et combinatoire aujourd'hui (Iordanskaja et Mel'čuk, 1990 ; Mel'čuk *et al.*, 1995 ; Mel'čuk et Wanner, 2001 ; Altman et Polguère, 2003 ; Barque, 2008 ; Barque et Polguere, 2009).

Cependant, il est nécessaire d'aboutir à une structuration aussi formelle que possible des définitions lexicographiques pour les applications en TALN, comme le justifie Wanner (2003). C'est pourquoi nous nous intéressons à la conceptualisation originellement privilégiée dans la TST, qui présente les définitions lexicographiques sous la forme d'une représentation sémantique qui encode le sens de l'unité lexicale définie (Mel'čuk *et al.*, 1995). Par exemple la figure 1 représente la définition de $\text{PEIGNE}_{\text{B.2.D}}$. Le nœud étiqueté ‹outil› est marqué comme communicativement dominant, il s'agit du genre prochain de $\text{PEIGNE}_{\text{B.2.D.}}$ D'autres nœuds sont marqués X, Y,… et représentent la structure actancielle sémantique de $\text{PEIGNE}_{\text{B.2.D.}}$

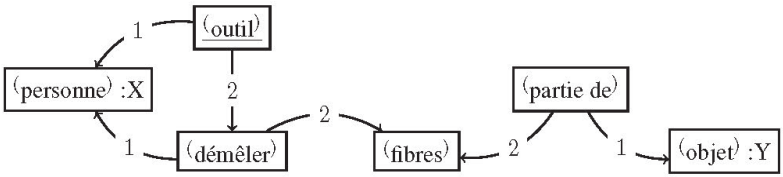


Figure 1 – *Représentation de la définition lexicographique de* $\text{PEIGNE}_{\text{B.2.D}}$ *sous la forme d'une représentation sémantique*

Jusqu'à présent, et de par la difficulté pratique de rédiger – à la main – un ensemble cohérent de telles représentations (Mel'čuk *et al.*, 1995, p. 75), ce type de représentation n'est pas utilisé dans les projets d'informatisation du DEC (Altman et Polguère, 2003 ; Apresian *et al.*, 2003 ; Wanner, 2009 ; Lux-Pogodalla et Polguère, 2011). Néanmoins, lorsque Polguère (1990) précise le

système de règles dans la composante sémantique du modèle Sens-Texte dans une perspective de génération automatique de textes, il justifie l'importance de représenter les définitions lexicographiques sous cette forme. On peut en effet en déduire deux règles de paraphrase qui devraient permettre de transformer une représentation sémantique en une autre équivalent, en remplaçant une unité sémantique par sa paraphrase, et vice-versa. Ces opérations élémentaires au niveau sémantique sont fondamentales pour faciliter le choix d'une lexicalisation appropriée pour un ensemble de représentations sémantiques équivalents.

3. La méthodologie

Lorsque Mel'čuk *et al.* (1995, p. 208) promeut l'informatisation du DEC, il met en évidence l'émergence de besoins spécifiques, tels que la recherche intelligente d'information, la mise à jour et la vérification automatique, ou encore la déduction logique. L'émergence de tels besoins existe en réalité lors de la production de connaissances par tout domaine de spécialité. L'ingénierie des connaissances se propose justement de répondre à ces besoins. Il s'agit de permettre par exemple de représenter, de manipuler, d'échanger, d'interroger, ou de raisonner avec les connaissances.

La tâche de l'ingénierie des connaissances, telle que Bachimont (2000) la définit, est de modéliser formellement un problème pour lequel les seules connaissances dont on dispose sont de nature linguistiques ou cognitives. La sortie de ce processus de modélisation est un formalisme de représentation de connaissances, qui peut alors être mis en œuvre par les machines. Nous nous sommes donc inspirés de la méthodologie en trois étapes proposée par Bachimont (2000), pour passer de l'expression linguistique des connaissances vers une représentation formelle et calculable.

3.1. Extension de la conceptualisation

La première étape consiste, à partir des connaissances de TST, à définir les primitives du domaine. Bien que la théorie des actants sémantiques et les définitions lexicographiques soient déjà hautement conceptualisés, nous avons identifié certaines limitations pour une formalisation ultérieure.

Conceptualisation des prédicats linguistiques. Pour une unité lexicale L, Mel'čuk (2004) distingue L dans la langue (i.e., dans le dictionnaire), ou dans les textes (i.e., dans les énoncés). Pour désactiver cette ambiguïté, nous nommons *type d'unité lexicale* l'unité lexicale décrite dans le dictionnaire, et

instance d'unité lexicale ou simplement *unité lexicale*, l'unité lexicale décrite dans les textes. De plus, nous associons à chaque type d'unité sémantique *de surface* une structure actancielle qui correspond à la structure actancielle sémantique de son type de lexie associé.

Nous avons montré (Lefrançois et Gandon, 2013c) que la hiérarchisation des prédicats sémantiques, avec héritage et spécialisation de leur structure actancielle, ne peut correspondre à une hiérarchisation des sens. Notons qu'en nous plaçant dans une démarche d'IC, nous ne souhaitions en aucun cas modifier la conceptualisation existante, mais seulement l'étendre. Nous avons donc introduit un nouveau niveau dit *sémantique profond* pour représenter les sens, et y avons défini les notions d'unité sémantique profonde et de leur structure actancielle de sorte qu'ils puissent être organisés en une hiérarchie qui satisfait les propriétés suivantes :

1. la hiérarchie des types d'unité sémantique profonde représente une hiérarchisation des sens ;
2. la structure actancielle des types d'unité sémantique profonde est héritée et éventuellement spécialisée au sein de la hiérarchie ;
3. il existe une correspondance simple entre la structure actancielle des types d'unité sémantique profonde, et celle des types d'unité sémantique de surface.

Finalement, nous proposons une visualisation graphique inspirée d'UML, comme l'illustre la figure 2. Les rectangles du haut représentent un type de lexie, un type d'unité sémantique de surface doté de sa structure actancielle, et un type d'unité sémantique profonde doté de sa structure actancielle. Les rectangles du bas représentent des instances de ces types.

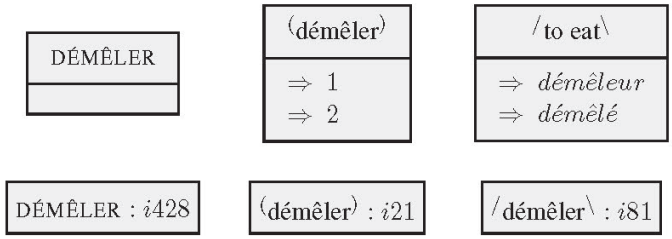


Figure 2 – *Visualisation des types et instances d'unité linguistiques.*

Conceptualisation des définitions lexicographiques. Une représentation d'une définition en langue naturelle, aussi formalisée soit-elle, reste une linéarisation qui encode le sens de la lexie définie. D'un autre côté, une définition lexicographique représentée sous la forme d'une représentation sémantique met en scène des (instances d') unités sémantiques. Il s'agit donc d'une instanciation prototypique de la définition d'un lexie L. Nous avons donc proposé une nouvelle conceptualisation pour les définitions lexicographiques :

- au niveau sémantique profond,
- et au niveau du dictionnaire.

La figure 3 représente la reconceptualisation de la définition de la lexie PEIGNE$_{\mathbf{B.2.D}}$.

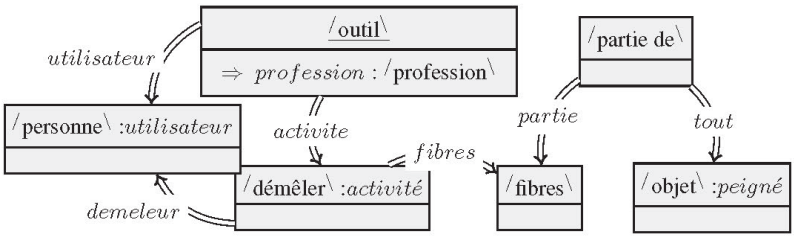


Figure 3 – *Repositionnement de la définition lexicographique de* PEIGNE$_{\mathbf{B.2.D}}$ *au niveau sémantique profond et au niveau du dictionnaire.*

3.2. Développement d'un prototype d'éditeur de définitions lexicographiques pour assister la conceptualisation

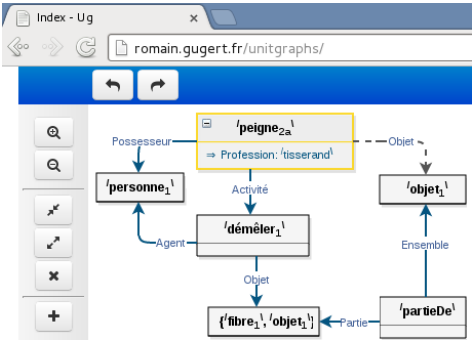
Nous avons proposé un workflow d'édition d'une définition lexicographique selon cette conceptualisation étendue (Lefrançois *et al.*, 2013), et développé un prototype d'éditeur qui implémente ce workflow (Gugert, 2013 ; Lefrançois *et al.*, 2014). Une démonstration est disponible à l'adresse http://wimmics.inria.fr/doc/video/ UnitGraphs/editor1.html. La figure 4 représente les deux premières étapes du processus d'édition d'une définition lexicographique dans ce prototype :

1. choix du genre prochain et de la structure actancielle sémantique profonde ;

2. construction de la définition lexicographique formelle par manipulation directe de graphes.

Nous avons évalué l'acceptation du processus et du prototype par la communauté de la TST. Nous avons ainsi pu confirmer les attentes des lexicographes de la TST pour un tel outil pratique, et identifier différentes directions d'amélioration pour notre prototype. Finalement, ce prototype offre une première base essentielle sur laquelle de futures discussions peuvent s'appuyer, et il permettra, nous l'espérons, l'intégration future de nos travaux dans des logiciels de lexicographie.

(A) Phase 1 : Sélection de la structure actancielle.



(B) Phase 2 : Édition de la définition formelle par manipulation de graphes.

Figure 4 – *Captures d'écran du prototype lors de l'édition de la définition de* PEIGNE$_{\text{B.2.D.}}$

3.3. Élaboration d'un formalisme de représentation des connaissances adapté

L'extension de la conceptualisation ainsi proposée peut être vue comme un objet frontière de type forme standardisée, au sens de Star et Griesemer (1989), entre la TST et le domaine de la représentation des connaissances.

Une fois la conceptualisation de la TST étendue, nous sommes passés à l'étape de formalisation des connaissances. Il s'agit de concevoir une ontologie du domaine, sa conception étant intimement liée au formalisme de représentation des connaissances que l'on choisit. Afin d'optimiser les possibilités de raisonnement dans le formalisme, notre étude s'est limitée à l'utilisation de fragments décidables de la logique du premier ordre. Ainsi, au premier abord, deux formalismes de représentation des connaissances existants semblaient adaptés pour la représentation des définitions lexicographiques :

- les logiques de description, qui sont intégrées aux standards de facto de la représentation des connaissances : les formalismes du Web Sémantique ;
- le formalisme des Graphes Conceptuels, qui avait originellement pour but la représentation des connaissances linguistiques à des fins de traitement automatisé, et qui fut inspiré des travaux de Tesnière (1959) à l'instar de la TST.

Montrons que ces deux formalismes ne sont pas adaptés pour représenter les connaissances de la TST à la lumière de deux des cinq critères de conception d'une ontologie proposés par Gruber (1995) : la clarté et la cohérence.

La clarté La définition d'un concept doit faire passer le sens voulu du terme, de manière aussi objective que possible. Dans la lignée de (Lefrançois et Gandon, 2011), nous avons montré que certains graphes simples peuvent être projetés sur la définition d'une seule classe en logique de description. Cependant, le manque de clarté de l'ensemble abscons d'axiomes ainsi obtenu justifie l'abandon de cette direction de recherche pour notre étude. Nous lui avons préféré dans un premier temps le formalisme des Graphes Conceptuels, qui propose une représentation des définitions de type de concepts et de relations très similaire aux définitions lexicographiques (Leclère, 1998).

La cohérence Rien qui ne puisse être inféré de l'ontologie ne doit entrer en contradiction avec les définitions des concepts. Bien qu'on puisse a priori représenter une définition lexicographique par un couple définition de type de concept, définition de type de relation dans le formalisme des GC, nous pouvons montrer que certaines inférences seraient incohérentes.

Nous avons donc décidé de modifier les bases du formalisme des Graphes Conceptuels, afin d'en dériver un nouveau formalisme, nommé formalisme des Graphes d'Unités (Lefrançois et Gandon, 2013a). La base de ce formalisme est une hiérarchie des types d'unité notée : l'ensemble minimal d'objets mathématiques nécessaires à la formation d'un ensemble préordonné de types d'unités munis de structures actancielles obligatoires, optionnelles, ou interdites. Un Graphe d'Unité est alors une combinaison de nœuds d'unité interconnectés.

3.4. Opérationnalisation de la formalisation sur le Web des données

La dernière étape de la méthodologie de l'ingénierie des connaissances est d'opérationnaliser le formalisme. Il s'agit de déterminer dans quelle mesure le formalisme choisi peut être utilisé pour les scénarios d'utilisation effectivement envisagés.

Nous avons tout d'abord étudié la déduction dans le formalisme des Graphes d'Unités. Nous lui avons attribué une sémantique logique (Lefrançois et Gandon, 2013b, d), et avons étudié les conditions de décidabilité du problème de conséquence logique. Notons que ce problème est décidable sous des conditions d'acyclicité que les lexicographes de la TST ont l'habitude de s'imposer dans l'élaboration du dictionnaire (Lefrançois, 2014).

C'est seulement après ce travail que nous avons choisi une modélisation du formalisme des Graphes d'Unités à l'aide des formalismes du Web Sémantique, afin de profiter des architectures existantes pour le partage, l'interopérationnalisation, et l'interrogation des connaissances sur le Web des données (Lefrançois, 2014). Nous utilisons le métamodèle OWL 2 RL, et nous étendons la base de règles OWL 2 RL/RDF par un ensemble de règles SPARQL pour capturer la sémantique de notre formalisme. Nous utilisons deux modèles hétérogènes :

- le modèle **gu-langue** adopte le point de vue de la compétence, et décrit les types d'unité au niveau assertionnel. Il s'agit du point de vue adopté par les ontologies lexicales du Web des données.
- Le modèle **gu-usage**, au contraire, adopte le point de vue de la performance. Les types d'unité sont décrits au niveau terminologique, laissant le niveau assertionnel aux graphes d'unités.

La figure 5 illustre ces deux modèles. Notons que ces modèles hérérogènes sont rendus interopérables par un ensemble de règles de transfert, qui définissent un morphisme d'ontologies au sens de Flouris *et al.* (2008).

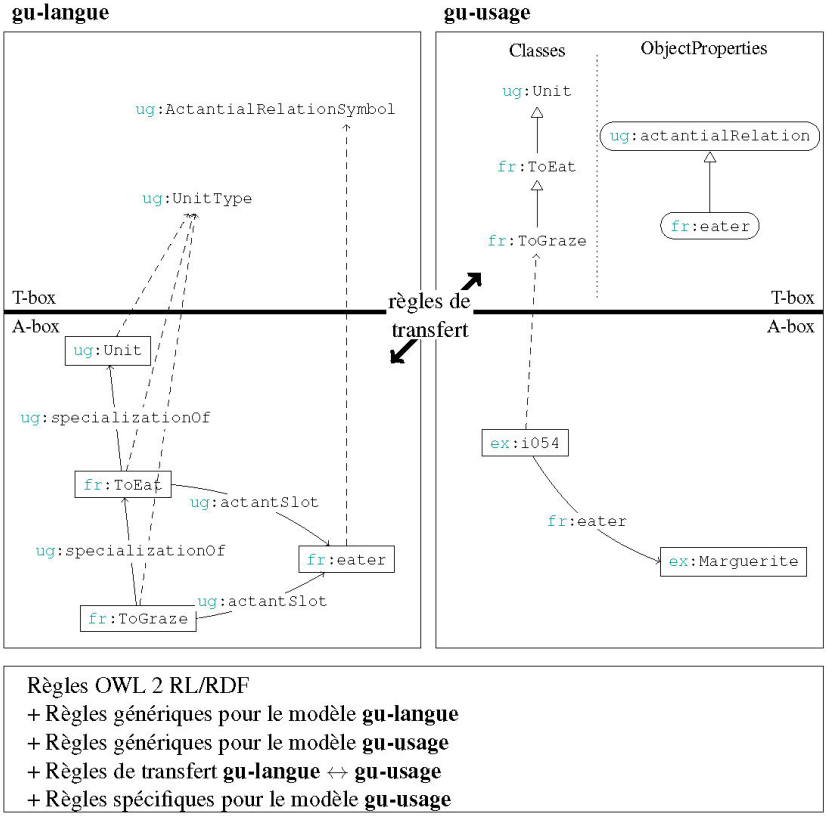


Figure 5 – *Architecture globale du modèle.*

La sémantique du formalisme des Graphes d'Unités est alors capturée par un ensemble de règles SPARQL :

- les règles SPARQL OWL 2 RL/RDF écrites en SPARQL, accessibles à l'URL http://ns.inria.org/ug/v1/owl2rl.rul ;
- des règles complémentaires génériques pour le modèle **gu-langue**, accessibles à l'URL http://ns.inria.org/ug/v1/ug-language.rul ;

- des règles complémentaires génériques pour le modèle **gu-usage**, accessibles à l'URL http://ns.inria.org/ug/v1/gu-usage.rul ;
- des règles de transfert entre les deux modèles, accessibles à l'URL http://ns.inria.org/ug/v1/ug-transfer.rul ;
- les règles spécifiques d'expansion et de contraction des types d'unité formellement définis. Notons que nous avons proposé un modèle des définitions formelles, et des transformations SPARQL TEMPLATE pour générer des règles SPARQL d'expansion et de contraction. Par exemple la représentation de la définition lexicographique de $\text{PEIGNE}_{\text{B.2.D}}$ est illustrée sur la figure 6, et la sortie du processus de transformation est illustrée sur la figure 7.

```
ex:peigneB2dug:definition_:outil.

_:outil{

    _:outilrdfs:label »outil»;
        aug:Participant;
        aex:Outil.

    _:personnerdfs:label »personne»;
        aug:Participant;
        aex:Personne;
        aug:obligatoryParticipant;
        ug:markerex:utilisateur.

    _:demelerrdfs:label »demeler»;
        aug:Participant;
        aex:Demeler;
        aug:obligatoryParticipant;
        ug:markerex:activite.

    _:fibresrdfs:label »fibres»;
        aug:Participant;
        aex:Fibres.

    _:partieDerdfs:label »partieDe»;
        aug:Participant;
        aex:PartieDe.

    _:objetrdfs:label »objet»;
        aug:Participant;
        aex:Objet;
        aug:obligatoryParticipant;
        ug:markerex:peigne.

    _:outilex:utilisateur_:personne
        .
    _:outilex:activite_:demeler.
    _:demelerex:demeleur_:personne.
    _:demelerex:fibres_:fibres.
    _:partieDeex:partie_:fibres.
    _:partieDeex:tout_:objet.
}
```

Figure 6 – *Représentation RDF de la définition lexicographique de* $\text{PEIGNE}_{\text{B.2.D.}}$

Règle d'expansion **def+**

```
# expansion rule
CONSTRUCT {
  ?u a ex:Outil .
  ?personne a ex:Personne .
  ?activite a ex:Demeler .
  ?fibres a ex:Fibres .
  ?partieDe a ex:PartieDe .
  ?peigne ex:Objet .
  ?u ex:utilisateur ?personne .
  ?u ex:activite ?activite .
  ?activite ex:demeleur ?personne .
  ?activite ex:fibres ?fibres .
  ?partieDe ex:partie ?fibres .
  ?partieDe ex:tout ?peigne .
}
WHERE {
  ?u a ex:peigneB2d .
  ?u ex:utilisateur ?personne .
  ?u ex:activite ?activite .
  ?u ex:peigne ?peigne .
  FILTER NOT EXISTS {
    ?u a ex:Outil .
    ?personne a ex:Personne .
    ?activite a ex:Demeler .
    ?fibres a ex:Fibres .
    ?partieDe a ex:PartieDe .
    ?peigne ex:Objet .
    ?u ex:utilisateur ?personne .
    ?u ex:activite ?activite .
    ?activite ex:demeleur ?personne .
    ?activite ex:fibres ?fibres .
    ?partieDe ex:partie ?fibres .
    ?partieDe ex:tout ?peigne .
  }
}
```

Règle de contraction **def-**

```
# contraction rule
CONSTRUCT {
  ?u a ex:peigneB2d .
  ?u ex:utilisateur ?personne .
  ?u ex:activite ?activite .
  ?u ex:peigne ?peigne .
}
WHERE {
  ?u a ex:Outil .
  ?personne a ex:Personne .
  ?activite a ex:Demeler .
  ?fibres a ex:Fibres .
  ?partieDe a ex:PartieDe .
  ?peigne ex:Objet .
  ?u ex:utilisateur ?personne .
  ?u ex:activite ?activite .
  ?activite ex:demeleur ?personne .
  ?activite ex:fibres ?fibres .
  ?partieDe ex:partie ?fibres .
  ?partieDe ex:tout ?peigne .
  FILTER NOT EXISTS {
    ?u a ex:peigneB2d .
    ?u ex:utilisateur ?personne .
    ?u ex:activite ?activite .
    ?u ex:peigne ?peigne .
  }
}
```

Figure 7 – *Règles d'expansion et de contraction générées automatiquement à partir de la représentation RDF de la définition lexicographique de* PEIGNEB.2.D.

Nous avons proposé une spécialisation du modèle des Graphes d'Unités pour la TST, que nous avons alignée avec le modèle **ontolex** proposé par le groupe communautaire Ontology-Lexica du W3C[1]. Nous espérons retirer de cette approche :

- une adoption facilitée de nos travaux par la communauté de la représentation des connaissances lexicales ;
- la possibilité de réutiliser des données existantes pour peupler notre formalisme.

Nos résultats sont publié en ligne à l'adresse http://ns.inria.fr/ug/v1#.

4. Conclusion

Ainsi, afin de rendre possible la représentation des définitions lexicographiques sous la forme d'une représentation sémantique dans un projet d'informatisation du Dictionnaire Explicatif et Combinatoire, nous avons appliqué une méthodologie en trois étapes :

- Nous avons proposé une extension de la conceptualisation de la théorie Sens-Texte, qui joue le rôle d'un objet frontière avec le domaine de l'ingénierie des connaissances ;
- Nous avons choisi un formalisme de représentation des connaissances adapté. Nous avons montré que les logiques de description et le formalisme des Graphes Conceptuels ne sont pas adaptés, et nous avons donc construit le nouveau formalisme de représentation des connaissances des Graphes d'Unités ;
- Nous avons étudié dans quelle mesure ce formalisme peut être opérationnalisé pour répondre aux besoins des linguistes. D'abord en étudiant la déduction logique et les conditions de décidabilité du formalisme, puis en proposant une modélisation avec les formalismes du Web Sémantique, afin de profiter des architectures existantes pour le partage, l'interopérationnalisation, et l'interrogation des connaissances sur le Web des données.

Ce travail offre de nombreuses perspectives : (i) en lexicographie Sens-Texte, (ii) en TALN où le besoin de formaliser les définitions lexicographiques a été noté par Wanner (2003), et (iii) en ingénierie des connaissances,

1 Ontolex - Ontology-lexical Community group - http://www.w3.org/community/ontolex.

car le formalisme des Graphes d'Unités pourrait apparaître le plus adapté pour représenter les connaissances d'un autre domaine de spécialité que la linguistique.

Références

Altman, J. et Polguère, A. (2003). La BDéf : base de définitions dérivée du Dictionnaire explicatif et combinatoire. *In Proceedings of the First International Conference on Meaning-Text Theory (MTT'2003)*, p. 43-54, Paris, France.

Apresian, J., Boguslavsky, I., Iomdin, L., Lazursky, A., Sannikov, V., Sizov, V. et Tsinman, L. (2003). ETAP-3 linguistic processor : a full-fledged NLP implementation of the MTT. *Proceedings of the First International Conference on Meaning–Text Theory (MTT'2003)*, p. 279-288.

Bachimont, B. (2000). Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en ingénierie des connaissances. *In* Charlet, J., Zacklad, M., Kassel, G. et Bourigault, D., éditeurs : *Ingénierie des connaissances, évolutions récentes et nouveaux défis*, chapitre 19. Eyrolles.

Barque, L. (2008). *Description et formalisation de la polysémie régulière du français*.

Thèse de doctorat, Université Paris 7.

Barque, L. et Polguere, A. (2009). Structuration et balisage sémantique des définitions du Trésor de la Langue Française informatisé (TLFi). *In Proceedings of the Fourth International Conference on Meaning-Text Theory (MTT'2009)*.

Flouris, G., Manakanatas, D., Kondylakis, H., Plexousakis, D. et Antoniou, G. (2008). Ontology change : classification and survey. *The Knowledge Engineering Review*, 23(02), p. 117-152.

Gruber, T. (1995). Toward principles for the design of ontologies used for knowledge sharing ? *International journal of human-computer studies*, 43(5-6), p. 907-928.

Gugert, R. (2013). Scénarisation d'interactions avec les objets du formalisme des Graphes d'Unités et prototypage d'un éditeur de définitions lexicographiques formelles. Rapport technique, Inria.

Iordanskaja, L. et Mel'čuk, I. A. (1990). Semantics of Two Emotion Verbs in Russian : BOJAT'SJA 'to be afraid' & NADEJAT'SJA 'to hope'. *Australian Journal of linguistics*, 10(2), p. 307-357.

Kahane, S. (2003). The Meaning-Text Theory. *Dependency and Valency, An International Handbooks of Contemporary Research*, 25(1), p. 546-569.

Leclère, M. (1998). Raisonner avec des définitions de types dans le modèle des graphes conceptuels. *Revue d'intelligence artificielle*, 12(2), p. 243-278.

Lefrançois, M. (2014). *Représentation des connaissances sémantiques lexicales de la Théorie Sens-Texte : Conceptualisation, représentation, et opérationnalisation des définitions lexicographiques*. Thèse de doctorat, Université de Nice-Sophia Antipolis.

Lefrançois, M. et Gandon, F. (2011). ILexicOn : Toward an ECD-Compliant Interlingual Lexical Ontology Described with Semantic Web Formalisms. *In* Boguslavsky, I. et Wanner, L., éditeurs : *Proceedings of the 5th International Conference on Meaning-Text Theory (MTT'2011)*, p. 155-164, Barcelona, Spain. INALCO.

Lefrançois, M. et Gandon, F. (2013a). Rationale, Concepts, and Current Outcome of the Unit Graphs Framework. *In Proceedings of the 9th International Conference on Recent Advances in Natural Language Processing (RANLP 2013)*, p. 382-388.

Lefrançois, M. et Gandon, F. (2013b). Reasoning with Dependency Structures and Lexicographic Definitions using Unit Graphs. *In Proc. of the 2nd International Conference on Dependency Linguistics (Depling'2013)*, Prague, Czech Republic. ACL Anthology.

Lefrançois, M. et Gandon, F. (2013c). The Unit Graphs Framework : A graph-based Knowledge Representation Formalism designed for the Meaning-Text Theory. *In Proceedings of the 6th International Conference on Meaning-Text Theory (MTT'2013)*, Prague, Czech Republic.

Lefrançois, M. et Gandon, F. (2013d). The Unit Graphs Framework : Foundational Concepts and Semantic Consequence. *In Proceedings of the 9th International Conference on Recent Advances in Natural Language Processing (RANLP 2013)*, Hissar, Bulgaria. ACL Anthology.

Lefrançois, M., Gandon, F., Giboin, A. et Gugert, R. (2014). Un éditeur de définitions formelles pour les connaissances lexicales de la théorie Sens-Texte. *In Proc. 25e Journées Francophones d'Ingénierie des Connaissances (IC'2014)*, Clermont Ferrand, France.

Lefrançois, M., Gugert, R., Gandon, F. et Giboin, A. (2013). Application of the Unit Graphs Framework to Lexicographic Definitions in the RELIEF project. *In Proceedings of the 6th International Conference on Meaning-Text Theory (MTT'2013)*, Prague, Czech Republic.

Lux-Pogodalla, V. et Polguère, A. (2011). Construction of a French Lexical Network : Methodological Issues. *In Proceedings of the International Workshop on Lexical Resources*, Ljubljana.

Mel'čuk, I. (2004). Actants in semantics and syntax I : Actants in semantics. *Linguistics*, 42(1), 1-66.

Mel'čuk, I., Clas, A. et Polguère, A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Duculot, Paris/Louvain-la-Neuve.

Mel'čuk, I. A. et Wanner, L. (2001). Towards a lexicographic approach to lexical transfer in machine translation (illustrated by the German–Russian language pair). *Machine Translation*.

Polguère, A. (1990). *Structuration et mise en jeu procédurale d'un modèle linguistique déclaratif dans un cadre de génération de texte*. Thèse de doctorat, Département de linguistique, Université de Montréal.

Star, S. L. et Griesemer, J. R. (1989). Institutional Ecology, 'Translations' and Boundary Objects : Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, 19(3), p. 387-420.

Tesnière, L. (1959). *Éléments de syntaxe structurale*. C. Klincksieck, Colombes, Impr. ITE.

Wanner, L. (2003). Definitions of Lexical Meanings : Some Reflections on Purpose and Structure. *In Proceedings of the First international conference on Meaning-Text Theory (MTT'2003)*, p. 16-28.

Wanner, L. (2009). Explanatory Combinatorial Lexicology and NLP Applications in Specialized Discourse. *In First International Workshop on Terminology and Lexical Semantics (TLS'09)*, p. 54-66.

Summary

The Meaning-Text Theory (MTT) is one of the most precise and exhaustive linguistic theories. Representing knowledge of the MTT is a challenge for applications in lexicography, but also for injecting fine grained linguistic knowledge in NLP systems. This paper describes the methodology and the main outcomes of our research in conducing a Knowledge Engineering approach to model the lexical knowledge of the MTT and formalize its semantics. We focus more specifically on linguistic predicates, linguistic representations, and lexicographic definitions. We drew our inspiration from Bachimont's methodology, and operated in three main steps : (i) we extended the domain conceptualization, (ii) we chose an adequate Knowledge Representation formalism, and (iii) we operationalized the formalism for logical reasoning, and for sharing and querying knowledge on the web of data.

Les concordanciers bi-multilingues en ligne à l'épreuve de l'approche d'« archive » : le cas des termes de l'égalité H/F

Rachele Raus

rachele.raus@unito.it

Résumé. Notre contribution porte sur l'analyse contrastive de trois concordanciers bi-multilingues en ligne (*WebiText*, *Tradooit* et *Linguee*) par le biais d'une approche que nous avons définie d'« archive » (Raus 2013). Après avoir présenté notre corpus, nous préciserons ce que nous entendons par approche d'archive et en tirerons des critères de testage spécifique. Enfin, nous testerons les concordanciers par ces critères à l'aide de la traduction de quelques syntagmes terminologiques de l'égalité H/F de l'anglais en français et proposerons un certain nombre d'améliorations possibles.

1. Introduction

Notre contribution porte sur l'analyse contrastive de trois concordanciers bi-multilingues en ligne (*WebiText*, *Tradooit* et *Linguee*)[1], analyse qui se situe dans le prolongement des recherches que nous avons menées à ce sujet dans le cadre plus large de la traduction des termes de l'égalité H/F (Raus 2013). Nous testerons notamment ces outils à l'aide d'une approche descriptive aux néotermes composés de l'égalité H/F, approche que nous avons dénommée d'« archive ».

Les travaux en linguistique textuelle (Bourigault et Slodzian 1999, Condamines 2005…) et ceux en « ontologie textuelle » (Roche 2005 : 60) ont permis de relier les termes avec la dimension textuelle. La parution du

1 La date de dernière mise à jour des données analysées dans cet article est avril 2015. Actuellement, Webitext n'est plus disponible gratuitement sur Internet et les deux autres concordanciers analysés ont subi des changements qui parfois vont dans le sens de ce que nous avons proposé dans cet article.

« web sémantique » a ultérieurement favorisé les approches descriptives en terminologie, mettant ainsi à disposition des chercheurs des systèmes d'accès transversaux aux corpus parallèles déjà disponibles sur Internet. D'ailleurs, les concordanciers dont il sera question ici exploitent des corpus numériques de textes alignés de grande dimension.

Cependant, tout en s'appuyant sur des corpus qui parfois peuvent être similaires – c'est le cas, par exemple, des textes du Parlement européen –, les concordanciers en question affichent souvent des résultats différents, voire opposés (Raus 2013 : 132). Le traducteur s'y trouve donc confronté, sans avoir à disposition les outils qui lui permettraient de choisir de manière consciente le « bon » traduisant parmi les résultats trouvés. C'est justement à ce type de problème que nous essayerons d'apporter des solutions possibles.

Après avoir présenté notre corpus, nous exposerons des préliminaires théoriques sur ce que nous entendons par approche « d'archive », de manière à en tirer des critères de testage spécifique. Nous testerons donc les concordanciers à l'aide des cinq critères retracés et proposerons un certain nombre d'améliorations possibles, en passant par l'analyse critique de quelques syntagmes terminologiques de l'égalité H/F traduits de l'anglais en français.

2. Présentation du corpus

Parmi le nombre des concordanciers disponibles en libre accès sur internet, nous en avons sélectionné trois (*Tradooit*, *Linguee* et *Webitext*) qui non seulement sont multilingues mais qui s'appuient sur des corpus variés au niveau international[2]. Nos intérêts de recherche, en effet, portent désormais, et ce depuis un certain nombre d'années, sur la circulation des termes multilingues de l'égalité H/F dans les discours internationaux.

Les trois concordanciers multilingues analysés s'adressent aux traducteurs et mettent à disposition des outils traduisant en 3 (*Tradooit*), 25 (*Linguee*) ou 30 langues (*Webitext*) différentes. Bien évidemment, l'interrogation de l'outil se fait par paires de langues, les résultats étant donc affichés de manière bilingue.

Tradooit, d'Okidoo Inc. et Logosoft au Québec, se présente comme « une suite d'outils de traduction assistée par ordinateur qui regroupe notamment

2 Notre analyse ne portera donc pas sur les concordanciers payants disponibles sur le marché, comme, entre autres, ParaConc, TransSearch, MultiConcord, etc. ni sur les concordanciers concernant une seule organisation, bien qu'elle soit internationale.

une mémoire de traduction, une banque de terminologie et un concordancier bilingue[3] », ce dernier étant disponible en ligne depuis 2011. Il permet de trier les données de sortie par rapport aux résultats affichés par ordre de fréquence et par source. Il affiche en outre les résultats dans des sources terminographiques canadiennes comme *Termium plus*, le Bureau de traduction du Gouvernement canadien, la base terminologique de l'Ontario (ONTERM) et Wikipedia francophone.

Voici comment il se présente :

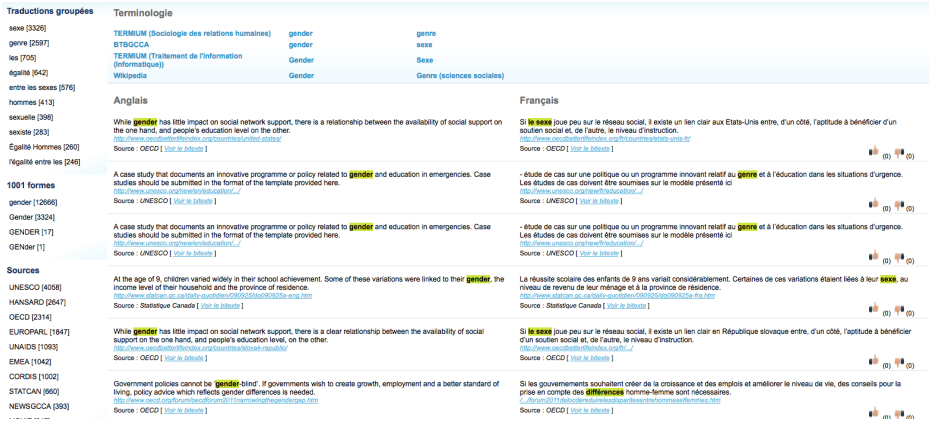


Figure 1 – *Affichage des résultats EN/FR dans* Tradoit *(entrée : «* gender *»).*

Linguee, produit par l'entreprise allemande *Linguee GmbH*, ne se pose pas comme concordancier, bien qu'il puisse être également utilisé de cette manière, mais comme « un outil de traduction unique en son genre », combinant un dictionnaire rédactionnel et un moteur de recherche qui peut interroger « plusieurs centaines de millions de textes bilingues[4] ». Tout en permettant, comme les autres concordanciers analysés, de récupérer les textes alignés d'où l'on tire les résultats, il ne permet pourtant pas de trier les résultats par

3 Citation tirée du site du concordancier. Nous constatons que les outils de TAO visent de plus en plus à aligner leurs différents outils, en essayant de dépasser en quelque sorte les décalages de méthode analysés par Bowker et Barlow (2008) entre concordanciers (bilingues) et mémoires de traduction.

4 *Idem*.

source, ne permet pas non plus de récupérer préalablement la liste des corpus interrogés, et n'affiche les fréquences des résultats que de manière intuitive, sous forme de camembert. Se posant également comme dictionnaire rédactionnel, il ne recourt pas à des sources terminographiques externes.

L'affichage des résultats dans *Linguee* se présente de la manière suivante :

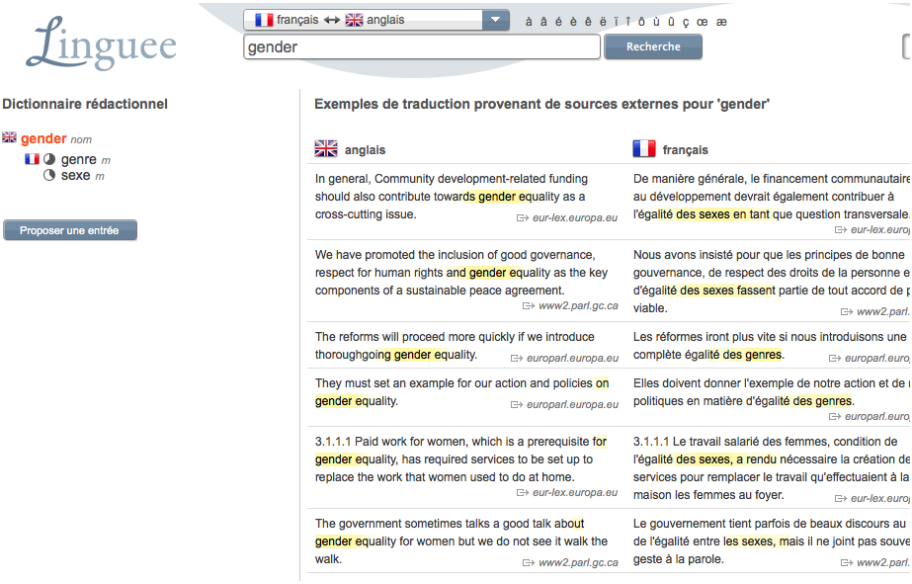


Figure 2 – *Affichage des résultats EN/FR dans* Linguee *(entrée : «* gender *»).*

Enfin, *Webitext*, de l'Université d'Ottawa au Canada, est décrit comme concordancier multilingue qui permet d'interroger « un corpus composé de plus de dix millions de pages Web du gouvernement [canadien][5] ». Comme *Tradooit*, il s'appuie sur *Termium plus* pour ce qui est de l'aspect terminographique, mais à la différence de celui-là, il permet de trier le corpus préalablement à l'interrogation de l'outil, en présentant une liste de sources très large.

5 *Idem*.

En fig. 3, nous représentons l'affichage des données dans cet outil :

Figure 3 – *Affichage des résultats EN/FR dans* Webitext *(entrée : «* gender *»).*

3. L'approche d'« archive » et ses critères

La littérature théorique sur les outils de traduction automatique ou assistée par ordinateur est désormais importante et ces outils sont devenus incontournables même à des fins didactiques (voir, entre autres, Bruno Osimo 2001, Marie-Claude L'Homme 2008…). Comme le souligne Froeliger (2013 : 216), la tendance actuelle est de favoriser une approche résolument numérique de la traduction. Cependant, force est de constater que l'acceptation

sans réserve de ces outils peut donner lieu à des dérives problématiques. Les approches « mixtes » en terminologie, qui essaient de concilier les méthodes linguistiques et statistiques avec l'extraction terminologique à partir de corpus numériques (voir, entre autres, Daille 1994, 1995 ; Ceausu & Despres 2004…), ne réfléchissent qu'en partie à ce type de problèmes. D'où la nécessité de proposer une approche qui puisse s'avérer complémentaire et que nous avons appelée d'« archive » pour nous rapprocher de l'analyse du discours à la française (voir à propos de cette notion Guilhaumou, Maldidier, Robin 1994 ; Maingueneau 1993, 1996…) et pour nous différencier des autres approches qui se sont développées en linguistique textuelle à la suite des travaux de Bourigault et Slodzian.

Notre approche insiste sur la nécessité de prendre en compte des conditions de production des termes en discours. En cela, nous poursuivons le questionnement des socioterminologues sur ce sujet (notamment Gaudin 2003, 2005), ainsi que les perspectives ouvertes par des études similaires qui ont justement commencé à poser la question du contexte discursif d'utilisation du terme (Delavigne 2003) et de la présence d'un certain degré de subjectivité des termes (Condamines et Rebeyrolles 1997).

Nous partons du constat que (Raus 2013 : 90) :

> Du point de vue de l'approche qui est la nôtre, la considération des instances énonciatives et des préconstruits lors de la circulation des termes en discours oblige à accepter l'idée de normes plurielles préalables aux variations terminologiques, variations dont l'analyste peut en outre retracer la genèse par l'analyse d'archive.

Ce constat nous permet de fixer au moins cinq critères d'archive à partir desquels nous testerons les outils en question :

1) L'affichage du cotexte pour repérer la présence éventuelle de *configuration d'énoncés*[6] ;

2) La prise en compte de la *source énonciative* et des *sous-ensembles énonciatifs*[7] ;

3) La prise en compte des sous-ensembles *texto-discursifs*[8] ;

4) La prise en compte de l'*interdiscours* (reprises interdiscursives), voire de la *mémoire discursive (préconstruit)*[9];

5) La prise en compte du *colinguisme*, à savoir le rapport entre langues instituées (Branca-Rosoff 2001 : 5), lors de la stabilisation de la norme linguistique.

Ces critères permettent de tenir compte des variations terminologiques (diastratique, diaphasique, diachronique et diatopique), comme nous le verrons par la suite, et s'ajoutent aux *Critères d'évaluation comparative des ressources terminologiques* que l'ISO recommande dans sa norme 23185 de 2009, notamment à l'égard de l'accès et de la sélection des données de la part de l'utilisateur (2009 : 10-11).

4. Comparaison des trois outils

Nous présentons en tableau 1 les résultats de la toute première comparaison des trois concordanciers, testés par rapport aux critères d'archives préalables.

6 Par configuration d'énoncé il faut entendre la « constitution de dispositifs d'énoncés nécessairement hétérogènes » à des fins d'analyse linguistique (Charaudeau, Maingueneau 2002 : 121).

7 La notion de sous-ensemble énonciatif est fortement liée à celle d'espace discursif (Maingueneau 1996 : 14 ; Charaudeau, Maingueneau 2002 : 97). Par exemple, parmi les discours de l'OCDE, la présence de discours de comités variés, tel le Comité d'aide au développement, produit l'émergence d'espaces discursifs précis qui créent des sous-ensembles énonciatifs spécifiques.

8 La notion d'ensemble texto-discursif est étroitement liée à celles de genres de textes et de discours (à propos de ces derniers, voir Maingueneau 1996 : 44).

9 Par « mémoire discursive » il faut entendre le fait qu'« à chaque moment le discours peut renvoyer à un énoncé précédent » (Maingueneau 1996 : 55-56 ; voir aussi Charaudeau, Maingueneau 2002 : 371). En ce sens, le « préconstruit » est la « trace, dans l'énoncé, d'un discours antérieur » (Charaudeau, Maingueneau 2002 : 464).

	Tradooit	*Linguee*	*Webitext*
1.	Il permet de récupérer le cotexte complet de l'occurrence	Il permet de récupérer le cotexte complet de l'occurrence	Il permet de récupérer le cotexte complet de l'occurrence
2.	Il permet de sélectionner les différentes sources textuelles	Ø	Il permet de sélectionner préalablement les corpus / sources à interroger
3.	Il a une bonne flexibilité des données de sortie	Ø	Il a une bonne flexibilité de sélection des catégories de données
4.	Il gère les syntagmes et les collocations et en enregistre les fréquences	Il gère partiellement les syntagmes et les collocations et affiche le pourcentage approximatif des mots simples	Il gère les syntagmes et les collocations mais sans en préciser les fréquences
5.	Il renvoie à *Termium plus*, au BT canadien, à Wikipedia (FR), à ONTERM pour la normalisation des termes	Il se présente aussi comme un dictionnaire rédactionnel	Il renvoie à *Termium plus* pour la normalisation des termes

Tableau 1 – *Comparaison des outils (*Tradooit, Linguee, Webitext*) par rapport aux critères d'archives.*

Bien évidemment, nous allons y revenir à travers l'analyse de cas d'études spécifiques, mais nous pouvons dire d'emblée que *Tradooit* apparaît le plus performant des trois outils en question : il permet de sélectionner les sources textuelles des données de sortie et gère les syntagmes et les collocations par ordre de fréquences. La structure présente une haute flexibilité des données de sorties, qui peuvent être triées non seulement par rapport à la source textuelle mais aussi par rapport aux résultats affichés. *Linguee*, par contre, est le moins performant, puisqu'il ne permet de récupérer que le seul cotexte des données de sortie et a des problèmes à gérer les syntagmes et collocations, comme nous verrons ensuite. *Webitext*, enfin, permettrait le tri des données préalablement à l'interrogation de l'outil par un système d'affichage de sélection des corpus et des sous-corpus ; en outre, il gère correctement les collocations mais sans en préciser les fréquences. Nous allons maintenant plus loin dans la comparaison et analyserons nos cinq critères dans autant de paragraphes

spécifiques, tout en interrogeant les outils à partir des termes de l'égalité H/F (EN>FR).

4.1. L'affichage du cotexte

Pour ce qui concerne le premier critère retracé, tous les concordanciers analysés permettent aisément de récupérer le cotexte d'occurrence des termes par l'affichage des segments-sources alignés ainsi que le site du document. De ce point de vue, donc, le système de repérage des concordanciers est normalement performant[10]. Nous remarquons juste deux problèmes, à savoir l'absence dans *Webitext* du surlignage des occurrences recherchées dans les textes d'arrivée, ce qui rend difficile parfois le repérage du terme, et l'affichage du seul segment d'occurrence du bitexte concerné dans *Linguee* alors que *Tradooit* et *Webitext* permettent de récupérer également les bitextes entiers. Nous passons donc aux quatre autres critères afin de tester ultérieurement les outils en question.

4.2. La prise en compte de la source énonciative et des sous-ensembles énonciatifs

Par rapport à la source énonciative, nous ne pouvons qu'évaluer la manière dont les outils permettent d'interroger la source de provenance des textes des corpus concernés, tout en précisant qu'elle ne correspond pas vraiment à la notion d'énonciateur. Par exemple, la source OCDE renvoie bien évidemment à la source Internet du site de cette organisation qui par là est instituée comme « garant » ultime du dit des textes concernés. Cependant, la source informatique ne correspond pas forcément aux différentes instances de l'OCDE ni, de ce fait, aux énonciateurs spécifiques (experts, comités…) qui ont rédigé ces mêmes textes. Ce n'est donc qu'en accédant au texte d'occurrence du terme qu'on peut en connaître l'énonciateur spécifique[11].

10 Remarquons que l'affichage du cotexte est utile au traducteur mais aussi à l'analyste du discours qui peut alors disposer aisément des configurations d'énoncés intéressants pour son analyse.

11 La notion d'énonciateur dans le discours institutionnel, notamment des institutions internationales, mérite, à vrai dire, des éclaircissements ultérieurs. En effet, il est souvent fort difficile d'en retracer les contours exacts, étant donné la tendance au désancrage énonciatif qui caractérise certains types de documents (voir Maingueneau 2002 : 130) ou, vice-versa la présence de remaniements à plusieurs mains dans d'autres types de textes (par exemple, les rapports parlementaires de l'UE, pour lesquels cf. Raus 2010b : 117).

À ce sujet, nous avons interrogé les trois outils quant à la traduction du syntagme anglais « *gender equality* ». Voici les résultats affichés :

Tradooit[11]	*Linguee*	*Webitext*
Traductions groupées l'égalité des genres [957] l'égalité des sexes [856] l'égalité entre les sexes [503] l'égalité hommes-femmes [311] Égalité Hommes / Femmes [192] l'égalité entre les hommes et les femmes [110] femme et de l'égalité des genres [71] l'égalité entre hommes et femmes [69] l'égalité hommes [48] l'égalité femme-homme [36] **Sources** UNESCO [1654] OECD [895] EUROPARL [662] HANSARD [419] UNAIDS [250] CORDIS [148] NEWSGCCA [74] CCCB [15] RESPECT [12] MOVIE [11]	**Résultats approchants :** equality *nom* égalité *f* gender *nom* genre *m* sexe *m*	Affichage des résultats dans le corpus bilingue aligné qu'on a sélectionné préalablement

Tableau 2 – *Comparaison des outils (*Tradooit, Linguee, Webitext*) par rapport à l'entrée «* gender equality *» (données de sortie).*

Tradooit affiche plusieurs équivalents français, ordonnés par fréquence. Son système s'avère en outre performant dans la mesure où il est ultérieurement flexible et permet de sélectionner les sources des résultats. Cependant, tout en ayant à disposition un système flexible, le traducteur ne peut s'orienter qu'en partie parmi ces résultats. Si, par exemple, le traducteur essaie de comprendre ce qui différencie l'utilisation de deux équivalents fréquents comme « égalité entre les sexes » et « égalité hommes-femmes » et qu'il les sélectionne pour en vérifier les contextes d'utilisation et, par là, les sources concernées, il verra que dans le premier cas, le système affichera des résultats provenant de sources variées (Parlement européen, Hansard, OCDE…) tandis que, dans

12 Par « Traductions groupées », *Tradooit* entend les résultats regroupés par variantes. Par exemple, la collocation « égalité hommes-femmes », qui affiche 311 occurrences dans le tableau, regroupe les variantes notationnelles « E/égalité homme-femme / femmes-hommes… ». En sélectionnant l'entrée en question, *Tradooit* permet le triage ultérieur des variantes concernées.

le deuxième, l'outil affichera presque exclusivement des résultats venant de la source OCDE.

Ces résultats sont d'ailleurs confirmés par la sélection des sources, en interrogeant l'outil à partir de la source OCDE : nous pouvons constater, en effet, que la fréquence de la collocation « égalité hommes-femmes » (274 occurrences) et de sa variante « égalité hommes / femmes » (165 occurrences) est largement majoritaire par rapport à « égalité des / entre les sexes », nettement plus rare (35 + 19 occurrences). Le traducteur est donc amené à croire que l'OCDE privilégierait le premier équivalent au deuxième et rien ne lui permet de différencier les sous-ensembles énonciatifs qui sont pourtant ceux qui déclenchent l'utilisation des équivalents concernés. En outre, la plupart de ces résultats dépendent du type de textes et de corpus que l'outil exploite.

Grâce à ce cas, nous sommes confrontés à la présence de la variation diastratique : en effet, c'est l'une des instances spécifiques de l'OCDE, à savoir le Comité d'aide au développement, qui a tendance à utiliser le syntagme « égalité hommes-femmes », notamment au niveau des statistiques (Raus 2010a : 209). D'ailleurs, si l'on interroge les bitextes d'occurrence des termes affichés par *Tradooit*, on retrouve normalement des statistiques ou des textes du CAD, ainsi que des documents cités, provenant justement de ce comité. Cependant, le traducteur n'est pas mis en mesure de choisir en connaissance de cause parmi les synonymes proposés au préalable par l'outil.

À ce sujet, il nous faut remarquer, en revanche, la possibilité qu'offre *Webitex*t de trier d'emblée des sous-corpus, ce qui permet de retracer des sous-ensembles énonciatifs plus précis. Il est vrai que pour l'OCDE, *Webitext* ne donne pas non plus la possibilité de trier les résultats par rapport à des instances spécifiques de l'Organisation. Cela dit, cet outil présente normalement des sous-corpus beaucoup plus variés et permet d'accéder aux données à partir de la source d'intérêt.

Voici, par exemple, comment se présentent les sous-corpus concernant la source « Gouvernement du Canada » dans *Webitext* alors que *Tradooit* considère seulement quelques-uns de ces sites :

Gouvernement du Canada
Tous les sites de la catégorie Gouvernement du Canada
Affaires autochtones et Développement du Nord Canada *(www.aadnc-aandc.gc.ca)*
Affaires étrangères et Commerce international Canada *(international.gc.ca)*
Agence canadienne d'inspection des aliments *(inspection.gc.ca)*
Agence des services frontaliers du Canada *(cbsa-asfc.gc.ca)*
Agence du revenu du Canada *(cra-arc.gc.ca)*
Agriculture et Agro-alimentaire Canada *(agr.gc.ca)*
Anciens combattants Canada *(www.vac-acc.gc.ca)*
Citoyenneté et Immigration Canada *(cic.gc.ca)*
Défense nationale *(forces.gc.ca)*
Environnement Canada *(ec.gc.ca)*
Gendarmerie royale du Canada *(www.rcmp-grc.gc.ca)*
Industrie Canada *(ic.gc.ca)*
Infrastructure Canada *(infc.gc.ca)*
Justice Canada *(justice.gc.ca)*
Ministère des Finances Canada *(fin.gc.ca)*
Patrimoine canadien *(pch.gc.ca)*
Pêches et Océans Canada *(dfo-mpo.gc.ca)*
Ressources humaines et Développement des compétences Canada *(rhdcc.gc.ca)*
Ressources naturelles Canada *(nrcan-rncan.gc.ca)*
Santé Canada
Service Canada *(servicecanada.gc.ca)*
Statistique Canada *(www.statcan.gc.ca)*
Sécurité publique Canada *(publicsafety.gc.ca)*
Transports Canada *(www.tc.gc.ca)*
Travaux publics et Services gouvernementaux Canada *(www.tpsgc-pwgsc.gc.ca)*

Figure 4 – Sous-corpus de la source « Gouvernement du Canada » dans Webitext.

Au niveau de l'Union européenne aussi, si *Tradooit* se limite au seul corpus du Parlement européen, comme nous le verrons, *Webitext* intègre également le portail de l'Union européenne et surtout le corpus législatif EUR-lex. Cet outil reste cependant limité du fait qu'il n'affiche pas les fréquences des résultats trouvés et qu'il ne permet pas d'interroger ultérieurement les données de sortie.

Enfin, *Linguee* a le défaut de ne pas gérer correctement les syntagmes (voir tab. 2), ce qui le rend moins performant par rapport au quatrième de nos critères, sur lequel nous reviendrons. Mais surtout, il ne permet de relier en aucune manière les résultats aux sources, si ce n'est par le renvoi à la source des bitextes d'occurrence, que le traducteur doit donc analyser cas par cas.

4.3. La prise en compte des sous-ensembles texto-discursifs

En proposant l'exemple de la «*gender equality*», nous avons souligné la tendance du CAD à utiliser l'équivalent «égalité hommes-femmes», notamment dans des sous-ensembles discursifs concernant les statistiques. La question des sous-ensembles texto-discursifs est d'ailleurs étroitement liée au critère des sous-ensembles énonciatifs et elle y ajoute la tendance ultérieure à la variation terminologique par la prise en compte de la diaphasie. En effet, le même énonciateur peut utiliser des équivalents différents en fonction des types de texte ou des discours spécifiques[13]. Nous avons démontré ailleurs (Raus 2013 : 48-50) que la tendance à la variation diaphasique caractérise, par exemple, le choix des équivalents des termes de l'égalité H/F des énonciateurs et traducteurs des commissions du Parlement européen dans des sous-discours spécifiques, comme la justification des amendements dans les rapports législatifs.

La prise en compte des sous-ensembles texto-discursifs commence à être considérée au niveau théorique en linguistique de corpus, bien qu'elle soit souvent réduite aux domaines du savoir[14]. Si l'on se réfère tout d'abord à ce critère comme à la possibilité de trier les résultats par domaine, il est clair que *Tradooit* et *Webitext* sont de nouveau les plus performants, *Linguee* ne permettant que l'affichage des sources sans prévoir le critère du triage par domaine.

13 La différence d'utilisation des termes par rapport au contexte d'utilisation est à la source de plusieurs recherches terminologiques concernant notamment les discours de vulgarisation (voir, entre autres, Delavigne 2003 et Druetta 2007).

14 Il ne faut surtout pas confondre le «domaine» du savoir avec le «domaine» renvoyant à l'adresse internet auquel nous ferons également référence dans notre article. En ce sens, nous marquerons le deuxième par l'étiquette «domaine internet».

Figure 5 – *Affichage de quelques résultats de traduction EN>FR de « gender equality » dans* Linguee.

À vrai dire, *Tradooit* et *Webitext* ne prévoient pas, eux non plus, un système de triage exclusif par domaine, mais la possibilité de sélectionner des sources spécifiques, ce qui va en partie dans ce sens. *Tradooit*, par exemple, exploite des sources concernant des domaines spécifiques. Citons, entre autres, « Musée Sans Frontières », « Producteurs laitiers du Canada », « Environnement Canada », « Cineuropa », « Société canadienne du sang », « La Conférence des évêques catholique du Canada »… *Webitext*, comme nous l'avons déjà souligné, permet l'accès à des corpus encore plus variés. En outre, il est le seul à permettre de les sélectionner au préalable, entre autres, par domaine :

Figure 6 – *Sélection des corpus dans* Webitext.

Bien que le critère disciplinaire ne soit pas exclusif, *Webitext* permet de choisir parmi quelques domaines spécifiques (en fig. 6, les domaines « légal, santé, finance »), ce qui en fait donc l'outil le plus performant en ce sens.

Faisons remarquer, pour finir, que le domaine ne correspond pas exactement aux sous-ensembles texto-discursifs, dont la prise en compte demande d'ajouter des filtres beaucoup plus fins et spécifiques.

4.4. La prise en compte de l'interdiscours

La notion d'interdiscours, à laquelle se rattache fortement celle de mémoire discursive, n'est présente au niveau des outils analysés que sous forme de collocations et de fréquences. En effet, les seuls éléments pris plus ou moins en compte à ce niveau par les concordanciers analysés sont justement la fréquence des termes et des collocations. Rien n'indique, par contre, la présence d'une occurrence terminologique venue d'ailleurs et reprise via l'interdiscours. Pour illustrer cette limite, nous allons analyser la traduction française du syntagme anglais « *gender mainstreaming* ».

Voici les résultats affichés par *Tradooit* :

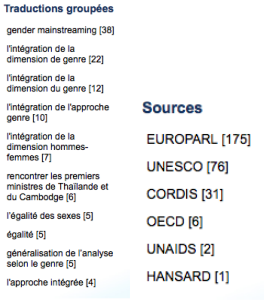


Figure 7 – *Affichage des résultats de traduction EN>FR de « gender mainstreaming » dans* Tradooit.

Mis à part l'emprunt, nous voyons que le calque français le plus diffusé est « intégration de la dimension de genre » et le moins diffusé est « approche intégré ». Nous remarquons, entre autres, le poids du discours parlementaire européen par rapport aux fréquences. En effet, le triage par le corpus EUROPARL, donne les résultats suivants :

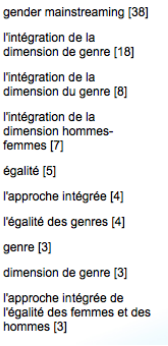


Figure 8 – *Résultats des équivalents français de «* gender mainstreaming *» dans le corpus EUROPARL (*Tradooit*).*

Encore une fois rien n'indique comment utiliser les différents résultats ni d'où ils viennent. De manière indirecte, c'est *Webitext* qui nous permet de retracer la source de l'équivalent le moins utilisé « approche intégrée (de

l'égalité entre les hommes et les femmes). Voici en effet, les toutes premières occurrences affichées par l'outil en question :

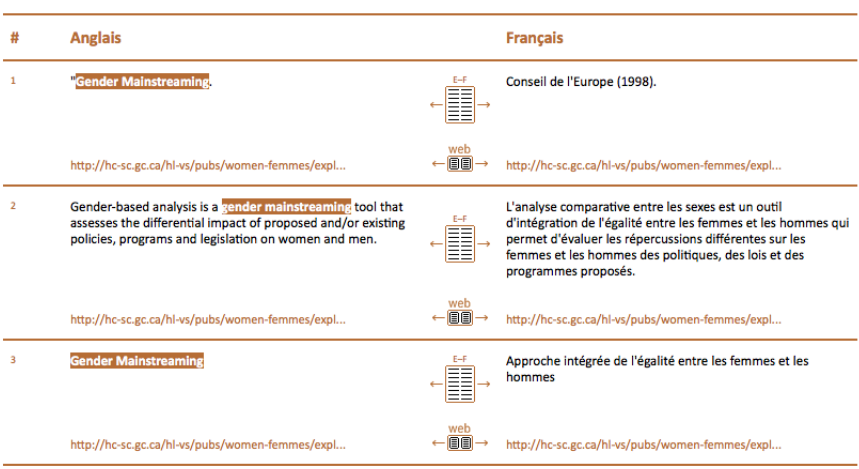


Figure 9 – *Résultats des équivalents français de «* gender mainstreaming *» dans tous les sites de la catégorie Gouvernement du Canada (*Webitext*).*

Nous n'avons pas encore interrogé le corpus du Parlement européen, mais les tout premiers résultats concernant les sites du Gouvernement du Canada nous permettent de comprendre en quoi consiste le problème de l'interdiscours. Si nous récupérons le texte complet du bitexte de la figure 9, nous comprenons qu'il s'agit du renvoi au document que le Conseil de l'Europe a rédigé en 1998 sur le «*gender mainstreaming*», où ce terme a été justement traduit par «approche intégrée de l'égalité entre les femmes et les hommes». Nous retrouvons là, en effet, l'équivalent le moins utilisé par le Parlement européen au niveau des résultats dans *Tradooit*. Il faut également savoir que c'est justement en 1998 que la Commission européenne rédige un glossaire des *100 mots de l'égalité* où l'équivalent proposé est «intégration de la dimension de genre». Ce terme deviendra l'équivalent utilisé au PE à partir de 2007, tandis qu'auparavant le Parlement avait privilégié l'équivalent venant du Conseil de l'Europe (Raus 2013 : 66-72). Rien n'indique donc le passage interdiscursif des termes, leur variation diachronique (vieillissement du terme «approche intégrée de l'égalité» dans le discours du PE de l'après 2007), ni la présence

de la double énonciation par la citation, si ce n'est que par la lecture du bitexte concerné.

Le choix du corpus du Parlement européen ou des autres corpus européens dans *Webitext* n'aide en rien le traducteur dans la mesure où l'outil n'est pas flexible pour ce qui est des données de sortie. Par exemple, parmi les résultats, *Webitext* affiche le *Rapport sur une approche intégrée de l'égalité des femmes et des hommes dans les travaux du Parlement européen (A7-0351/2011)*, présenté le 27 octobre 2011 par Mikael Gustafsson, rapporteur fictif de la Commission des droits de la femme et de l'égalité des genres. Dans ce rapport d'initiative du PE, nous trouvons la réutilisation de l'équivalent désormais le moins fréquent. Rien ne permet au traducteur de voir que ce document remanie en fait un document du 21 décembre 2006, période à laquelle le PE utilisait encore l'équivalent du Conseil de l'Europe, présenté par la rapporteure Anna Záborská sur l'approche intégrée de l'égalité des femmes et des hommes dans le cadre des travaux des commissions. Ce rapport sera repris ensuite plusieurs fois, entre autres, toujours par Záborská le 22 avril 2009[15]. Lors du remaniement, plusieurs segments du texte sont entièrement repris tels quels, comme, par exemple, « l'importance de l'utilisation d'une terminologie et de définitions précises conformes aux normes internationales, lors de l'emploi des termes relatifs à l'approche intégrée de l'égalité », extrait qui est présent aussi dans *Webitext*.

4.5. La prise en compte du colinguisme

Du point de vue du colinguisme, à savoir de la relation entre langues instituées, tels l'anglais et le français au niveau des corpus bi-multilingues exploités par les outils, le français se caractérise par la tendance à adapter les termes de l'égalité venant de l'anglais, notamment en ce qui concerne le concept « *gender* » et ses dérivés par composition (voir, entre autres, Sénac-Slawinski 2006, Raus 2012…). Des contaminations majeures entre ces deux langues caractérisent, par contre, certaines variantes diatopiques du français et de l'anglais lors de leur coexistence bilingue sur un territoire. C'est ce qui arrive par exemple dans le cas de « *gender parity* », syntagme anglais obtenu par calque à partir du terme français « parité ». Cette forme est particulièrement fréquente dans l'anglais canadien, comme on peut également le

15 À ce sujet, soulignons l'importance grandissant de l'utilisation des mémoires de traduction au niveau de l'Union européenne et, par conséquent, la tendance à retraduire des textes sur la base de textes précédents qui ont été précédemment traduits.

constater du fait de sa stabilisation au niveau de la norme linguistique dans les résultats affichés par les banques terminographiques canadiennes présentes dans *Tradooit* et dans *Webitext*, notamment *Termium plus*. De même, le dictionnaire rédactionnel de *Linguee* se réfère en particulier aux usages du Gouvernement canadien et des sources onusiennes, ces dernières privilégiant souvent les variantes canadiennes (Raus 2013 : 92)[16].

Au sujet de la norme linguistique, nous faisons remarquer :

- la tendance de *Linguee* au remplissage semi-automatique des syntagmes recherchés, ce qui facilite l'interrogation de l'outil de la part du traducteur ;
- les désalignements existant entre les traductions empiriques et les traductions normalisées dans *Tradooit*, ce qui pouvait déjà se déduire lors de l'analyse de l'entrée « *gender equality* ». Si les banques terminographiques canadiennes privilégient « égalité des sexes » pour traduire « *gender equality* » et « parité hommes-femmes » pour traduire « *gender parity* », les usages enregistrés par *Tradooit* d'après le critère des fréquences restituent des résultats décalés (« égalité des genres », « parité des sexes »).

Si l'on passe aux usages, et donc aux bitextes affichés par les outils pour l'entrée en question, les résultats de *Tradooit* concernant « *gender parity* » montrent qu'effectivement le syntagme est présent en prévalence dans le corpus canadien Hansard et encore plus dans le corpus Unesco. Ceci confirme la tendance des traducteurs des instances onusiennes à privilégier, au niveau colingue, les variantes américaines, entre autres canadiennes, des termes. La rareté de la variante diatopique au niveau des corpus européens (PE, Cineurope, News européennes Cordis) ne doit donc pas étonner, étant donné le choix colingue effectué par les Européens qui consiste à privilégier l'anglais britannique.

Ce dernier cas nous permet de réfléchir sur les moyens mis à la disposition du traducteur pour tenir compte de la diatopie, notamment dans des outils où la présence de corpus canadiens, et donc des variantes canadiennes (aussi anglaises que françaises), est importante. En effet, bien que les sources canadiennes aient un poids remarquable non seulement au niveau des corpus mais aussi à l'égard de la norme terminographique, rien ne permet au traducteur de détecter la présence des variantes diatopiques. Le seul indice affiché par

16 Faisons remarquer la tendance différente qui caractérise le Québec. En effet, le *Grand dictionnaire terminologique* n'affiche aucun résultat pour ce syntagme.

Tradooit en ce sens est celui de la fréquence (ex. rareté de « *gender parity* » dans les sites de l'UE), qui, par ailleurs, n'est pas nécessairement liée à la diatopie et pourrait s'expliquer différemment, par exemple en raison de la diaphasie.

En revanche, *Webitext* semble prendre en compte la variation diatopique par la sélection préalable de tous les sites de la catégorie domaine internet «.ca». Si cette « astuce » résout en partie la question, elle ne permet pas de retracer les variantes canadiennes présentes, par exemple, dans les sites des instances onusiennes et qui n'ont pas un nom de domaine internet «.ca».

5. Des améliorations possibles

D'après notre analyse, les trois concordanciers en question pourraient donc être améliorés par rapport aux critères repérés, de manière à aider le traducteur lors du choix du traduisant à utiliser dans les différents discours.

Tout d'abord, à l'égard des énonciateurs et des sous-ensembles énonciatifs, la possibilité de trier des sources en amont ou en aval de l'interrogation devrait être possible aussi pour les outils qui, comme *Linguee*, ne le permettent pas encore. Cela dit, on pourrait essayer de tenir compte de sous-corpus de plus en plus spécifiques et variés, comme le fait *Webitext*. Si cet aspect ne résout pas la question des énonciateurs, puisque la source, comme nous l'avons dit, ne correspond pas nécessairement à l'énonciateur, c'est pourtant la seule manière pour l'instant de s'y rapprocher et de tenir compte de la variation diastratique des termes considérés. Il nous faut appréhender aussi un autre aspect intéressant à ce sujet, dans le parcours qui nous mène à l'adoption de critères de plus en plus sémasiologiques, à savoir le fait que *Tradooit* permet à l'utilisateur enregistré d'afficher le traducteur des textes d'arrivée, puisque « ce filtre peut aider à évaluer la fiabilité des équivalents trouvés, selon l'expérience, la spécialité, la qualité et le style des traducteurs[17] ». À ce sujet, nous nous posons la question de savoir si, finalement, il ne serait pas également possible d'ajouter un filtre concernant justement l'énonciateur spécifique des textes constituant les corpus ou de prévoir tout au moins des filtres qui permettent de trier de manière informatique les sous-ensembles énonciatifs, en s'appuyant sur le parcours spécifique indiqué dans l'adresse URL de la source. De cette manière, par exemple, le système pourrait trier les résultats présentant l'acro-

17 Citation tirée du site du concordancier.

nyme « cad /dac » du Comité d'aide au développement au niveau de l'adresse URL de la source OCDE.

Au niveau des sous-ensembles texto-discursifs, la possibilité de trier les résultats par domaine, comme le fait en partie *Webitext*, va dans le sens de la prise en compte de la variation diaphasique. Ceci dit, ne pourrait-on pas aller plus loin, en permettant à l'outil de trier les corpus par types de discours ? À ce sujet, les recherches de Delavigne (2003) et de Condamines et Dehaut (2011) s'avèrent prometteuses, grâce à la focalisation effectuée, non seulement sur les similitudes des ensembles lexicaux-sémantiques, mais aussi sur le repérage automatique de marqueurs discursifs qui pourraient renvoyer à la présence d'un type spécifique de discours. La démarche que ces chercheurs proposent sur le discours de vulgarisation et / ou didactique pourrait s'appliquer également à d'autres discours. Citons également les recherches que Ward Van de Velde (2013) a conduites sur un corpus médico-administratif et qui s'appuient sur les critères linguistico-discursifs de Van Campenhoudt, ouvrant, là encore, des perspectives intéressantes.

Remarquons que la possibilité ultérieure de croiser les résultats obtenus par domaine, et, dans une situation idéale, par discours spécifiques avec le critère diastratique des sous-ensembles énonciatifs permettrait finalement de voir comment un même énonciateur peut éventuellement opter pour des équivalents différents par rapport aux contextes discursifs variés et rendre compte des usages au sens le plus large.

À l'égard de l'interdiscours, la prise en compte de la diastratie, comme le font en partie *Tradooit* et encore plus *Webitext* lors du triage des sources, ne permet pas de rendre compte de la circulation des termes. Bien évidemment, la possibilité d'afficher les fréquences, comme le fait *Tradooit* et de manière générique *Linguee*, est un élément important au niveau de la mémoire discursive, mais cet élément à lui seul ne permet pas de rendre raison des circulations discursives des termes. Par contre, la possibilité de trier les documents des corpus par rapport à la diachronie serait une première amélioration souhaitable et relativement facile à implémenter sous l'aspect informatique. En effet, ce filtre permettrait, non seulement de retracer les changements terminologiques des instances sur l'arc temporel, mais aussi de retracer raisonnablement la source de la première utilisation du terme et, par là même, de suivre sa circulation dans d'autres espaces discursifs. Dans le cas du « *gender mainstreaming* », la possibilité de disposer des sous-corpus des instances Conseil de l'Europe et Commission européenne, outre celui du PE, et de pouvoir y accéder non seulement par le filtre des fréquences mais aussi par la date

des documents exploités permettrait de suivre la provenance et la circulation à l'intérieur du discours parlementaire des équivalents français venant des deux autres instances.

Enfin, quant au tout dernier critère que nous avons pris en compte, à savoir les choix colingues des institutions – et par conséquent des rédacteurs et/ou des traducteurs – qui peuvent privilégier des variantes diatopiques à d'autres, nous proposons les réflexions suivantes :

(a) il faudrait d'abord s'interroger sur le rôle joué par les sources canadiennes, notamment au niveau de la terminographie. En effet, par rapport au critère qualitatif de l'officialité des sources, recommandé par l'ISO quant à l'utilisation des sources faisant autorité (ISO 2009 : 7), il faudrait ajouter un critère quantitatif qui tienne compte des variétés diatopiques. En ce sens, nous nous demandons par exemple pourquoi *Tradooit*, qui est un outil québécois, ne prend pas également en compte le *Grand dictionnaire terminologique* comme source de la norme. Remarquons, à cet égard, que la tendance colingue québécoise se différencie de la tendance canadienne en général, du moins pour ce qui est des traductions des termes de l'égalité en français (Raus 2014) ;

(b) le fait qu'à la différence de *Tradooit* et de *Webitext*, qui affichent clairement leur choix de privilégier des sources terminographiques canadiennes, *Linguee* privilégie le dictionnaire rédactionnel ne résout pas le problème de la diatopie, puisque son dictionnaire affiche les données les plus fréquentes à l'intérieur des corpus concernés et que dans ces derniers, la présence des documents canadiens peut rester importante ;

(c) le fait de pouvoir trier les résultats par domaine internet «.ca» est en effet un remède souhaitable, bien qu'il soit insuffisant à lui seul, étant donné, entre autres, la tendance colingue des instances onusiennes à privilégier les variantes américaines.

6. Conclusions

Après une toute première comparaison des trois outils (fig. 1), l'analyse plus fine que nous avons conduite par le biais de quelques cas de termes de l'égalité H/F traduits de l'anglais en français, démontre que *Webitext* est un concurrent à la hauteur de *Tradooit*, notamment au niveau des critères que nous avons ajoutés, tandis que *Linguee* se confirme comme l'outil le moins performant.

En donnant la possibilité à l'utilisateur de trier les sources préalablement à l'interrogation de l'outil, *Webitext* permet à l'utilisateur d'être conscient d'emblée de la présence de normes plurielles, issues des pratiques de rédaction / traduction des textes. Ces normes varient, en effet, en fonction des sous-énonciateurs (variation diastratique) et des sous-ensembles texto-discursifs (variation diaphasique). La présence de sous-corpus variés est pour l'instant le seul moyen qui permette de tenir compte de ces deux aspects. En outre, la possibilité de trier le corpus par le domaine internet «.ca» a pour conséquence que *Webitext* est actuellement le seul des trois outils à poser la question de la variation diatopique des termes, du moins en ce qui concerne les occurrences empiriques dans les bitextes.

La majeure flexibilité des données de sortie, notamment la possibilité de trier les résultats par rapport aux sources et aux fréquences, reste l'atout principal de *Tradooit*, qui, en outre, tient également compte des variantes notationnelles des syntagmes concernés.

Les deux concordanciers pourraient donc améliorer leurs performances à condition de :

- ajouter des sous-corpus ultérieurs ;
- améliorer la flexibilité de l'un (*Webitext*), pour ce qui est des données de sortie, et de l'autre (*Tradooit*) lors de la sélection préalable des données ;
- donner la possibilité de trier les données par chronologie ;
- permettre de trier les données par domaine internet (par exemple, non seulement «.ca», mais également «.org», «.eu»…).

Bien évidemment, il reste beaucoup à faire dans la voie de la prise en compte des critères discursifs concernés, mais il nous semble que le fait d'aller vers des critères de triage sémasiologiques s'impose dans la mesure où le web sémantique, qui favorise entre autres les approches d'ontologie textuelle, demande des démarches nouvelles. Or, dans le contexte actuel, le normalisateur ne peut plus négliger les usages que le web contribue à stabiliser par la réitération aux dépens de tout principe onomasiologique préalable.

Bibliographie

Bourigault, D., Slodzian M. (1999) « Pour une terminologie textuelle », *Terminologies nouvelles*, n°19, p. 29-32.

Bowker, L., Barlow, M. (2008) « A comparative evaluation of bilingual concordancers and translation memory systems », in Elia Yuste Rodrigo (ed) *Topics in Language Resources for Translation and Localisation*. Amsterdam/Philadelphia : John Benjamins Publishing Company, p. 1-22

Branca-Rosoff, S. (éd.) (2001) *L'institution des langues – Autour de Renée Balibar*. Paris : Ed. de la Maison des Sciences de l'Homme.

Ceausu, V., Despres, S. (2004) « Une approche mixte pour la construction d'une ressource terminologique »

URL : <http://liris.cnrs.fr/~ic04/programme/articles/Ceausu-IC2004.pdf>

Charaudeau, P., Maingueneau, D. (2002) *Dictionnaire d'analyse du discours*. Paris : Éditions du Seuil.

Condamines, A. (2005) « Linguistique de corpus et terminologie », *Langages*, n°157, p. 36-47.

Condamines, A., Dehaut, N. (2011) « Mise en œuvre des méthodes de la linguistique de corpus pour étudier les termes en situation d'innovation disciplinaire : le cas de l'exobiologie », *Meta. Journal des traducteurs*, n°56/2, p. 266-283.

Condamines, A., Rebeyrolle, J. (1997) « Point de vue en langue spécialisée », *Meta. Journal des traducteurs*, n°42/1, p. 174-184.

Delavigne, V. (2003) « Quand le terme entre en vulgarisation », Actes des cinquièmes rencontres « *Terminologie et Intelligence artificielle* ». Strasbourg : LIA- Ensais, p. 80-91.

Druetta, R. (2007) « Les *Lettres aux Actionnaires*, ou ce que la vulgarisation peut nous apprendre de la langue de spécialité », *Synergies Italie*, n° 3, p. 67-77.

URL : <http://gerflint.fr/Base/Italie3/druetta.pdf>

Froeliger, N. (2013) *De la traduction pragmatique. Les noces de l'analogique et du numérique*. Paris : Les Belles Lettres.

Gaudin, F. (2003) *Socioterminologie. Une approche sociolinguistique de la terminologie*. Bruxelles : De Boeck.

– (2005), « La socioterminologie », *Langages*, n°157, p. 81-93.

Guilhaumou, J., Maldidier, D., Robin, R. (1994) *Discours et archive*. Liège : Mardaga.

ISO (2009) *Norme 23185. Critères d'évaluation comparative des ressources terminologiques*, Genève, ISO.

L'Homme, M.-C. (2008) *Initiation à la traductique*. Montréal : Linguatech éd., 2e édition.
Maingueneau, D. (1993) « Analyse du discours et archive », *Semen*, n°8
URL : <http://semen.revues.org/4069>
– (1996) *Les termes clés de l'analyse du discours*. Paris : Éditions du Seuil.
– (2002) « Les rapports des organisations internationales : un discours constituant ? », G. Rist, *Les mots du pouvoir. Sens et non-sens de la rhétorique internationale*. Paris : PUF, p. 119-132.
Osimo, B. (2001) *Traduzione e nuove tecnologie. Informatica e Internet per i traduttori*. Milan : Hoepli.
Raus, R. (2010a) « Synonymie et autres mécanismes de co-référence dans la traduction des textes internationaux mixtes : vers une 'synonymie d'adaptation' ? », *Translation*, Vol. 28, 2009, n° 4, p. 206-214.
– (éd.) (2010b) *Multilinguismo e terminologia nell'Unione europea. Problematiche e prospettive*. Milano : Hoepli.
– (2012) « Terminologie de l'égalité entre les hommes et les femmes », *La banque des mots*, n°83, p. 5-23
– (2013) *La terminologie multilingue. La traduction des termes de l'égalité H/F dans le discours international*. Bruxelles : De Boeck.
– (2014) « Pour une approche discursive de la variation lexicale en terminologie : les termes français de la *gender equality* dans les glossaires des institutions transnationales », in Annick Farina, Valeria Zotti (eds), *La variation lexicale des français : dictionnaires, bases de données, corpus*, Paris : Champion, p. 333-344.
Roches, C. (2005) « Terminologie et ontologie », *Langages*, n°157, p. 48-62.
Sénac-Slawinski, R. (2006) « Le *gender mainstreaming* à l'épreuve de sa genèse et de sa traduction dans l'action publique en France », *Politique européenne*, n°20, 3, p. 9-33.
Van de Velde, W. (2013) « Le corpus BABELIRIS : d'une méthodologie de constitution aux principales caractéristiques linguistiques », *Équivalences*, n° 40/1-2, p. 163-189.

Sitographie[18]

Linguee : http://www.linguee.com

Tradooit : http://www.tradooit.com

18 Dernière date de consultation des sites : 1er juin 2014.

Webitext :

http://www.webitext.com

Summary

In this article, we analyse three online bi/multilingual concordancers (*WebiText*, *Tradooit* et *Linguee*) by an "archive" approach (Raus 2013). First of all, we describe our corpus, then we define what we mean by "archive" approach and we select some criteria to test the concordancers. Finally, we analyse some case studies from gender terminology in translation (EN> FR) in order to propose suggestions to improve concordancer performances.

Étude des termes diagnostiques des troubles du langage en orthophonie

Frédérique Brin-Henry

Centre Hospitalier de Bar-le-Duc
Cellule Recherche
1, boulevard d'Argonne - 55012 Bar le Duc
ATILF UMR 7118 - CNRS/Université de Lorraine
frederique.henry@atilf.fr
https://perso.atilf.fr/frederiquebrinhenry/

Résumé. À travers une étude terminologique de la langue de l'orthophonie nous contribuons à la réflexion sur le lien entre les représentations des troubles du langage chez les orthophonistes et la réalité des troubles langagiers observés chez les patients. L'annotation et l'extraction terminologique semi-automatique au sein de 436 comptes rendus de bilan orthophonique ont permis d'élaborer une analyse syntactico-sémantique de 558 termes diagnostiques. L'examen de ces variations montre la présence remarquable d'unités lexicales non spécifiques (« difficultés ») et la scission entre deux groupes de syntagmes correspondant à une perception exclusive d'atteintes d'un 'système' ou d'atteintes d'un 'processus', pour ce qui concerne les troubles du langage. Les répercussions pour la terminologie et la nosologie du domaine sont discutées.

1. Introduction

L'orthophonie est une discipline scientifique et une pratique clinique de soins bien établie en Europe, qui s'est structurée au cours du XX^{e} siècle sur des fondamentaux théoriques interdisciplinaires (Morin 1994). Cette 'langue de spécialité' utilise un ensemble d'unités lexicales émanant de la langue générale, complétée par les domaines des sciences humaines et sociales (sciences du langage, psychologie en particulier) et des sciences médicales (neurologie, otorhinolaryngologie…). Ces unités restent assez peu explorées sur un plan lexicographique (Brin-Henry *et al.* 1997, 2004, 2011). Sur un plan terminolo-

gique, les variations diagnostiques puisent dans les nomenclatures nationales (Nomenclature Générale des Actes Professionnels, 2012) et les classifications internationales reconnues. Ainsi on retrouve :

La Classification Internationale des Maladies (CIM-10), publiée par l'Organisation Mondiale de la Santé (2008), et ses dérivés (SNOMED http://esante.gouv.fr/snomed/snomed/)

- Le Manuel diagnostique et statistique des troubles mentaux (*Diagnostic and Statistical Manual)* de *l'American Psychiatric Association*, dont la 5e version est parue en 2013 (DSM5).

Il faut noter que la littérature du domaine s'intéresse de plus en plus à une perspective attachée à l'interaction de l'individu avec son environnement comme dans :

- La Classification Internationale des Fonctionnalités et des Handicaps (CIF, 2001)
- Le Modèle de Développement Humain – Processus de Production du Handicap (MDH-PPH) (Fougeyrollas 1998).

La terminologie orthophonique bénéficie des évolutions parallèles à celles du domaine, qui touchent à la conception même de la maladie et à sa frontière avec la normalité. Nous nous interrogeons sur les termes diagnostiques utilisés en orthophonie en ce qu'ils objectivent la relation entre la réalité des difficultés du patient et la représentation que l'orthophoniste a de la pathologie. Ainsi nous explorons par un abord linguistique le lien entre le « concept » - au sens ontologique du terme (Roche 2005) - et le syntagme - au sens terminologique du terme (Cabré 2000) - et nous l'étudions grâce à l'identification semi-automatique et le recueil des réalisations linguistiques des termes diagnostiques dans les écrits orthophoniques.

La terminologie orthophonique et la langue des orthophonistes sont constituées tout d'abord des mots de la langue française, organisés selon les modalités linguistiques, phraséologiques et discursives du français, soumis aux contraintes intra et extralinguistiques habituelles. Il faudra également y ajouter des termes de spécialité, et une organisation collocationnelle spécifique au domaine pour ce qui concerne le « jargon technique ». Cette langue orthophonique particulière comprend à la fois des termes du français courant mais également ceux d'une langue scientifique (Baylon et Mignot 2005 : 330), comme des termes affixés (« dysphonie »), des composés (« anosognosie ») ou

des collocations[1] (« retard de parole »). La méthodologie de l'étude a consisté à isoler au cœur de cette langue orthophonique les termes diagnostiques ayant servi à l'orthophoniste pour étiqueter les troubles ou difficultés mis en évidence au cours du bilan orthophonique. Ces termes peuvent être regroupés selon leur morphologie et également selon leur usage en lien avec une pathologie spécifique. Ils sont parfois apparentés à des termes non spécifiques (« retard, difficultés ») mais ont été traités au même titre que ceux ayant une connotation plus directe à la pathologie (« dysphonie, bégaiement »).

Dans cet article nous expliquerons en quoi le « terme diagnostique » (*diagnostic labels* chez Plug *et al.* 2009) nous paraît particulièrement intéressant dans cette démarche d'exploration des représentations des pathologies chez le praticien en précisant la notion de diagnostic. Puis nous présenterons le corpus étudié. Les résultats les plus remarquables seront exposés afin d'introduire la considération termino-ontologique qui suivra.

2. Intérêt du « terme diagnostique » en orthophonie

Les prémices de l'exercice de l'orthophonie ont servi de creuset à une démarche clinique rigoureuse se fondant sur la mise au point d'hypothèses sur le fonctionnement cognitif et linguistique du patient, sous tendues par les conceptions contemporaines de la « bonne santé » linguistique et communicationnelle. Le bilan orthophonique est la première étape permettant une rencontre entre un(e) orthophoniste et son patient (Kremer & Lederlé 2009). Dans cette démarche, la phase diagnostique représente une étape indispensable mettant en correspondance les pathologies et les termes pour décrire la réalité du patient.

Le Trésor de la Langue Française informatisé TLFi propose une définition médicale du diagnostic[2] ouverte sur un usage dans la langue courante. Le dia-

1 **Collocation :** de façon générale, une collocation représente l'apparition consécutive et fréquente de deux mots (Halliday, 1962, Dubois *et al.,* 2007, Charaudeau et Mainguenau, 2002). Polguère (2003) en donne une définition plus spécifique en évoquant une combinaison syntagmatique (expression semi-idiomatique), modifiable par choix sémantique, dans laquelle le sens originel de la base est modifié.

2 « MÉD., usuel. Art d'identifier une maladie d'après ses signes, ses symptômes. Par analogie : Conclusion, généralement prospective, faisant suite à l'examen analytique d'une situation souvent jugée critique ou complexe. (Quasi-) synon. jugement » TLFi *(*http://atilf.atilf.fr/)

gnostic médical fait suite à un processus d'identification et produit une forme de prédiction quant à la suite des événements.

Au XX^{e} siècle, plusieurs professions paramédicales se sont vues reconnaître la capacité à établir un diagnostic ; on parle à présent de diagnostic infirmier, de diagnostic kinésithérapique, et depuis 2002 de diagnostic orthophonique[3]. Cette reconnaissance permet d'offrir une place essentielle au praticien paramédical dans le système de santé, face à ses pairs et face au patient, en démontrant la responsabilité du professionnel dans le système de soins et face à une personne affectée dans sa communication et/ou ses fonctions d'alimentation. La réflexion sur les terminologies et les classifications professionnelles mérite toute notre attention pour constituer des bases intégratives. Celles-ci permettent d'élaborer une terminologie fiable et d'améliorer la communication interlangue et la fiabilité des démarches d'identification des pathologies, comme le montre le projet de création de l'*International Classification For Nursing Practice* (ICNP®) (Nytun & Fossum 2014).

Le diagnostic orthophonique ne se résume donc pas à une simple conclusion fondée sur des mesures quantitatives évaluant un fonctionnement ou l'intégrité d'un système, mais procède bien d'un processus d'identification, d'interprétation de signes menant à la labellisation d'une affection et attribuant à une personne un statut pathologique.

Cette étape diagnostique peut se concevoir en deux temps. Le premier temps pourrait être considéré comme l'aboutissement d'un processus le plus souvent hypothéticodéductif, comme c'est le cas dans bon nombre de processus de recherches sociales et linguistiques. L'orthophoniste est à même de décrire et de comprendre, dans une dynamique d'observation et de recherche de conclusions, la façon dont « fonctionne » son patient. Il peut distinguer les performances mais aussi les procédures qui les sous-tendent et les stratégies éventuelles mises en place pour contourner les difficultés. Dans un deuxième temps, qui reste souvent optionnel et parfois implicite, le diagnostic est également l'issue d'un processus onomasiologique à part entière, dépendant de la terminologie (Dubois *et al.* 2007 : 481) et qui induit la mise en place d'une démarche thérapeutique répondant à des contraintes individuelles et sociales. Le choix terminologique permet de sélectionner les étiquettes correspondant au diagnostic posé sur la condition du patient, appelées *diagnostic labels* par Plug *et al.* (2010). Dans leur article, ces auteurs ont étudié les différences d'éti-

Décret n°2002-721 du 2 mai 2002 relatif aux actes professionnels et à l'exercice de la profession d'orthophoniste, modifié par le décret 2004-802

quetage, par le patient et par le médecin, d'une crise convulsive (qu'elle soit ou non causée par de l'épilepsie). Ainsi ils distinguent l'utilisation des étiquettes suivantes : *fit, attack*, en comparant leurs occurrences et leurs significations chez le patient et chez le médecin. Ils démontrent que cet étiquetage va bien au-delà d'une simple dénomination, et représente la validation d'un état de santé ou d'une pathologie au sein d'un dispositif d'énonciation bien particulier qu'est la consultation. Cette démarche diagnostique s'enrichit par la dimension supplémentaire de l'observateur sur sa propre pratique.

Nous parlons donc ici d'un praticien réflexif et d'un processus dynamique de labellisation des pathologies observables chez un individu.

Il existe ainsi dans la phase diagnostique une position épistémique de l'orthophoniste qui sélectionne un terme puis écrit son compte rendu. Le professionnel utilise une modalité subjective dans sa rédaction ; il décrit ce qu'il pense de la personne qui est venue le consulter, et des résultats de l'entretien et de l'évaluation. Il adapte donc son discours en jouant sur les modalisations syntactico-sémantiques dans le choix terminologique auquel il est confronté dans une réalisation linguistique contextualisée et située sur un plan diachronique et synchronique. Dans cet article nous nous intéressons à la langue orthophonique du début du XXI^{e} siècle.

Les variations dans la terminologie orthophonique (Walsh 2005) sont donc le résultat de :

- L'évolution de la conception de la bonne santé et des pathologies du langage
- L'évolution des classifications et des nomenclatures
- La démarche onomasiologique personnelle et collective du praticien, qui recherche dans les classifications et ses représentations les termes correspondant le mieux à sa perception des troubles de la personne qui consulte.

Dans ce sens, l'étude de la terminologie orthophonique pourra alimenter un modèle conceptuel et une classification originale des troubles du langage, éclairant les concepts à modéliser en fonction des réalisations linguistiques des professionnels eux-mêmes. Il s'agira alors d'une démarche termino-ontologique qui nous préoccupe depuis la première édition du Dictionnaire d'orthophonie (Brin-Henry *et al*. 1997, 2004, 2011) et notre thèse (2011).

2.1. Vers une étude de la terminologie orthophonique

Étudier les termes diagnostiques en orthophonie dans cette perspective nous pousse à nous intéresser aux variations terminologiques intervenant au cours de l'étiquetage des troubles du langage par le professionnel à l'issue de l'étape d'évaluation des troubles de son interlocuteur.

Nous cherchons donc à faire se rejoindre une analyse de la démarche onomasiologique des orthophonistes au cours de la phase diagnostique, avec une démarche descriptive et sémasiologique (Gaiffe *et al.* 2009) permettant d'extraire les diverses variations des termes à partir des comptes rendus de bilan. Il s'agit pour nous d'un double enjeu appliqué et fondamental.

- L'enjeu professionnel est épistémique et terminologique, visant à établir la pratique orthophonique et les réalisations linguistiques des praticiens dans leur langue de spécialité comme la fondation d'une réflexion conceptuelle originale.
- L'enjeu fondamental est descriptif et termino-ontologique vise à établir la frontière entre langue de spécialité et langue naturelle par l'examen syntactico-sémantique des unités linguistiques saillantes.

Les deux objectifs principaux de ce projet sont :

1. L'observation et l'annotation des candidats-termes en texte intégral pour étudier la phase d'étiquetage à partir de critères linguistiques et non pas à partir de critères issus de considérations biomédicales ou théoriques, et une étude du comportement linguistique des occurrences fréquentes et du format des termes diagnostiques : analyse sémantico-syntaxique sans standardisation
2. L'élaboration d'un système conceptuel pour alimenter une classification orthophonique fondée sur la pratique clinique (COFOP).

2.2. Étude terminologique sur un corpus de comptes rendus de bilan orthophonique

Le temps rédactionnel (Charaudeau et Maingueneau 2002) auquel se consacrent les praticiens (par convention avec l'assurance maladie) est concrétisé par les comptes rendus de bilan orthophonique. Ce passage à l'écrit révèle le processus réflexif du praticien.

La constitution d'un corpus de 436 comptes rendus de bilan orthophonique (CRBO) a été possible grâce à la coopération de collègues exerçant en France

métropolitaine et dans les départements et territoires d'Outre-mer, qui ont envoyé leur réponse entre le 20 février et le 6 mai 2007.

Le travail d'extraction et d'analyse s'est fait en plusieurs temps comme le montre le tableau ci-dessous :

Étapes manuelles de traitement du corpus	Anonymisation des comptes rendus
	Prise d'indices et de filtres[4]
	Sélection et constitution des métadonnées
Étapes semi-automatiques	Utilisation de concordanciers : AntConc, Lexico3
	Codage XML projet CLARIN-DATIST (http://www.clarin.eu/external/)
	Analyses quantitatives et qualitatives (AFC) avec R® sur 1 puis 2 variables
	Extraction, analyse formelle et classement des candidats-termes et de leurs propriétés (caractères)

Tableau 1 : Étapes de traitement des données

2.2.1. Étapes manuelles

La première étape a été une phase d'anonymisation. En effet il était important que le secret médical qui avait permis le recueil des CRBO soit réellement préservé tout au long du traitement des données. Ce corpus informatisé est à l'heure actuelle protégé et a été rendu accessible uniquement par une identification sécurisée.

La mise au format des textes a permis de les préparer à l'utilisation complémentaire semi-automatique de logiciels.

Des métadonnées ont été extraites et classées (informations sur les textes eux-mêmes, les âges des patients, les tests utilisés, les lieux d'exercice des orthophonistes, etc.) pour établir une analyse descriptive du corpus.

Chacun des CRBO a ensuite été annoté pour faciliter l'identification des termes diagnostiques et des chapitres des CRBO (anamnèse, bilan, diagnostic orthophonique, plan thérapeutique, examens complémentaires).

4 phase d'introspection décrite par Condamines (2007)

2.2.2. Traitement semi-automatique

Des concordanciers (Lexico3, AntConc) ont été utilisés pour extraire des cooccurrences (liste des termes fréquents, dans quelle proximité textuelle), afin d'obtenir une photographie des termes utilisés dans le corpus de CRBO.

Un projet collaboratif (CLARIN-DATIST)[5] a permis la mise en place d'un système de balisage des termes (codage XML) et d'une interface permettant l'affichage À l'écran de l'ordinateur à la fois des métadonnées et des CRBO comme le montre l'exemple ci-dessous (certaines données ont été effacées).

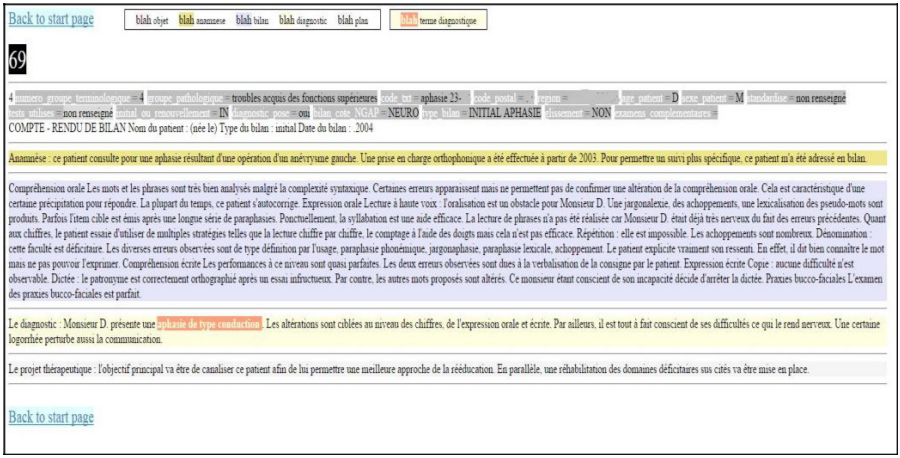


Figure 1 : Capture d'écran illustrant l'interface d'affichage des CRBO

Cet affichage permet également de vérifier les informations produites et de situer en contexte les termes diagnostiques identifiés.

2.2.3. Présentation résumée des métadonnées

L'ensemble des CRBO constitue un corpus de textes comprenant certaines données permettant de décrire une pratique orthophonique du bilan et de la phase diagnostique. L'examen des métadonnées du corpus fournit les résultats suivants :

5 http://www.clarin.eu/external

- 98,6% des orthophonistes ayant participé à cette étude sont des femmes. Les deux régions les plus représentées sont l'Île de France et la région Rhône-Alpes.
- Les groupes pathologiques évoqués dans les bilans montrent une répartition relativement classique d'après les données évoquées par L. Tain dans son ouvrage de 2007, soit environ 58 % du temps de travail hebdomadaire consacrés aux troubles du langage oral et écrit.
- Les groupes d'âge des patients sont répartis entre 1,4 % d'enfants de 0 à 3 ans et 10,1 % pour les plus de 65 ans. La majorité des patients ont entre 3 et 12 ans. Le nombre croissant de personnes de plus de 65 ans bénéficiant de rééducation orthophonique met en exergue les besoins dus à l'évolution de la population générale vieillissante.
- La répartition par genre de la population montre une légère disproportion puisque 53 % des patients sont de sexe masculin et 46 % de sexe féminin. Cela reflète la réalité actuelle en matière de prévalence des troubles du langage et de la communication (notamment la dysphasie, l'autisme), comme le montrent par exemple Choudhury et Benasich en 2003.
- Dans 69,8 % des textes du corpus est mentionnée une utilisation explicite d'évaluation standardisée pour effectuer le bilan.

3. Résultats

Cet article reprend les résultats les plus saillants concernant trois points principaux :

- Fréquence remarquable de termes non spécifiques : le cas de 'difficulté(s)'.
- Particularités sémantico-syntaxiques des mots appartenant aux textes orthophoniques et notamment la scission entre l'utilisation dans les CRBO de syntagmes faisant référence à une atteinte d'un système langagier (« dysphasie, troubles spécifiques du langage oral, déglutition atypique ») ou à une atteinte du fonctionnement de ce même système (« troubles de la compréhension du langage élaboré »).
- Nous terminerons la présentation de ces résultats par une évocation des propriétés des termes diagnostiques extraits.

3.1. Fréquence remarquable du terme « difficulté(s) » dans le corpus

La terminologie explorée montre l'occurrence fréquente du nom « difficulté(s) » dans la distribution du corpus orthophonique étudié (concordances). Cette fréquence importante d'un terme utilisé par ailleurs hors de la notion de pathologie montre le besoin d'une formulation par une unité lexicale complexe dans ce contexte.

	Nombre d'occurrences	**Rang (sur 12607)**
Difficultés	1004	39e
Difficulté (*pas de difficulté)*	226	140e

Tableau 2 : Fréquence d'apparition de "difficulté(s)" dans le corpus

Ce terme est associé à des adjectifs, des quantificateurs (notamment d'intensité), et des prépositions particulières (à+N, en+N, de+N, à+V, pour+V). De telles observations montrent les variations des usages chez les professionnels, qui corrompent parfois ainsi des syntagmes plus habituels dans les classifications non orthophoniques citées plus haut. Nous nous intéressons actuellement à la signification de ces variations pour la représentation de la « difficulté » chez l'orthophoniste (Condamines 2005) et pour la langue de spécialité elle-même (Kjaër 2005).

Nous avons procédé à une extraction par expressions régulières (contexte qui suit) par le concordancier AntConc. Les occurrences ont ensuite été triées selon le modèle suivant (résumé dans le tableau ci-dessous) : classe syntaxique, valeur, domaines, propriétés sémantiques. Nous avons observé les adjectifs taxinomiques et qualificatifs, les syntagmes prépositionnels (SP), associés à des noms seuls (N), à des déterminants et des noms (Dét+N) ou à des verbes (V).

Difficultés	+ Adj TAXINOMIQUE[6]		linguistique cognitif physique moteur socio-relationnel
	+ Adj QUAL		intensité incidence durée absence visibilité comparaison spécificité partie/tout
	SP	SP de+N SP à+V SP en SP dans SP pour V SP à+Dét+Nom	temporel
			topographique

Tableau 3 : Types d'occurrences du terme "difficulté(s)"

3.1.1. Occurrences de difficultés + adjectifs taxinomiques

Le tableau ci-dessous montre que les occurrences de « difficultés » dans le corpus permettent d'évoquer des classes taxinomiques diverses induisant la cause ou le contexte dans lequel la difficulté va s'exprimer chez le consultant (Roché 2006). Il ne s'agit pas de considérations étiologiques sur l'origine du trouble, mais bien de la façon dont l'orthophoniste aura qualifié l'expression des troubles de son patient.

6 Au sens de Cruse (1986)

Classe	Nombre	Exemple
Cognitif	53	*difficultés attentionnelles*
Linguistique	44	*difficultés lexicales, pragmatiques*
Socio-relationnel	19	*difficultés familiales, scolaires*
Physique	13	*difficultés ORL, pulmonaires*
Moteur	8	*difficultés graphiques*

Tableau 4 : Difficultés + Adj.Tax (137 occurrences)

Les difficultés cognitives et linguistiques sont les plus fréquemment mentionnées dans le corpus.

3.1.2. Occurrences de difficultés + type de qualification

Le tableau ci-dessous présente la répartition des adjectifs qualificatifs associés au terme « difficultés ». Dans la plupart des cas l'orthophoniste aura utilisé le terme dans sa forme au pluriel. Cependant dans un cas unique (lié justement à l'absence de difficulté), ce terme est utilisé dans sa forme au singulier. L'intensité est la fonction la plus souvent représentée.

Type de Qual	Nombre	Exemple
Intensité	31	*des difficultés majeures*
Incidence	12	*des difficultés rencontrées, présentes*
Durée	10	*des difficultés persistantes*
[Absence	8	*aucune difficulté]*
Visibilité	6	*des difficultés évidentes, manifestes*
Comparaison	6	*des difficultés semblables*
Spécificité	3	*des difficultés particulières*
Partie/tout	2	*des difficultés plus globales*

Tableau 5 :Difficultés + Adj Qual (78 occurrences)

3.1.3. Occurrences de difficultés + valeur topographique et temporelle

Le tableau ci-dessous présente la répartition des adjectifs qualificatifs associés au terme « difficultés » dans une perspective topographique ou temporelle. En effet l'exemple de difficultés « dans » nous permet d'évoquer soit une difficulté dans un système d'acquisition (ici la notion numérique) qui se déroule au cours du temps, soit dans un processus quasi instantané mais sectorisé (ici le fait de convertir un son en graphème).

	Nombre	**Exemple**
SP dans	27	*difficultés dans la construction du nombre*
SP au niveau	12	*difficultés au niveau de la parole*
SP avec	10	*difficultés avec les graphies complexes*
SP sur le plan	5	*difficultés sur le plan du langage écrit*
SP concernant	3	*difficultés concernant l'orthographe*

Tableau 6 : SP+ Valeur topographique (66 occurrences)

	Nombre	**Exemple**
SP dans	18	*difficultés dans la conversion phonographémique*
SP lors de	6	*difficultés lors de la prise d'indices visuels*
SP au cours de	1	*difficultés au cours du premier trimestre*

Tableau 7 : SP+ Valeur temporelle (33 occurrences)

3.1.4. Occurrences de difficultés + syntagme prépositionnel

Le tableau ci-dessous présente la répartition des syntagmes prépositionnels dépendant du nom « difficultés ». Le regroupement a été effectué en fonction des prépositions présentes (Sarda 2010). Dans la majorité des cas (394 occurrences), l'orthophoniste aura utilisé la préposition « de », ce qui est un comportement classique dans la langue naturelle commune, comme dans la langue orthophonique.

Classe	Nombre	Exemple
SP de+N	394	*difficultés de compréhension, de lecture, de flexibilité*
SP à+V	87	*difficultés à retenir les mots, à traiter les informations, à prononcer*
SP en	63	*difficultés en lecture, en mathématiques*[7]
SP dans	45	*difficultés dans l'élaboration d'un raisonnement, dans l'espace*
SP pour V	36	*difficultés pour fixer les règles, pour nettoyer la bouche, pour parler*
SP à+Det+Nom	15	*difficultés à l'endormissement, à l'écriture*[8]

Tableau 8 : Difficultés + SP (739 occurrences)

La poursuite de ce travail doit permettre d'évaluer la différence de signification entre les diverses variations. Les questions posées pourront porter sur la signification des variantes entre les constructions verbales ou nominales (comme « difficultés de compréhension », et « difficultés à comprendre ») notamment dans une perspective de description des noms et des verbes d'activité (par exemple la notion de stativité (Heyd & Knittel 2009)). Même si le linguiste considère classiquement que cette variation n'est pas significative sur un plan syntaxique, nous pensons qu'elle est significative dans la langue de spécialité, montrant une conception plus active de la difficulté dans les représentations des orthophonistes dans le cas de l'utilisation de la forme verbale.

3.2. Répartition entre les deux catégories Forme/Nature de termes diagnostiques

Le format de l'étude a permis de dégager des typologies formelles pour les termes diagnostiques utilisés dans les CRBO par les praticiens. Le tableau ci-dessous propose de distinguer les variantes morphologiques utilisées, qui peuvent également se regrouper dans deux catégories exclusives de termes.

7 Appartenant donc plutôt à un domaine scolaire
8 Appartenant donc plutôt à un domaine scolaire ou physique

termes diagnostiques		**NATURE : Atteinte du fonctionnement**	**FORME : Atteinte du système**
	Syntagme collocationnel ou locution	*troubles spécifiques du langage écrit*	*raucité infantile, retard de parole*
	Affixés	*dysfonctionnement (exécutif de l'articulation temporo-mandibulaire)*	*bégaiement dysarthrie, dysphonie*
	Composés	*(trouble du raisonnement) logico-mathématiques*	*anosognosie*

Tableau 9 - Typologie des termes

Il existe selon nous une distinction franche, et une utilisation exclusive de l'une ou l'autre des deux catégories de termes décrivant la pathologie (et qui serviraient de fondement à la base de la construction d'une nouvelle classification orthophonique):

- D'une part des termes faisant référence à l'atteinte du fonctionnement d'un système lié à la communication humaine (ex : « retard massif de développement du langage oral, difficultés importantes dans l'accès au langage écrit »), appartenant à une catégorie que j'ai nommée 'Nature'.
- D'autre part des termes (collocations voire locutions) faisant état de l'atteinte du système responsable du fonctionnement humain (ex : « bégaiement, raucité infantile, dyslexie mixte sévère »), appartenant à la catégorie 'Forme' nommée ainsi dans ma thèse.

Le tableau ci-dessous présente brièvement la distinction faite entre les termes des deux catégories, au travers de certaines connotations caractéristiques.

	NATURE	**FORME**
Relation à la conception	Processus	Système
Capacités et compétences, habiletés	Métalinguistiques et méta-cognitives	Linguistiques et cognitives
Dynamique	Acquisition, évolution	Stabilité, précision
	Obstacle au processus et son accomplissement	Déviance, Perturbation d'un système
Ancrage causal	Externe	Interne
Exemples	*troubles de la compréhension des énoncés, déficit de gestion des informations complexes*	*bégaiement, trouble d'articulation, retard de langage, dyslexie*

Tableau 10 : termes diagnostiques issus des catégories « nature » et « forme »

Cette répartition révèle la distinction entre les termes montrant une conception fonctionnelle du trouble d'une part et une conception structurelle du trouble d'autre part, s'éloignant ainsi d'une analyse du fonctionnement cognitif humain pour alimenter un raisonnement linguistique décrivant la démarche onomasiologique des orthophonistes.

Les termes diagnostiques du groupe « NATURE » relèvent d'une perception de processus, liée à une mécanique de fonctionnement. L'aspect dynamique de la pathologie, son impact, sa persistance et sa durabilité, l'importance (quantifiée) en sont les critères principaux. Le recours à ce type de terme (par exemple « troubles de l'acquisition du langage écrit », « déficit dans la compréhension d'énoncés complexes »), montre une approche des compétences (structuration, organisation, gestion, manipulation, programmation) partant vers le métalinguistique et le métacognitif (habiletés discursives, métaphonologiques) donc favorisant plutôt une analyse individuelle des compétences du patient dans un contexte et une interaction spécifique avec son environnement.

Pour ce qui concerne les termes diagnostiques du groupe « FORME », qui découlent de la perception d'une perturbation d'un système, ils sont délimités et plus facilement qualifiés et associés les uns aux autres. Ces termes diagnos-

tiques (par exemple « dysphonie », « trouble d'articulation », « bégaiement », « dyslexie », « retard de langage ») servent à spécifier l'étiologie et les contours de la pathologie, et donc plutôt ce qu'il en est (aspects formels, mécanicistes). Les termes sont plus stables de façon générale, et le recours à ces étiquettes se conçoit comme une approche des capacités et des habiletés, dans les fonctions motrices cognitives et linguistiques (mémoire, langage écrit, conscience phonologique…), dont un substrat quasi anatomique, objectivement évaluable.

3.3. Propriétés sémantiques des termes.

L'étude des termes diagnostiques permet d'évoquer les propriétés qui révèlent la perception des troubles chez l'orthophoniste et sur lesquelles nous nous appuierons dans la construction de la COFOP :

- La stabilité et précision justifiant l'apparition de variantes synchroniques comme « dyslexie /alexie », « dysphasie/aphasie »
- La durabilité justifiant le caractère synchronique des termes, comme « aphasie » vs « alalie » de Lordat (1820)
- La flexibilité et la diversité permettant la modalisation et l'expression individuelle du diagnostic
- La temporalité ou dynamique :

Du terme (qui s'expose dans l'évolution synchronique comme dans « dysphonie dysfonctionnelle » vs « dysphonie fonctionnelle »)

Du trouble (« séquelles », « retard global ») qui exprime l'évolution des performances et des difficultés, de la pathologie du patient

- La déviance par rapport à une norme, mise en évidence notamment grâce au préfixe dys, altération, atteinte, troubles… (notamment dans les termes appartenant à la catégorie 'Nature').

4. Conclusion

Si l'on considère le continuum entre normal et pathologique dans la maîtrise du langage et du fonctionnement de la sphère orofaciale (au sens large de ces termes) chez les patients consultant un(e) orthophoniste, l'étape diagnostique représente au travers du choix terminologique un marqueur essentiel de la démarche thérapeutique, une clé de voûte de la démarche méthodologique orthophonique.

Pour l'orthophoniste, il s'agit de la détermination d'un processus d'identification des troubles et de mise en œuvre d'une rééducation, dans une démarche rigoureuse, dynamique et évolutive de l'exercice de la profession. Le recours aux termes diagnostiques est une étape de labellisation importante dans le cadre du bilan orthophonique et guide le praticien dans sa relation à l'autre (patient, professionnel médical ou paramédical, enseignant, entourage…) grâce à sa maîtrise des concepts mais aussi des termes de la discipline.

Pour le patient et son entourage, l'étape diagnostique, par cet étiquetage, représente la reconnaissance d'un état « pathologique » de son langage et de la sphère orofaciale. Comme le montrait déjà Canguilhem (1966), le passage d'un état ordinaire à un état pathologique relève non seulement d'une identification médicale ou paramédicale conventionnelle d'un fonctionnement hors de l'ordinaire ou de l'intégrité d'une structure, mais aussi de la conception elle-même de ce qui est pathologique (Delahaie 2009, Tran 2001). Le diagnostic orthophonique est un point important du bilan dans la mesure où cette étape met en jeu l'ensemble des paramètres contribuant à l'établissement de la relation orthophoniste/patient. Elle est le moment où les troubles sont nommés, identifiés, grâce à l'étape préalable de sélection des indices pertinents, et à la suite de laquelle s'engage la phase où le patient, informé de la démarche thérapeutique à sa disposition, pourra participer à la relation de soin orthophonique à proprement parler.

Cette étude a permis de dégager des tendances pour les termes diagnostiques. Elle montre que les correspondances entre le terme diagnostique, la réalité des troubles et la conception de la pathologie elle-même ne sont pas si univoques que l'on pourrait le croire. L'angle choisi a montré que la poursuite d'une réflexion approfondie sur la nature et la forme des troubles intéressant les orthophonistes peut s'appuyer sur une analyse terminologique de corpus de termes diagnostiques et de comptes rendus de bilan orthophonique. Pour conclure, nous pensons que la prise de conscience de ces phénomènes doit encourager le praticien à faire reconnaître la place de ce diagnostic, en prenant confiance et affirmant la légitimité de cette démarche spécifique.

La démarche terminologique utilisée ici permet de révéler les représentations des orthophonistes sur les pathologies objectivées par leur sélection du terme, et en raison de la scission atteinte du système/du fonctionnement (Nature/Forme). De plus, elle renseigne sur les spécificités de la langue orthophonique (typologies, propriétés). La construction d'une nouvelle classification orthophonique fondée sur la pratique clinique (COFOP)

en est à sa première étape. L'établissement de ces deux types de termes, y compris leurs propriétés sémantiques, leur construction morphologique et la publication des résultats des analyses sémantico-syntaxiques donne naissance à d'autres études visant à valider la perception de cette phase d'étiquetage que représente le diagnostic orthophonique. Elle pousse à structurer une classification des troubles du langage fondée sur une identification approfondie de la langue de spécialité, liée à des phénomènes de réalisation dans une langue et un contexte donnés.

Les intérêts multiples de ce type d'approche sont à présent reconnus dans le domaine médical et paramédical. Les usages recouvrent par exemple des intérêts concernant la qualité des soins (Nytun 2014), la facilitation de l'alignement des terminologies dans des perspectives de soins et de recherche (Kahn *et al.* 2014, Ma *et al.* 2014).

À long terme, nous travaillons à l'émergence d'une terminologie de langue de spécialité (l'orthophonie) à partir d'une base intégrative d'envergure européenne. Ce projet requiert la mise en place d'un alignement multilingue sur les termes (hors contexte), et l'organisation et la mise à disposition de ressources linguistiques pour le chercheur et le praticien-chercheur. L'objectif en est de :

- fournir des données enrichies à la communauté scientifique et professionnelle tout en validant une démarche complexe de mise en œuvre d'une plateforme de données protégées sur le plan juridique et éthique (anonymisation, formatage, annotation manuelle, analyse statistique et export de données, concordances)
- permettre l'observation du comportement linguistique particulier de certains termes fréquents

L'originalité de la démarche se situe dans l'étude des usages à partir de l'extraction des termes à partir de textes authentiques écrits par des professionnels, représentant la pratique clinique située sur le plan diachronique et synchronique.

Acknowlegments

Nous remercions Évelyne Jacquey et Marie Laurence Knittel de l'ATILF pour leurs conseils et relecture précieuse de cet article.

Références bibliographiques

Baylon, C. & Mignot, X. (2005). *La Communication*. Paris : Armand Colin.

Brin-Henry, F. (2014). Using corpus-based analyses in specialised paramedical French. *Revue Française de Linguistique Appliquée : Langues de spécialité : problèmes et méthodes* Vol. XIX, n°1, p.103-15.

Brin-Henry, F. (2011). *La terminologie crée-t-elle la pathologie ? Le cas de la pratique clinique de la pose du diagnostic orthophonique*. Thèse pour l'obtention du Doctorat en Sciences du Langage, Université de Lorraine. Disponible en ligne sur http://tel.archives-ouvertes.fr/tel-00655952/

Brin-Henry, F.,Courrier, C., Lederle, E., Masy V. (2011). (1re édition 1997). *Dictionnaire d'orthophonie*. Isbergues : Ortho-Edition.

Cabré, M-T. (2012). *Disciplinarisation de la terminologie : contribution de la linguistique*.

Dossiers d'HEL 2012. SHESL [en ligne]. [Consulté le 12 septembre 2014]. Disponible sur http://htl.linguist.univ-paris-diderot.fr/num5/articles/Cabr%C3%A9_2012.pdf

Canguilhem, G., (1966) (11e édition quadrige 2009). *Le normal et le pathologique*. Paris : Presses Universitaires de France.

Charaudeau P. & Maingueneau D., (2002). *Dictionnaire d'analyse du discours*, Paris : Seuil.

Choudhury N. & Benasich A. (2003). A family aggregation study : the influence of family history and other risk factors on language development. *Journal of Speech, Language, and Hearing Research*, Vol. 46, n°2, p.261-72.

Condamines A., (2005). Linguistique de corpus et terminologie. *Langages*, n° 57, p.36-47.

Condamines, A. (2007). L'interprétation en sémantique de corpus : le cas de la construction de terminologies. *Revue Française de Linguistique Appliquée*, Vol XII, n°I, p.39-52.

Cruse, D.A. (1986). *Lexical Semantics*. Cambridge University Press

Delahaie, M. (2009). *L'évolution du langage de l'enfant. De la difficulté au trouble. Guide ressources pour les professionnels*. Saint-Denis : Inpes.

Dubois J., Giacomo M., Guespin L., Marcellesi J.-N., Mevel J-P. (2007). *Grand dictionnaire Linguistique et Sciences du Langage*. Paris : Larousse.

Fougeyrollas, P., Cloutier, R., Bergeron, H., Côté, J. & Michel, G. (1998). Classification québécoise Processus de production du handicap, Québec, *Réseau international sur le Processus de production du handicap (RIPPH)/SCCIDIH*.

Gaiffe, B., Jacquey, E., Kister, L. (2009). Approche lexico-sémantique de l'extraction terminologique : utilisation de ressources lexicographiques et validation sur corpus. *TOTh 2009. Actes de la troisième conférence TOTh* - Annecy – 4 & 5 juin 2009, p. 19-32.

Heyd, S. & Knittel, M-L. (2009). Les noms d'activité parmi les noms abstraits ; propriétés aspectuelles, distributionnelles et interprétatives. Linguisticae Investigationes Vol. 32 n°1, p.124-148.

International Council of Nurses (ICN). International Classification For Nursing Practice (ICNP®). [en ligne]. [Consulté le 28 août 2014]. Disponible sur http://www.icn.ch/fr/piliers-et-programmes/international-classification-for-nursing-practice-icnpr/

Kahn, MG., Bailey, LC., Forrest, CB., Padula, MA., Hirschfeld, S. (2014). Building a common pediatric research terminology for accelerating child health research. *Pediatrics*, Vol 133, n°3, p. 516-25.

Kjaër, B.E. (2005). Terminology and conception of the profession. *Advances in Speech-Language Pathology*, Vol. 7, n°2, 115-17.

Kremer J.-M., Lederle E., (2009) (1re édition 1991). *L'orthophonie en France*. Paris :

Presses Universitaires de France (Que sais-je, n°2571).

L'Homme, M-C. (2004) : *La terminologie : Principes et Techniques*. Presses Universitaires de Montréal

Ma. L., Liu, B., Xie, Q., Mao, S., Cui, Z. (2014). Ontological reconstruction of the clinical terminology of traditional Chinese medicine. *Front Med*. (ahead of print).

Morin, E. (1994). Sur l'interdisciplinarité. *Bulletin Interactif du Centre International de Recherches et Études transdisciplinaires* n° 2 - Juin 1994. [en ligne]. [consulté le 15 avril 2014]/ Disponible sur http://perso.clubinternet.fr/nicol/ciret/bulletin/b2c2.htm

Nytun, JP. & Fossum, M. (2014). Information Model for Learning Nursing Terminology. *E-Health- for continuity of care*. European Federation for Medical Informatics and IOS Press, p. 181-5.

Plug L., Sharrack B., Reuber M. (2009). Seizure, Fit or Attack ? The use of diagnostic labels by patients with Epileptic or Non-Epileptic Seizures, *Applied Linguistics*, Vol 31, n°1, p.94-114.

Roche, C. (2005). Terminologie et ontologie. *Langages,* n° 157, p. 48-62

Roché, M. (2006) « Comment les adjectifs sont sémantiquement construits ». Cahiers de Grammaire n° 30 Spécial Anniversaire, p. 373-87.

Sarda, L. (2010) Les adverbiaux prépositionnels en dans : exploration en corpus de la notion de contenance CORELA, Espace, Préposition, Cognition.

[en ligne]. [Consulté le 15 avril 2014). Disponible sur http://corela.edel.univ-poitiers.fr/index.php?id=911

Tain L. (Dir.), 2007. *Le métier d'orthophoniste : langage, genre et profession*. Rennes : ENSP.

Tran TM., 2001. Les accidents dans le langage ordinaire et aphasique : du normal au pathologique. *Revue Française de linguistique appliquée*, Vol. I, n°1, p. 35-46.

Walsh, R. (2005). A response to eight views on terminology : is it possible to tame the wild beast of inconsistency ? *Advances in Speech-Language Pathology*, Vol. 7, n°2, p.105-11.

Summary

This article introduces a terminological study of 436 speech and language therapy (SLT) reports. 558 diagnostic labels were extracted in order to examine how SLT professionals consider and label the disorders they encounter in their patients and prepare to a more general reflexion on language disorders classification.

La terminologie *in statu nascendi* et ses implications au niveau sémantique. Exemple de la terminologie française et polonaise du génie biomédical

Kaja Gostkowska

kaja.gostkowska@gmail.com
http://www.ifr.uni.wroc.pl/fr/home.html

Résumé. Dans notre article, nous décrivons une partie de la terminologie biomédicale qui n'a pas fait l'objet jusqu'à présent d'études approfondies ni en français, ni en polonais. Et pourtant, le génie biomédical s'avère très propice à l'étude terminologique étant donné son caractère totalement nouveau comme domaine de connaissance, et son interdisciplinarité. Nous nous concentrons sur ce premier aspect de la terminologie biomédicale. L'objectif de cet article est d'étudier les phénomènes qui sous-tendent la création d'un champ de recherche nouveau. Il s'agit d'observer deux phénomènes linguistiques spécifiques d'une terminologie qui se crée sous nos yeux : dans un premier temps, la synonymie et, deuxièmement, la polysémie. De plus, l'étude contrastive des termes utilisés dans les communautés scientifiques française et polonaise permet de comparer deux langues de statut différent, leur rôle dans la science, ainsi que leur comportement vis-à-vis de l'anglais.

1. Introduction

Depuis plusieurs années, la terminologie en tant que domaine qui étudie les terminologies, c'est-à-dire des ensembles de termes utilisés dans des domaines précis, développe et adapte ses méthodes pour faire face à une multiplication sans précédent des termes dans la science contemporaine. Avec l'apparition de nouveaux secteurs de la recherche – dont le génie biomédical – de nouvelles perspectives s'ouvrent pour les études terminologiques. Dans cet article, nous décrivons une partie de la terminologie biomédicale qui n'a pas

fait l'objet jusqu'à présent d'études approfondies ni en français, ni en polonais. Il faut insister sur le fait que la terminologie biomédicale ce n'est pas la même chose que la terminologie médicale. Le génie biomédical s'occupe de la conception de toutes sortes de dispositifs médicaux (par exemple une endoprothèse, un cœur artificiel, une orthèse ou un implant dentaire). Les objectifs des études biomédicaux sont les mêmes que ceux de la médecine, mais ce qui distingue le génie biomédical c'est, tout d'abord, son caractère totalement nouveau comme domaine de connaissance. On peut dire sans trop exagérer que c'est un domaine qui se crée « sous nos yeux », de même que sa terminologie. En plus, c'est un domaine interdisciplinaire qui résulte de la coopération des ingénieurs et des médecins. Et cela a des implications importantes au niveau linguistique, et plus particulièrement terminologique (voir Gostkowska 2011). Vu donc son caractère totalement nouveau comme domaine de connaissance et son interdisciplinarité, le génie biomédical s'avère très propice à l'étude terminologique.

Dans notre article, nous nous concentrons sur ce premier aspect de la terminologie biomédicale. L'objectif que nous nous posons est d'étudier les phénomènes qui sous-tendent la création d'un champ de recherche nouveau. Il s'agit d'observer les traits linguistiques et sémantiques spécifiques d'une terminologie qui se crée sous nos yeux. De plus, l'étude contrastive des termes utilisés dans les communautés scientifiques française et polonaise permet de comparer deux langues de statut différent, leur rôle dans la science, ainsi que leur comportement vis-à-vis de l'anglais.

Comme le génie biomédical constitue un domaine de connaissance très vaste, nous avons choisi d'analyser les termes appartenant à deux de ses disciplines, à savoir la biomécanique et les biomatériaux. En ce qui concerne la biomécanique, elle focalise ses recherches sur le fonctionnement moteur du corps pour concevoir les dispositifs artificiels (les prothèses, les implants, etc.) pouvant remplacer les éléments défaillants de l'appareil moteur humain. Les biomatériaux en tant qu'une discipline scientifique s'occupent de la conception des biomatériaux, c'est-à-dire de tout matériau utilisé pour remplacer une partie ou une fonction du corps de manière sûre et fiable, acceptable d'un point de vue physiologique et… économique. Dans notre analyse, nous nous sommes bornée à étudier les termes appartenant aux deux micro-systèmes conceptuels très caractéristiques des disciplines en question : le micro-système des dispositifs orthopédiques pour la biomécanique, et celui des propriétés des biomatériaux pour les biomatériaux (en tant que discipline scientifique). Pour le faire, nous avons adopté une perspective onomasiologique pour

vérifier de quelle manière les concepts sont dénommés dans la langue. Nous nous concentrons donc, ici, sur les questions de la synonymie et la polysémie. Les résultats de cette analyse seront comparés dans les deux langues étudiées, à savoir le français et le polonais.

2. Synonymes dans la terminologie

Depuis un certain temps, l'existence de la synonymie dans les langues spécialisées, jadis fortement critiquée ou même niée, ne suscite plus de doutes ni de réticences. L'étude des terminologies biomédicales française et polonaise a confirmé que c'est un phénomène bien présent dans la communication entre les spécialistes. Et c'est le moins que l'on puisse dire. Le nombre de synonymes (et il faut souligner que, dans une grande partie, ce sont des synonymes parfaits, donc remplaçables dans tous les contextes) et leur complexité formelle inattendue sont tels qu'on les a divisés en quatre groupes différents. Nous distinguons donc quatre types de variantes : étymologiques, syntaxiques, de conceptualisation et des variantes en contexte. Le terme de *variante* que nous utilisons est emprunté à Daniel Gouadec.

2.1. Variantes étymologiques

Ce qui distingue les paires de termes réunies dans ce groupe c'est leur forme, et plus précisement l'étymologie des morphèmes qui les composent. Le plus souvent, il s'agit d'un terme emprunté et de son équivalent formel dans la langue d'arrivée. Dans les termes empruntés à une autre langue, les morphèmes viennent soit de l'anglais, soit du latin ou du grec. Pour ce groupe de variantes, le polonais est beaucoup plus riche en exemples. En outre, les trois derniers exemples polonais montrent que les équivalents des termes empruntés sont, souvent, plus complexes du point de vue formel. Les spécialistes essaient ainsi de « traduire » la siginification du terme emprunté et de la rendre par une expression descriptive.

FR	PL
- effets **carcino**gènes / effets **onco**géniques - **stress-shielding** / **bouclier anticontraint** - activité ostéo**génique** / biomatériau ostéo**formateur**	- bio**kompatybilność** / bio**zgodność** - **implant** / **wszczep** - **implant**acja / **wszczep**ianie - endoproteza **kondylarna** / endoproteza **kłykciowa** - **osteosynteza** / **zespolenie** - aparat **kompresyjno-dystrakcyjny** / aparat **rozciągowo-dociskowy** - klamra do **osteosyntezy** / klamra do **łączenia złamań kości** - strefa **stress-shieldingu** / **naprężeniowa strefa ochronna** - **remodeling** / **przebudowa tkanki kostnej**

Tableau 1 – Variantes étymologiques.

2.2. Variantes syntaxiques

Les termes de ce groupe sont composés des mêmes morphèmes, des mêmes unités sémantiques, mais celles-ci sont représentées par des parties du discours différentes. On a donc affaire aux constructions syntaxiques différentes dont le sens est identique. Dans les deux terminologies, les exemples abondent. Quant à la forme des termes on a pu réléver certaines constructions syntaxiques qui se répètent dans les deux langues. Le plus souvent, on a une paire dans laquelle l'un des termes ne se compose que d'un substantif (**N**), et l'autre est composé d'un substantif et d'un adjectif (**N + Adj**). Dans ce deuxième cas, c'est l'adjectif qui véhicule le sens du terme, le substantif qui l'accompagne n'étant qu'un terme général.

FR	PL
- **composite** / matériau **composite** - **ductilité** / matériau **ductile** - **bioactivité** / matériau **bioactif** - **biocompatibilité** / matériau **biocompatible** - **biodégradabilité** / matériau **biodégradable** - **biorésorbabilité** / matériau **biorésorbable** - **biotolérance** / matériau **biotoléré** - **porosité** / matériau **poreux**	- **przeciążenie** / zmiany **przeciążeniowe** - **kompozyt** / materiał **kompozytowy**, **polimer** / materiał **polimerowy** - **biokompatybilność** / materiał **biokompatybilny** - **bioresorbowalność** / materiał **bioresorbowalny** - **biotolerancja** / materiał **biotolerancyjny** / materiał **biotolerowalny** - **biozgodność** / materiał **biozgodny** - **porowatość** / materiał **porowaty**

Tableau 2 – *Variantes syntaxiques : N vs. N + Adj.*

Dans le deuxième groupe de synonymes syntaxiques, s'opposent, d'un côté, les termes composés d'un substantif et d'un adjectif (**N + Adj**) et, de l'autre côté, les termes dans lesquels le sens de l'adjectif est rendu par un groupe prépositionnel (**N + Prép + N**).

FR	PL
- phase **oscillante** / phase **d'oscillation** - prothèse **non cimentée** / prothèse **sans ciment** - fixation **prothétique** / fixation **de la prothèse** - fracture **spiroïde** / fracture **en spirale**	- trzpień **bezkołnierzowy** / trzpień **bez kołnierza**, trzpień **kołnierzowy** / trzpień **z kołnierzem** - odporność **korozyjna** / odporność **na korozję**

Tableau 3 – *Variantes syntaxiques : N + Adj vs. N + Prép + N.*

Les paires des synonymes suivantes n'apparaissent qu'en français : il s'agit des éléments morphologiques et sémantiques donnés qui peuvent apparaître

soit en tant qu'un terme générique soit comme un terme spécifique, déterminant le terme plus général (**N1 + Adj2** vs. **N2 + Adj1**).

FR	PL
- ligament prothétique / prothèse ligamentaire - genou prothétique / prothèse de genou - néoformation osseuse / os néoformé	-

*Tableau 4 – Variantes syntaxiques : **N1 + Adj2** vs. **N2 + Adj1**.*

Dans cette quatrième section, nous avons réuni les paires : terme complexe *vs*. son abréviation. Dans la terminologie biomédicale, ce type de paires syntaxiques renvoie principalement aux différents types de biomatériaux ou d'endoprothèses. Le plus grand nombre de tels exemples a été observé pour le français, surtout en ce qui concerne les types de prothèses. En polonais, on emploie uniquement les abréviations et les symboles des types de matériaux utilisés, par contre les abréviations dénotant les types de prothèses n'existent guère.

FR	PL
- enclouage centromédullaire / ECM	- fosforan trójwapniowy / ß-TCP
- prothèse avec conservation du **ligament croisé postérieur** / prothèse avec conservation du **LCP** - prothèse totale du genou / PTG - prothèse unicompartimentale / PUC - prothèse totale de hanche / PTH - alliages **chrome cobalt** / alliages **Cr-Co** - polyéthylène à haute densité / PEHD / UHMWPE - polyéthylène à très haute masse moléculaire / PEHM - polymétacrylate de méthyle / PMMA - hydroxyapatite / HA	- hydroksyapatyt / HA - stal chromowo-niklowo-molibdenowa / stal 316L - polietylen o bardzo wysokiej gęstości cząsteczkowej / UHMWPE - polietylen o wysokiej gęstości / HDPE - polimetakrylan metylu / PMMA

*Tableau 5 – Variantes syntaxiques : **terme complexe** vs. **abréviation**.*

2.3. Variantes de conceptualisation

Pour ce type de variantes les deux termes de chaque paire désignent le même concept mais leur motivation est différente. Autrement dit : les spécialistes utilisent à chaque fois un autre schéma de conceptualisation. Dans les deux langues, les exemples sont très nombreux. Nous ne présentons ici que des exemples choisis, les plus caractéristiques pour le domaine en question.

FR	PL
- appareil **de contention** externe / **orthèse** externe - implant **orthopédique** / prothèse **endosquelettique** - **implant de comblement** osseux / **substitut** osseux - implant d'**ostéosynthèse** / **orthèse** interne - **fixation** dynamique / **montage** dynamique - **articulation artificielle** / **prothèse articulaire** - **composant** fémoral / **implant** fémoral	- proteza **implantowana** / proteza **stawowa** / proteza **wewnętrzna** - aparat **rozciągowo-dociskowy** / aparat **stabilizujący** / **stabilizator zewnętrzny** - stabilizator **kombinowany** / stabilizator **przestrzenny** - **osteosynteza śródszpikowa** / **stabilizacja śródszpikowa** / **stabilizacja wewnętrzna** - **gwóźdź** Endera / **osteosynteza metodą** Endera - **alloplastyka** / **endoprotezoplastyka** - **częściowa** alloplastyka stawu biodrowego

/**pièce** fémorale - prothèse **avec sacrifice des deux ligaments** croisés / prothèse **à stabilisation postérieure** / prothèse **postéro-stabilisée** - tige **fémorale** / tige **prothétique** - ancrage **biologique** / ancrage **sans ciment** - os **porteur** / os **receveur** - **colonisation** osseuse / **réhabitation** osseuse - matériau **biologique** / matériau **prothétique** - **alliages chrome cobalt** / **stellites** - **comportements** biologiques / **propriétés** biologiques - **activité** ostéogénique / **effet** ostéogène / **propriétés** ostéogéniques - **compatibilité** tissulaire / **tolérance** tissulaire - **caractère** ostéoconducteur / **effet** ostéoconducteur	wego / **połowicza** protezoplastyka stawu biodrowego - **endoproteza** / **sztuczny staw** - **łożysko** ślizgowe / **płytka** ślizgowa - osteosynteza **neutralizująca** / osteosynteza **zabezpieczająca** - węzeł **ruchowy** / węzeł **trący** - cement **chirurgiczny** / cement **kostny** - moduł **sprężystości** / moduł **Younga** - **bio**materiał / materiał **implantacyjny** - bio**kompatybilność** / bio**tolerancja**

Tableau 6 – *Variantes de conceptualisation*.

Nous proposons de s'arrêter sur quelques exemples mentionnés ci-dessus pour démontrer leur spécificité. Dans la paire *articulation artificielle* vs. *prothèse articulaire*, les spécialistes accentuent deux aspects différents de la même prothèse : soit on souligne les conséquences de l'implantation, c'est-à-dire une articulation qui n'est plus naturelle, soit on met l'accent sur l'objectif d'une telle solution, à savoir une prothèse qui vise à remplacer une articulation (et non pas une autre partie du corps). Les termes *tige fémorale* et *tige prothétique* designent la même partie de la prothèse, sauf que le premier insiste sur l'entité anatomique qui doit être remplacée, le deuxième quant à lui signale que la tige (en tant qu'une partie de la hanche) n'est plus naturelle mais

qu'elle est artificielle. Deux points de vue différents sont également marqués dans la paire *os porteur* vs. *os receveur* : on parle de l'os qui porte ou qui reçoit l'implant. D'autres exemples, également en polonais, démontrent aussi des différences pertinentes dans la motivation. Ce qu'il faut souligner c'est le fait que les exemples des synonymes de conceptualisation dans les deux langues ne sont pas les mêmes, à quoi on aurait pu s'attendre vu le nombre de termes empruntés à l'anglais par les deux communautés de spécialistes. Les paires françaises décrites ci-dessus n'ont pas de paires équivalentes dans la terminologie polonaise et *vice versa*. En outre, d'après l'analyse des textes il nous semble que la fréquence d'emploi de chaque terme d'une paire est comparable. Tout ceci nous emmène à la constatation que malgré l'influence indéniable de l'anglais les deux langues spécialisées se développent chacune à sa façon. Nous allons y revenir encore dans la suite de notre propos.

2.4. Variantes en contexte

Dans ce dernier groupe de variantes, il s'agit de termes qui ne sont pas remplaçables, leur usage dépond du contexte. Parmi tous les synonymes trouvés, le récouvrement sémantique de ceux-ci est le plus limité. Dans les deux terminologies étudiées, nous n'avons trouvé qu'un seul exemple de ce type de synonymes. En polonais trois termes différents sont utilisés pour signaler que c'est une seule partie de l'articulation naturelle qui sera remplacée : *endoproteza częściowa*, *endoproteza jednoprzedziałowa* et *endoproteza jednostronna*. Le premier terme peut être utilisé par rapport à la hanche et au genou, les deux autres termes – uniquement par rapport au genou. En français, la différence concerne deux termes employés pour désigner une prothèse remplaçant toute une articulation : le terme *prothèse totale* peut se référer aussi bien à la hanche qu'au genou, le terme *prothèse tricompartimentale* quant à lui ne se réfère qu'au genou.

Le grand nombre de synonymes dans la terminologie biomédicale, en polonais ainsi qu'en français, prouve sans aucun doute que la synonymie est un phénomène tout à fait naturel pour les langues spécialisées. Ce phénomène semble d'autant plus visible sur l'exemple d'une nouvelle terminologie qui est en train d'être créée. Les spécialistes cherchent sans cesse comment exprimer de nouveaux concepts dans leurs langues. Si on prend en compte la rapidité du développement du génie biomédical et le fait que le réseau des chercheurs qui s'en occupent (et qui, en plus, représentent des disciplines différentes) est très large, le nombre et la complexité des synonymes ne devraient pas nous surprendre. Il faut ajouter encore que parmi les termes biomédicaux on peut

rencontrer non seulement des paires de synonymes mais également toutes les séries de synonymes qui réunissent les différentes variantes décrites plus haut. Dans les exemples ci-dessous, on peut distinguer les séries réunissant les variantes étymologique et de conceptualisation (1, 3), étymologique et syntaxique (2), syntaxique et de conceptualisation (4). Dans la série (5) toutes les variantes sont représentées.

1. *endoproteza / proteza implantowana / proteza stawowa / proteza wewnętrzna*
2. *bioinertność / materiał bioinertny / biomateriał obojętny / materiał obojętny biologicznie*
3. *articulation artificielle / articulation prothétique / implant articulaire / prothèse articulaire*
4. *prothèse totale du genou / PTG / prothèse tricompartimentale*
5. *cancérogénité / effets carcinogènes / effets oncogéniques / potentiel cancérigène / pouvoir carcinogène*

Pour terminer cette partie de notre article, citons un commentaire d'un des spécialistes polonais du génie biomédical qui a repondu à notre enquête concernant l'évaluation des termes et des langues utilisés dans son travail :

« J'ai toujours des doutes concernant le terme anglais *scaffold* – c'est une sorte d'échafaudage sur lequel des cellules sont posées et qui sert à assurer une forme géometrique voulue de tissu cultivé hors de l'organisme vivant. J'utilise conjointement les termes : *base spatiale*, *échafaudage*, *matrice* ou même la forme polonisée *skafold*. Je pense que le terme *skafold* devrait être admis dans la langue polonaise en tant qu'un terme dénotant un matériau spécialement conçu pour les besoins du génie tissulaire. […] Les autres le traduisent comme *porteur de cellules* mais, selon moi, c'est une notion trop étroite parce les *scafolds* peuvent être implantés également sans cellules, directement dans la lacune tissulaire. »[1].

Cette opinion du spécialiste montre bien comment la nouvelle terminologie est créée et travaillée. L'apparition des synonymes dans ce processus

1 Notre traduction du commentaire polonais : « Mam zawsze wątpliwości jeśli chodzi o termin 'scaffold' - rusztowanie, na którym osadza się komórki, które ma zapewnić właściwą formę przestrzenną tkance hodowanej poza organizmem człowieka. Używam zamiennie form : przestrzenne podłoże, rusztowanie, matryca a nawet 'skafold'. Myślę, że termin 'skafold' powinien być dopuszczony w j. polskim, jako opisujący materiał specjalnie zaprojektowany dla celów inżynierii tkankowej. […] Niektórzy tłumaczą go jako nośnik komórek, ale jest to

semble être le pain quotidien, surtout dans le cas d'un domaine d'études interdisciplinaire.

3. Polysémie dans la terminologie

Un autre phénomène linguistique jugé par l'école wüsterienne comme négatif c'était l'existence des termes polysémiques. L'analyse des termes biomédicaux montre que les langues spécialisées ne diffèrent pas de la langue générale en ce qui concerne la polysémie, même si les termes polysémiques ne sont pas tellement nombreux que les mots polysémiques. Parmi les termes biomédicaux recueillis nous avons distingué seulement quelques exemples des termes ayant plus d'une signification. En plus, les termes polysémiques se correspondent d'une langue à l'autre, inversement que dans le cas des synonymes.

Les termes *proteza* et *prothèse*, dans les deux langues, possèdent deux significations : dans un sens plus large, ils désignent un dispositif orthopédique utilisé à l'exterieur du corps pour remplacer un membre du corps humain (1), ainsi qu'un dispositif orthopédique implanté à l'intérieur du corps pour remplacer un organe (2).

(1) « Une aide technique qui remplace une partie du corps humain est une **prothèse**. Une prothèse tibiale remplace le pied et une partie de la jambe d'une personne amputée. » (TB : 131)

(2) « Toutes trois sont de forme plus ou moins hémisphérique calquée sur la forme du cotyle naturel. La **prothèse** cotyloïdienne métallique de Mac Kee-Farrar est hérissée sur sa surface convexe de petites tiges qui assurent son ancrage dans le ciment. » (PHG : 42-43)

Dans un sens plus étroit, ils dénotent uniquement les dispositifs utilisés à l'extérieur du corps. Pour éviter cette ambiguité, dans la terminologie française, on a introduit la distinction entre *prothèse externe* et *prothèse interne* mais dans la pratique le terme *prothèse* est utilisé sans adjectif et seul le contexte permet de déterminer de quel sens il s'agit. Dans les textes polonais,

wg mnie za wąskie pojęcie, bo scafoldy można używać bez komórek implantując je bezpośrednio w ubytek tkanki. »

les spécialistes préfèrent utiliser, dans ce deuxième sens, le terme *endoprothèse*[2].

Un autre terme polysémique dans la terminologie française est l'*implant*[3]. Il peut désigner soit un dispositif orthopédique inséré à l'interieur du corps (1), soit, par métonymie, une partie de ce dispositif dont les synonymes seraient les termes *composant* et *pièce* (2).

(1) « Les orthèses internes comme les prothèses internes sont également appelées **implant**, car toutes deux sont insérées ou implantées dans le corps humain. » (TB : 132)

(2) « **L'implant** cotyloïdien est une cupule métallique (du diamètre de la tête anatomique) remplie par un cotyle (habituellement rétentif) en polyéthylène serti qui reçoit la tête prothétique. » (CHO : 211)

Dans ce deuxième sens, le terme *implant* est le plus souvent utilisé avec un adjectif qui précise de quelle partie de la prothèse il s'agit (comme dans l'*implant cotyloïdien* plus haut). Dans ce contexte, apparaît notamment le terme *implant fémoral* qui, lui-même, est polysémique : il peut renvoyer à une partie de la prothèse du genou ou à une partie de la prothèse de la hanche, toutefois chacune d'elles se caractérise par un emplacement et une construction différents.

Les deux derniers termes polysémiques, fonctionnant dans les deux langues, sont la *biomécanique* et les *biomatériaux*. Le premier renvoie, en premier lieu, à une discipline scientifique qui examine le fonctionnement du corps humain du point de vue de la mécanique (1). En deuxième lieu, il désigne la manière dont fonctionnent les différentes structures du corps humain (2).

(1) « La **biomécanique** a progressé en cascade […] pour émerger comme champ d'intérêt vers la fin du XIX^{e} siècle. Cela explique pourquoi les faits saillants de la biomécanique sont encore mal définis et sont ouverts à discussion. » (QSJ : 5)

(2) « Au travers des différentes investigations reprise dans ce chapitre consacré à la **biomécanique de la marche** normale, nous avons tenté

2 Le terme français *endoprothèse* et le terme polonais *endoproteza* semblent être des équivalents parfaits mais, en réalité, ce ne sont que des faux-amis. Le terme polonais renvoie à un implant utilisé pour les os et les articulations tandis que le terme français denote un implant du système circulatoire. [3] On peut signaler en passant que son équivalent formel en polonais – *implant* – ne garde qu'un seul sens qui correspond à la première signification du terme français.

> de lever un bout du voile qui recouvre un phénomène complexe, peu accessible à l'étude, et qui pourtant nous est si familier. » (BG : 81)

Le même rapport peut s'observer entre les deux significations du terme *biomatériaux*. D'un côté, c'est le nom de la discipline qui conçoit les matériaux utilisés, par la suite, dans les différentes prothèses. De l'autre, ce terme denote l'objet même de ces recherches, à savoir tous les matériaux qui peuvent être implantés dans le corps humain. Dans la terminologie polonaise, le terme équivalent– *biomateriały* – semble gagner encore une troisième signification : par plusieurs auteurs, il est utilisé en tant qu'un hyperonyme de l'implant, un dispositif médical qui, sous diffèrentes formes, peut être inséré à l'intérieur du corps.

4. Conclusions

Nous avons montré deux phénomènes linguistiques significatifs pour une terminologie qui n'est pas encore figée et qui ne cesse à se développer : l'existence de la synonymie et de la polysémie.

Le nombre considérable de synonymes dans la terminologie étudiée confirme la thèse que le postulat wüsterien d'univocité des termes ne se reflète pas dans la réalité linguisitique. En plus, la synonymie semble tout à fait typique pour une terminologie qui est en train de naître. N'oublions pas que le génie biomédical est un domaine de recherches très vaste et très dynamique, se trouvant au carrefour de plusieurs disciplines traditionnelles comme la physique, la biologie ou la médecine. Ses spécialistes représentent donc des milieux scientifiques différents et, ce qui va forcement avec, des points de vue souvent très divergeants. Si on prend en compte tous ces facteurs, le grand nombre de synonymes dans les deux terminologies, et surtout leur complexité, deviennent compréhensibles et justifiables.

Les exemples des termes polysémiques présentés dans la deuxième partie de notre article prouvent que le phénomène de la polysémie existant dans les terminologies ne devrait surprendre personne, même s'il n'est pas tellement fréquent que la synonymie. L'analyse plus détaillée de ces exemples, et surtout de certaines nuances sémantiques et contextuelles qui commencent à apparaître entre les équivalents français et polonais signale que le développement de chaque terminologie suit son propre cours. Même si, au départ, les spécialistes des deux communautés linguistiques empruntent les termes à une seule langue (l'anglais en l'occurence), au fur et à mesure ces termes s'adaptent au nouveau système conceptuel et sémantique dans lequel ils fonctionnent.

Ce qui explique les divergences qui commencent à paraître entre certains équivalents français et polonais. Cela prouve aussi que tous les emprunts qui commencent à être utilisés dans un nouveau système terminologique, par une nouvelle communauté des spécialistes, acquièrent une nouvelle vie, ils sont automatiquement adaptés au niveau sémantique aux besoins linguistiques et terminologiques de nouveaux utilisateurs.

Dans ce qui précède, nous avons voulu présenter certaines implications sémantiques qui sous-tendent la terminologie *in statu nascendi*. Les conclusions qu'on a pu tirées nous emmènent à poser quelques questions : comment les termes présentés dans cet article vont-ils évoluer dans la suite, est-ce que le nombre de synonymes sera maintenu ou peut-être certains d'entre eux vont disparaître ou leur sens changera, est-ce que le sens des polysèmes evoluera encore. Ce sont des questions qui, pour l'instant, doivent rester sans réponses.

Références

Des fondements théoriques de la terminologie (2004). (dir.) C. Cortès, Cahier du C.I.E.L.

Gostkowska K. (2011). « Les terminologies française et polonaise du génie biomédical face au défi de la diversité », [dans :] *Passeurs de mots, passeurs d'espoir : lexicologie, terminologie et traduction face au défi de la diversité.* Actes des 8es Journées scientifiques du Réseau de chercheurs « Lexicologie, terminologie, traduction », Lisbonne, 15-17 octobre 2009, (dirs.) M. Van Campenhoudt, T. Lino, R. Costa, Éditions des archives contemporaines, Paris, pp. 139-148.

Le sens en terminologie (2000). (dirs.) H. Béjoint, Ph. Thoiron, Presses universitaires de Lyon, Lyon.

Pergnier M. (1989). *Les anglicismes*, Presses universitaires de France.

Sources d'exemples :

BG : *Biomécanique générale* (1985). (dir.) C. Kénési, Cahiers d'enseignement de la SOFCOT, vol. 22, Expansion Scientifique Française, Paris.

CHO : *Chirurgie orthopédique. Principes et généralités* (2004). (dir.) A.-Ch. Masquelet, Masson, Paris.

PHG : *Prothèses de hanche et de genou. Actualités et perspectives* (1988). (dirs.) J. Vidal, L. Simon, Collection de pathologie locomotrice 15., Masson, Paris.

QSJ : Allard P. et Blanchi J.-P. (1999). *La biomécanique*, Collection *Que sais-je ?*, Presses Universitaires de France, Paris.

TB : Dalleau G., Allard P. (2009). *Traité de biomécanique. Mécanique articulaire et tissulaire*, Presses Universitaires de France, Paris.

Summary

The aim of this article is to describe a part of biomedical engineering terminology the studies on which both in Polish and French literature are scarce so far. Biomedical engineering as a field appears to be a very promising area for studying the mechanisms of the formation of new terms and whole conceptual systems mostly because it is a rapidly developing field and, moreover, it is interdisciplinary enterprise. We decided to focus on this first feature of biomedical terminology. The aim of this article is to examine the linguistic phenomena that accompany the emergence of new discipline, in other words what role do synonymy and polysemy play in the terminology under study. What is more, we aimed to compare the tendencies arising in the two languages and communities whose social and political status differs vast- ly, as well as their culture and traditions with regard to science.

Objet, concept, terme, discours… Repenser le domaine dans le cadre de la terminologie comparée

Héba Medhat-Lecocq

32, Avenue Charles de Gaulle
91630 Marolles-en-Hurepoix
Institut national des langues et civilisations orientales
hebalecocq@yahoo.fr

Résumé. Depuis quelques décennies, les principes fondateurs de la terminologie ne font pas l'unanimité au sein même de leur propre domaine de recherche. La présente étude a pour objectif de faire la lumière sur les différentes interprétations de ces principes qui fixent les grandes lignes de notre discipline. Nous nous attacherons ici à diviser le travail terminologique en plusieurs étapes afin de mieux comprendre les différentes positions prises par le spécialiste vis-à-vis des éléments constitutifs du domaine. Suivre la même logique que ce dernier nous permettrait de dissiper certains malentendus qui pèsent depuis un certain temps sur la recherche terminologique, et plus particulièrement sur le domaine de la terminologie comparée.

1. Introduction

Établir un état des lieux de la science de la terminologie à l'heure actuelle n'est pas chose aisée. Bien que relativement moderne et malgré tous les moyens technologiques qui sont à sa disposition et qui devraient contribuer à son développement, la terminologie passe aujourd'hui par une période de remise en question, y compris par ses propres adeptes. Vacillant entre monosémie wüstérienne et polysémie de la langue naturelle ; partagée entre les deux démarches, onomasiologique et sémasiologique ; divisée entre les deux approches, normative et descriptive ; fragmentée par l'intervention de plusieurs acteurs qui n'ont pas tous les mêmes objectifs dans leur ligne de mire, la terminologie se trouve aujourd'hui à la croisée des chemins.

La présente recherche s'interroge sur les sources de toutes ces divergences et contradictions qui affectent l'évolution de la discipline et qui mettent constamment le terminologue face à un choix cornélien. Doit-il bannir la monosémie wüstérienne au profit de la polysémie de la langue ? Comment peut-il adopter une démarche onomasiologique alors que la première étape de son travail terminologique – celle du dépouillement – est réalisée à partir d'un corpus ? S'il adopte une approche normative, que faire des signes non normalisés ? Et si au contraire, il opte pour une approche descriptive, ne risque-t-il pas de transgresser les fondements, voire la raison d'être de la terminologie ?

Dans une tentative de donner quelques éléments de réponse à ces interrogations, nous poserons dans un premier temps le cadre de notre étude. Sous l'intitulé « *Le domaine, le spécialiste et le profane* », nous mettrons l'accent sur l'importance de la connaissance du domaine spécialisé avant tout projet d'étude terminologique.

Ensuite, nous proposerons une nouvelle vision du travail terminologique en le divisant en plusieurs étapes. Celles-ci seront avancées en fonction du rapport établi, au cours des processus cognitif et discursif, entre les éléments constitutifs de la science de la terminologie, à savoir **:** *l'objet*, *le concept*, *le terme* et *le discours scientifique*, mais aussi au fur et à mesure du changement de rôle de l'un des deux acteurs concernés directement par ce domaine, le spécialiste et le terminologue. La nouvelle division du travail terminologique que nous proposons ici et qui sera présentée sous le titre « *Le travail terminologique par étapes* », fera la distinction entre « *une langue émettrice du savoir* » et « *une langue réceptrice du savoir* », ces deux dernières n'adoptant pas toujours la même démarche, ni la même logique, pour conceptualiser et/ou décrire le domaine.

2. Le domaine, le spécialiste et le profane

Tout d'abord, nous tenons à signaler que l'emploi à ce stade du mot ″profane″ et non pas celui de ″terminologue″ n'est pas sans raison, ce dernier – au premier contact avec le domaine scientifique – étant considéré comme un profane. Néanmoins, en acquérant les connaissances nécessaires (à travers la documentation, la recherche sur le terrain, etc.) dans l'objectif d'appréhender le domaine, il passe du statut de ″profane″ à celui de « spécialisé » dans ce domaine.

Vectrice de connaissances sur les différents domaines scientifiques, la langue spécialisée est l'un des outils indispensables du spécialiste. Celui-ci

en a besoin pour appréhender les concepts scientifiques et les termes qui les véhiculent. Il en fait usage également pour transmettre son savoir à sa communauté scientifique, mais aussi à d'autres acteurs qui peuvent y être concernés : terminologues, traducteurs, simples usagers, bref, aux non-spécialistes du domaine.

Mais il arrive quelquefois que ces derniers peinent à saisir les informations scientifiques émises par le spécialiste. Cela est dû évidemment à leur manque du niveau d'expertise et du savoir-faire acquis par celui-ci et lui permettant une intégration conceptuelle bien particulière du domaine. L'exemple suivant illustre notre propos.

Prenons dans le domaine de l'architecture le concept d'< arc >, cet « élément architectural dont la forme et le mode d'appareil lui donnent une certaine flexibilité et lui permettent de résister à la poussée née de la charge qu'il supporte » (H. Medhat-Lecocq, 1997 : 855-856). Les architectes distinguent d'une manière générale entre les différents types d'arcs suivant leurs tracés : < arc en plein-cintre >, <arc en ogive>, <arc surbaissé>, etc. Or, ce mode de conceptualisation n'est pas absolu dans le domaine concerné, puisque l'on trouve certains arcs conceptualisés selon leur destination, *ie* leur fonction, comme c'est le cas de l'<arc de décharge>. Il s'agit d'un « arc surmontant un linteau dans le but de soulager la charge portée par celui-ci » (*Ibid.*, 1997 : 881) Il est, selon Pérouse de Montclos, « bandé dans un mur plein » (1989 : 267). Mais ce type d'arc doit épouser une forme quelconque, un tracé. Il prend souvent la forme d'un arc surbaissé, rarement celle d'un arc en mitre.

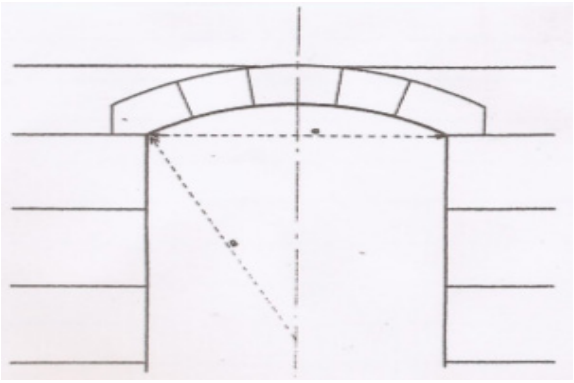


Figure 1 – Arc surbaissé

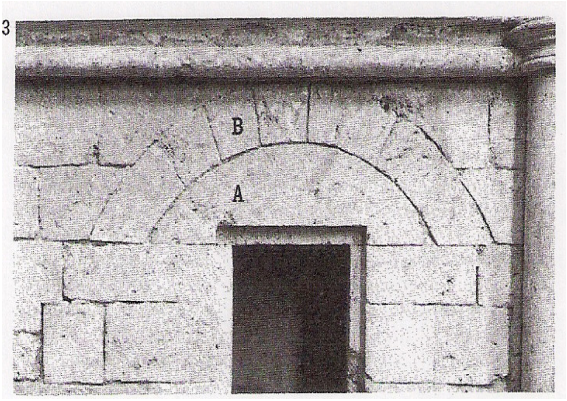


Figure 2 – Arc de décharge
(A-Linteau ; B-Arc de décharge)

Nous attirons l'attention ici sur les légendes qui accompagnent ces deux figures. Nous lisons *arc surbaissé*[1] sous la première, et *arc de décharge* sous la deuxième. Cela dit, à observer la deuxième figure sans la légende, une personne peu initiée et s'étant limitée à l'étude des arcs selon leurs tracés, risque de s'arrêter au concept d'<arc surbaissé>, puisque cet *arc de décharge*

1 Nous soulignons que le terme *arc surbaissé* a pour synonyme parfait celui d'*arc segmentaire*. Nous ne nous y attarderons pas ici.

épouse en même temps la forme d'*un arc surbaissé*. Cependant, le spécialiste découvrira deux concepts à travers cette deuxième illustration : un <arc de décharge> dont le tracé est celui d'un <arc surbaissé>.

- Sa connaissance du domaine lui permettra de distinguer entre deux différents concepts liés par un rapport d'implication logique : <arc surbaissé> et <arc de décharge>, puisque cet élément architectural est un arc surbaissé destiné à soulager la charge portée par un linteau.
- Il intégrera également cet <arc de décharge> dans un champ conceptuel le liant avec les concepts <linteau> et <mur plein> par une relation associative de contiguïté de type (objet → élément contigu), cet objet étant placé au-dessus d'un *linteau* et construit dans un *mur plein*.
- Pour aller encore plus loin, il classera < l'arc de décharge> et le < linteau > dans une autre relation d'implication partitive de type (objet → constituant) avec le concept de < baie > (une fenêtre, une porte, etc.), les deux premiers éléments faisant partie d'une baie et lui servant de couvrement. En d'autres termes, la présence d'une baie est une condition nécessaire pour la construction de ces deux éléments.

Bref, il aura toujours une vision plus claire, plus large et plus précise du champ conceptuel abritant le concept à délimiter. En voici la partie qui vient d'être évoquée :

- Implication logique (superordonné logique → subordonné logique)

(<arc surbaissé > → < arc de décharge >)

- Relation associative de contiguïté (objet → élément(s) contigu(s))

<arc de décharge> → <linteau> ; <mur plein>

- Implication partitive (objet →constituants)

<baie> → < arc de décharge > ; < linteau>

Nous pouvons bien évidemment pousser la démonstration encore plus loin en explorant ce champ conceptuel liant l'<arc de décharge> aux autres concepts du domaine. Mais contentons-nous de cette partie qui met en évidence le fait suivant : nous les non-spécialistes, lorsque nous abordons un domaine que nous ne maîtrisons pas, nous réussissons certes à appréhender une partie de son système conceptuel, mais cette dernière reste beaucoup moins importante que celle intégrée par le spécialiste de ce domaine. Au surplus, le mode de conceptualisation du domaine scientifique, reconnaissons-le, n'est pas toujours à la portée du non-initié. Et c'est là que réside le problème auquel font face les traducteurs de textes spécialisés, entre autres acteurs,

lorsqu'ils ont affaire à un domaine spécialisé. Ce fait amène Nicolas Frœliger, traducteur spécialisé et traductologue, à la réflexion suivante :

> «[…] *à notre vision, il manque le gros plan et les gros plans et les plans d'ensemble. [...] ; ce que nous savons, nous ne le savons en général pas sur le même mode que les spécialistes des domaines dans lesquels nous traduisons et qui sont, pour partie, nos interlocuteurs.*» (2013 : 74 et 78).

De ce qui précède, nous pouvons dégager les constatations suivantes :

- L'importance de l'intellection dans l'appréhension d'un domaine scientifique.
- L'expérience et le savoir-faire du spécialiste lui permettent une appréhension plus large, plus précise et plus organisée du réseau conceptuel de son domaine.
- Le terme ne révèle pas tout sur le concept.

C'est à la lumière de ces constatations que nous avons mené nos réflexions sur le travail terminologique. Ce faisant, nous l'avons divisé en plusieurs étapes

3. Le travail terminologique par étapes

Il convient de souligner tout d'abord que nous ne prétendons pas ici procéder à une analyse approfondie des éléments constitutifs de la terminologie – *objet*, *concept*, *terme* et *discours* – évoqués dans le libellé de notre article, travail déjà accompli de manière approfondie par d'autres recherches. Nous nous appuierons seulement sur ce qui nous paraît essentiel dans la réflexion que nous menons sur le travail terminologique. Notre objectif est de tenter de dégager la position du spécialiste par rapport à ces quatre éléments afin d'obtenir une vision plus claire du domaine de la terminologie et de trouver des éléments de réponse à nos interrogations évoquées *supra*.

Le travail terminologique est constitué d'un vaste ensemble de tâches. Il comporte entre-autres : la création des termes, la sélection entre les termes dans le cadre d'une activité au sein d'une instance de normalisation, le repérage et l'analyse des termes d'un domaine à travers le dépouillement d'un corpus textuel, la confection des dictionnaires et des bases de données, etc. Pour comprendre le travail terminologique sans brouiller les frontières entre toutes ces tâches que peut accomplir un terminologue, nous proposons de le

diviser en plusieurs étapes tout en tenant compte du rapport entre le domaine du savoir et la langue qui le véhicule.

3.1. Dans une langue émettrice du savoir

Une langue émettrice du savoir est une langue dans laquelle ont été créés les – ou une partie des – concepts du domaine. Dans cette catégorie de langues, le travail terminologique pourrait être réparti en trois étapes : la genèse terminologique, l'épreuve terminologique et l'observation terminologique.

3.1.1. La genèse terminologique

Deux sous-étapes sont nécessaires pour l'accomplissement de ce que nous proposons
d'appeler « la genèse terminologique », à savoir : la conceptualisation d'un nouvel objet et sa désignation. Au cours de la conceptualisation d'un objet scientifique, deux cas de figure peuvent se présenter :

- l'objet déjà existant est perçu pour la première fois par le spécialiste. Il en forme une image mentale : objet → concept ;
- lorsque le spécialiste invente un objet, il en construit d'abord une image mentale avant de le créer : concept → objet (cf. M. Diki-Kidiri et E. Atibakwa Baboya, 2008 : 36-37).

Dans les deux cas de figure, tout processus de conceptualisation implique une catégorisation du nouveau concept, ce dernier étant loin d'être saisi isolément. Il est plutôt délimité et classé par rapport aux autres concepts qui lui sont liés par une relation logique, partitive ou associative, à l'intérieur d'un même champ conceptuel relevant du domaine auquel ils appartiennent tous. Or, ce travail mental de perception et de catégorisation, auquel viennent s'ajouter la langue du spécialiste et, par conséquent, la mémoire collective des usagers de cette langue, joue un rôle essentiel dans la désignation du nouveau concept.

Prenons l'exemple du terme *minerve* utilisé dans le domaine de la médecine pour désigner ce collier cervical qu'on porte autour du cou et qui sert à « maintenir la tête en extension (en cas de traumatisme des vertèbres cervicales, etc.) » (PR)[2]. Comme nous le savons, ce mot tire son origine de « *Minerva*, déesse romaine de l'intelligence et de la sagesse » (TLFi)[3]. Pour

2 Le Petit Robert.
3 Trésor de la langue Française informatisé.

couvrir le concept scientifique de <minerve> avec ce terme, l'esprit humain établit une correspondance entre les deux images mentales : celle de la déesse de l'intelligence et celle de l'appareil médical concerné, puisque celui-ci est destiné à maintenir la tête – considérée comme le siège de l'intelligence et de la sagesse chez l'homme – « *en extension et en rectitude* » (TLFi).

Sur ce travail d'imagination effectué par le cerveau humain pour admettre les nouveaux concepts scientifiques avec les termes qui les couvrent et qui renvoient à d'autres objets bien ancrés dans la mémoire collective des usagers d'une langue, les exemples sont légion. Nous pouvons citer à titre exemplatif l'*éolienne*, mot qui tire son origine du nom grec Éole, dieu des vents, ainsi que le *nectar de fruit*, terme ayant fait l'objet d'un article publié par Pierre Lerat et dont le mot nectar nous renvoie, selon l'auteur, aux dieux des Olympes. (Lerat, 2009). Notons également le terme *nymphe* qui, dans le domaine de l'entomologie, représente le concept correspondant à l'un des stades de la métamorphose de l'insecte. Ce terme trouve son origine dans la mythologie grecque où le mot nymphe porte le sens d'une « *déesse d'un rang inférieur, qui hantait les bois, les montagnes, les fleuves, la mer, les rivières* » (PR)

Si l'on quitte le domaine de la mythologie pour se tourner vers d'autres types de termes motivés, beaucoup d'exemples ne peuvent manquer de venir à l'esprit, tels le *serpent*, cet instrument de musique ainsi nommé à cause de sa forme ondulante ressemblant à celle d'un serpent ; les *dents d'une scie* ; la *louve*, instrument qui, dans le domaine de la technologie, sert à soulever les pierres de construction (TLFi), etc.

Comme nous pouvons le constater à travers tous ces exemples, à cette étape-là, celle de la « genèse terminologique », la terminologie se trouve entièrement dépendante de la langue naturelle avec toutes les charges socio-culturelles et historiques que celle-ci véhicule et qui restent ancrées dans la mémoire collective de ses usagers.

3.1.2. L'épreuve terminologique

Elle commence lorsque le nouveau terme, concept/désignation, est présenté à l'ensemble des spécialistes du domaine, d'abord afin que ceux-ci l'approuvent et l'intègrent, ensuite pour qu'ils en fassent usage dans leur communication. Par conséquent, « l'épreuve terminologique » est divisée elle aussi en deux sous-étapes, à savoir : l'intégration du nouveau terme avec le concept qu'il couvre d'une part, et son usage dans le discours scientifique, d'autre part.

L'intégration du nouveau terme par les spécialistes du domaine

Certes à ce stade-là, le discours, majoritairement définitoire, est axé sur la description du nouvel objet scientifique, et, par la suite, centré sur le concept. Mais pour intégrer le terme représentant ce dernier, les spécialistes procèdent de la même manière qu'à l'étape précédente. Ils interpellent leur mémoire collective et établissent un rapport analogique entre les deux mondes : celui de leur domaine scientifique et celui appartenant à la mémoire collective des usagers de leur langue. Ils acceptent le terme *minerve* comme vecteur du « collier cervical qu'on porte autour du cou » en établissant une correspondance entre celui-ci et la déesse de la sagesse Minerva. Il en est de même pour l'*éolienne* qui les renvoie à Éole, dieu des vents. Ils font également le rapport entre la forme ondulée du serpent, cet instrument de musique, et le mouvement ondulant du reptile qui porte le même nom, etc..

Pour accepter ce type de termes, vecteurs de nouveaux concepts et considérés comme polysémiques dans la langue naturelle, les usagers de la langue ont besoin d'un repère qu'ils trouvent dans leur mémoire collective. C'est à ce stade-là, au moment de l'intégration du nouveau terme et du concept le véhiculant, ainsi qu'à l'étape précédente, celle de la genèse terminologique, qu'il nous vient à l'esprit la constatation faite par Loïc Depecker et selon laquelle : « [le] *terme*, dit-il, *est donc pour nous pleinement signe. Et c'est un signe vivant. »* (L. Depecker, 2000 : 92).

L'usage du terme dans les différents discours scientifiques

Une fois les nouveaux termes intégrés, les spécialistes, dans un objectif de décrire leurs objets scientifiques, commencent à en faire usage dans leur discours. À cet égard, plusieurs remarques méritent d'être signalées au cours de cette deuxième sous-étape.

- Lorsque le terme renvoie à une image collective propre aux usagers de la langue, les spécialistes refusent d'activer cette image susceptible de les détourner du concept scientifique véhiculé par ce terme.

Le pharmacien qui nous vend une *minerve* ne pense pas à la déesse de la sagesse. Il *met en veille* cette charge culturelle que porte le terme *minerve*. L'opticien qui fabrique les « lunettes », lui non plus, ce faisant, ne songe pas à la lune. Lorsque le spécialiste du domaine de l'entomologie parle de *nymphe*, il la présente uniquement dans le cadre de sa définition scientifique, à l'exemple de : *« Troisième état du cycle biologique caractéristique des Holométaboles*

qui précède l'état imaginal ; il est marqué par l'immobilité et l'absence d'alimentation. » (d'Aguilar et Fraval). Ainsi nous pouvons déduire que, lorsque le terme est intégré dans le discours scientifique, discours axé sur la description de l'objet scientifique, il ne représente que le concept scientifique.

- *Donc, convenons-en, dans leur discours scientifique, les spécialistes du domaine se servent du terme comme d'une « étiquette ».*

De ce fait, des mots comme *louve, serpent, nymphe*, etc., polysémiques dans la langue naturelle, deviennent des termes monosémiques à l'intérieur de leurs domaines scientifiques. Cette prise de position conceptuelle est une particularité de la langue spécialisée qui, pour créer de nouveaux termes, emprunte souvent des mots appartenant au lexique général de ses usagers, et pour se les approprier et leur donner le statut de termes, neutralise les autres sens qu'ils portent, ceci au profit des seuls concepts qu'ils devront véhiculer.

Ce qui place la terminologie dans un combat perpétuel avec la polysémie au profit de la monosémie.

- Lorsqu'un terme porte quelques ambiguïtés pour les futurs spécialistes, il est accompagné d'une glose désambiguïsante. Il est courant par exemple, dans un cours écrit ou oral destiné à des étudiants en médecine, que le professeur soit amené à accompagner le terme technique d'une glose ou d'une définition.

Prenons l'exemple du terme *colectomie*. Si les étudiants ne sont pas encore familiarisés avec le suffixe -*ectomie*, le terme technique *colectomie* leur sera présenté accompagné de sa glose : *ablation du côlon.*

- *Par conséquent, dans certains discours scientifiques, le terme n'est pas - et ne peut-être - le seul représentant du concept.*

- Cependant, il y a des situations de communication précises dans les différents domaines scientifiques où l'emploi d'un seul terme – du terme normalisé – s'impose. Ceci pour éviter toute éventuelle confusion.

Prenons le même exemple précédent. Dans un compte rendu opératoire, le seul terme utilisé est celui de *colectomie*, la glose *ablation du côlon* n'étant pas tolérée dans ce type de document.

- *Aussi, dans quelques types de discours scientifiques, le spécialiste et, par conséquent, le domaine de la terminologie, ont-ils besoin de la normalisation.*
- Pour éviter la répétition du terme, bien que celui-ci soit le seul normalisé, voire le seul admis par l'ensemble des spécialistes du domaine, en

tant que vecteur du concept, ces derniers le remplacent par son synonyme, par son quasi-synonyme ou, le plus souvent, par son hyperonyme.

Citons à titre d'exemple le lexème *engin*, terme générique qui vient à notre secours pour remplacer dans nos discours des termes comme *bulldozer*, *tracteur*, *char d'assaut*, *grue,* etc. Bien que ces derniers soient plus précis, leur répétition alourdit le discours sur les concepts qu'ils véhiculent. D'où le recours à un terme véhiculant son concept générique comme le montre l'exemple suivant :

> « *La principale mission du conducteur de* ***grues mobiles*** *est l'approvisionnement et l'alimentation en matériaux des différentes équipes intervenant sur un site (chantier, port, industrie…), avec lesquelles il communique par radio. Il doit également positionner et déployer* ***l'engin*** *sur le site, vérifier son fonctionnement, contrôler les poids de chargement qu'il peut supporter.* »[4]

- *Ainsi, le terme normalisé ne suffit pas à lui tout seul pour répondre à tous les types de discours scientifiques.*

- Une connivence tacite existe entre les spécialistes d'un même domaine et fait que l'implicite n'est pas totalement absent de leurs discours.

Appuyons-nous cette fois-ci sur un exemple relevant du domaine de la terminologie, celui concernant les deux termes : *concept* et *notion*. Imaginons qu'un terminologue écrit : « La terminologie selon la théorie wüstérienne donne la priorité au **concept** et par conséquent plaide pour une étude synchronique des langues spécialisées ». Ensuite, pour étayer son propos, il avance une citation de Wüster : « *La priorité accordée aux* ***notions*** *entraîne la terminologie par la force des choses, à considérer la langue d'un point de vue synchroniqu*e » (E. Wüster, 1981 : 64)

S'adressant à sa communauté professionnelle, aux terminologues francophones dans le cas présent, cet auteur n'a pas besoin d'expliquer quand, comment et pourquoi ils sont passés du terme *notion* à celui de *concept*, ceci pour la simple raison qu'il estime que tous les spécialistes du domaine de la terminologie en sont au fait.

4 http://www.ecf.asso.fr/Formation-pro/Les-metiers/Batiments-et-travaux-publics/Conducteur-de-grues-mobiles.

- *Par conséquent, l'implicite n'est pas absent dans le discours des spécialistes du domaine.*

Ce dernier exemple montre également que dans leur communication scientifique, les spécialistes d'un même domaine donnent la priorité au concept.

3.1.3. L'observation terminologique

Si les deux étapes précédentes concernent le spécialiste du domaine, l'acteur principal de cette troisième étape est le terminologue. Pour analyser le discours scientifique dans l'objectif d'en étudier la terminologie, quelle approche ce dernier devrait-il adopter ? Partant des idées et principes dégagés à partir des deux étapes précédentes, nous pouvons trouver la réponse à cette question à travers le raisonnement suivant :

- puisque le spécialiste a besoin du terme normalisé dans quelques types de discours scientifiques ;
- puisque ce terme normalisé ne suffit pas à lui tout seul dans toutes les situations de communication scientifique[5] ;
- pour respecter la monosémie par rapport à la langue naturelle, phénomène que s'impose le spécialiste lorsqu'il décrit son objet scientifique ;
- pour comprendre en même temps son recours aux différents représentants discursifs du terme dans d'autres types de discours scientifiques ;
- pour pouvoir répondre à toutes ces exigences de la terminologie descriptive, exigences qui paraissent au premier abord contradictoires, mais à notre sens, ne le sont pas point,

notre seul îlot de sûreté est le concept.

Si nous parvenons à délimiter et catégoriser les différents concepts du domaine, nous pourrons ajouter pour chacun, le terme normalisé, son synonyme, son quasi-synonyme, bref tous les représentants discursifs de ce terme normalisé cher à la science de la terminologie. Ceci évidemment en prenant en considération son degré de technicité, ainsi que le type de discours susceptible de l'abriter. À cet égard, nous tenons à souligner le manque de recherches en terminologie sur les différents types de discours scientifiques, à

5 Pour plus de détails sur le rapport entre la terminologie normative et la terminologie descriptive, cf. Héba-Medhat Lecocq, 2009 : 157-234.

l'exemple de celle de Gérard Cornu sur le domaine du droit, dans son ouvrage « Linguistique juridique » (2005).

Pour étayer notre propos, nous proposons le schéma suivant à travers lequel nous présentons le concept comme une étoile au centre d'un système planétaire et autour de laquelle gravitent, chacun sur son orbite, les différents signes qui la représentent. Chaque orbite devrait définir le degré de technicité et de légitimité du signe linguistique, sa fréquence d'emploi, et plus particulièrement le type de discours susceptible de l'abriter.

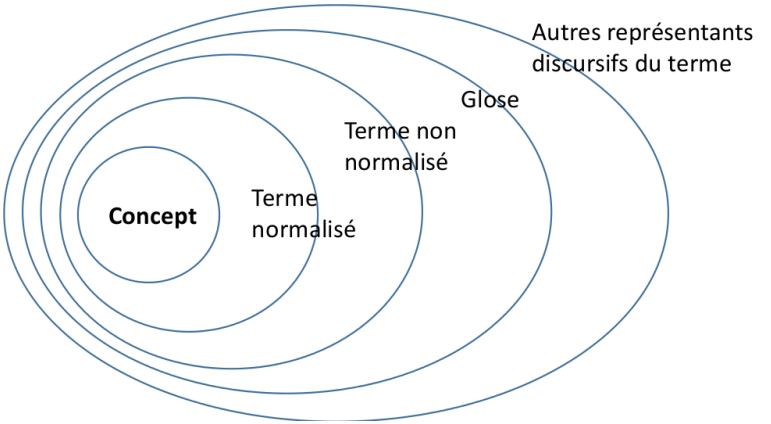


Schéma 1 – Recherche terminologique axée sur le concept

Nous tenons à souligner ici que la terminologie formelle, *i.e.* l'ontoterminologie, trouve sa place dans ce schéma, puisque les tenants de cette dernière branche de la terminologie cherchent « à mettre en regard, c'est-à-dire à associer, les termes d'usages et les termes normés avec les concepts de l'ontologie correspondants. » (Marie Calberg-Challot *et al.* 2009 : 37)

Dès lors, une question s'impose.

Comment, dans la pratique, un terminologue pourrait-il arriver à suivre ce schéma, alors qu'il commence sa recherche terminologique par le dépouil-

lement d'un corpus et, par conséquent, a affaire en premier lieu aux signes linguistiques ?

En fait, le repérage des termes d'un corpus tel que nous le concevons et c'est ce que d'ailleurs beaucoup d'entre nous le font inconsciemment, peut se résumer à travers les deux formules suivantes :

- signe linguistique ↔ concept (2 étapes indispensables)
- objet→ concept → signe linguistique (terme et non terme) (3e étape, si nécessaire)

L'exemple évoqué *supra*, celui des deux termes *concept* et *notion*, corrobore d'une manière remarquable la première formule. Lorsque le terminologue fait usage des deux termes – le premier dans son propre discours et le second dans une citation de Wüster – les spécialistes du domaine ne s'attarderont pas sur la différence entre les deux termes pour comprendre ce discours, puisque leur processus de compréhension du texte est axé sur le concept. Certes, leur première démarche sera forcément sémasiologique : signe linguistique → concept. Mais ils finiront par renverser ce processus de compréhension : concept → terme, en inférant que le concept évoqué dans les deux discours, celui du terminologue contemporain et celui de Wüster, est le même. C'est la seule raison pour laquelle ils acceptent cette quasi-synonymie, d'ordre diachronique dans le cas présent.

Quant à la deuxième formule, elle représente une partie de la démarche adoptée par le terminologue au cours de l'étape de « l'observation terminologique ». Elle s'avère indispensable lorsque ce dernier se trouve dans l'incapacité de délimiter les concepts scientifiques véhiculés par les termes qu'il avait repérés. La recherche sur le terrain et le recours au spécialiste lui permettent de voir l'objet concret, et/ou de percevoir l'objet abstrait, seul moyen pour le conceptualiser et l'intégrer avec le (ou les) signes(s) linguistique(s) qui le couvre(nt).

3.2. Dans une langue réceptrice du savoir

Le travail terminologique dans une langue réceptrice du savoir passe, comme pour la langue émettrice du savoir, par les trois étapes évoquées *supra*. *Cependant,* la différence entre les deux types de langues se manifeste plus particulièrement au cours de la première étape, celle de *la genèse terminologique* qui devient ici soit une *réception terminologique,* soit *une deuxième genèse terminologique*. Analysons les trois étapes à partir de ce nouvel angle, celui de la langue réceptrice du savoir

3.2.1. La réception terminologique / La genèse terminologique

Lorsqu'une langue accueille sur son territoire un nouvel objet scientifique ou technique, il lui arrive de le percevoir selon la même vision du monde, mais aussi d'adopter la même conceptualisation du domaine concerné que celle de la langue d'origine. Pour ce qui est des nouveaux concepts récemment importés, s'ils sont représentés par des emprunts ou des calques à partir de cette dernière langue, il s'agirait dans ce cas-là d'une ***réception terminologique***.

Nous pouvons prendre en exemple la terminologie de l'appareil digestif dans trois langues différentes, le français, l'anglais et l'arabe, où – une influence de l'ancienne médecine grecque en est certaine – plusieurs organes sont conçus et perçus de la même manière :

Langue française	**Langue anglaise**	**Langue arabe**
Langue	Tongue	لسان (lisân) (littéralement : langue)
Pancréas	Pancreas	بنكرياس (bankeriâs)
Duodénum	Duodenum	الاثناعشري (al-ithna-ʿachri) (littéralement : qui se compose de douze…)
Côlon	Colon	القولون (al-qawloun)
Côlon transverse	Transverse colon	القولون المستعرض (al-qawloun al-mustaʿrad) (littéralement : côlon transverse)
Côlon descendant	Descending colon	القولون النازل (al-qawloun al-nâzil) (Littéralement : côlon descendant)
Côlon ascendant	Ascending colon	القولون الصاعد (al-qawloun al-ṣâʿid) (Littéralement : côlon ascendant

La réception terminologique, comme nous pouvons le constater, traduit donc une sorte de symétrie conceptuelle et terminologique entre les différentes langues en contact. Il peut en découler certes des problèmes d'un autre ordre qu'une langue émettrice du savoir ne connaît pas ou rencontre rarement, comme celui de l'implantation de nouveaux termes dans la langue d'accueil, ainsi que leur concurrence avec les emprunts introduits par la langue émettrice du savoir. Pour ce qui est des termes indiqués *supra*, ce sont les calques

et les emprunts qui ont réussi à évincer les anciens termes arabes. Mais pour d'autres domaines plus récents comme l'informatique ou la téléphonie mobile, le problème de la concurrence entre les emprunts et les nouveaux termes créés par la langue d'accueil est loin d'être simple. Dans le monde arabe par exemple, des termes d'origine anglo-saxonne (kombyoutar (de computer) ; mobayel (de mobile), etc.) continuent à concurrencer leurs successeurs créés par les instances de normalisation arabes.

Néanmoins, il arrive quelquefois que cette langue d'accueil, en recevant le nouveau domaine, résiste à la conceptualisation d'origine et décide de conceptualiser ce dernier selon sa logique et sa propre vision du monde. Dans ce cas-là, il est question ***d'une deuxième genèse terminologique***. Deux causes peuvent être à l'origine de cette dernière : une différence de conceptualisation du domaine et une perception différente d'un même objet.

Nous pouvons parler d'*une différence de conceptualisation* lorsque le champ conceptuel concernant un objet est modifié. Il n'est que de penser aux concepts de <Yin> et de <Yang> en Médecine traditionnelle chinoise qui n'existent pas dans les autres médecines du monde. Si, dans la première, plusieurs maladies sont à l'origine d'« *un excès d'énergie Yin* » et d'autres reviennent à « *un excès d'énergie Yang* », (Christian Brun, 2011 : 125), le champ conceptuel d'une maladie, voire le système conceptuel de la pathologie chinoise traditionnelle sera différent de celui des autres médecines.

En revanche, *percevoir le même objet différemment* n'implique pas forcément une modification du champ conceptuel concernant cet objet – bien que nous ayons coutume de parler de différence de conceptualisation dans ce cas de figure – mais peut, par contre, influer sur le choix du terme qui le représente.

Pour nous expliquer, appuyons-nous sur l'exemple d'un objet relevant du domaine de l'architecture orientale. Il s'agit d'un support architectural qui peut servir également d'élément décoratif. Il constitue une évolution de la *trompe* et du *pendentif* et permet le passage du plan carré au plan circulaire dans la construction des coupoles. Lorsque plusieurs éléments de la trompe se répètent et se resserrent les uns à côté des autres, cet élément prend le nom de *muqarnas*, une forme altérée de *muqarfas*, qui signifie *accroupi*, le spécialiste arabe l'ayant perçu à l'image de plusieurs hommes assis en accroupi les uns à côté des autres.[6]

6 Nous signalons qu'il existe une autre interprétation selon laquelle le terme *muqarnas* tire son origine de *corniche*.

D'un autre côté, au premier contact avec ce support, le spécialiste français l'avait perçu différemment. Il a vu plutôt ces éléments juxtaposés comme *des alvéoles* (d'une ruche d'abeilles). Et lorsque leurs piédroits s'allongent pour se suspendre dans l'air, il les a perçus comme des *stalactites* qui se forment sur les voûtes des grottes (Fig. 3 & 4). Aussi a-t-il appréhendé le domaine, à la lumière de cette nouvelle perception des objets concernés, selon *une deuxième genèse terminologique.*

Mais comme le mot *alvéoles* n'a pas été terminologisé, il est a été progressivement remplacé par l'emprunt *muqarnas,* tout en l'accompagnant quelquefois comme une glose afin de le désambiguïser s'il en est besoin. Cela démontre que pour un Français, les *muqarnas* continuent à être perçus comme des objets évoquant la forme des alvéoles dans une ruche d'abeille.

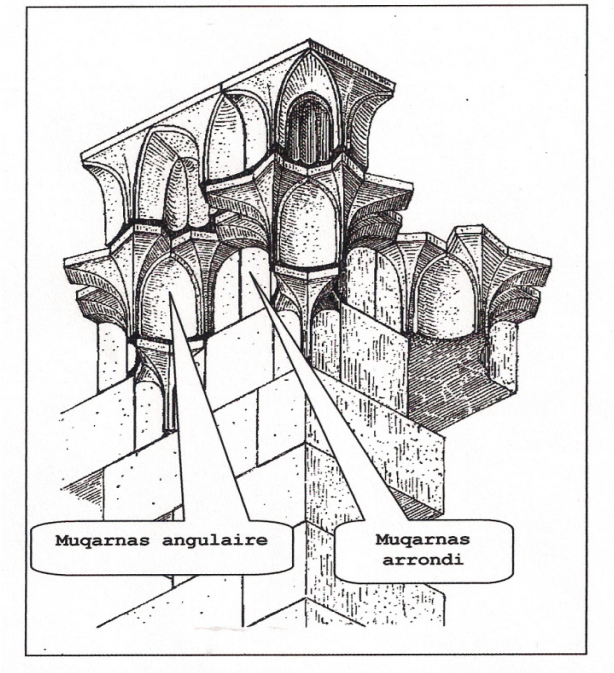


Figure 3. Muqarnas angulaires et arrondis

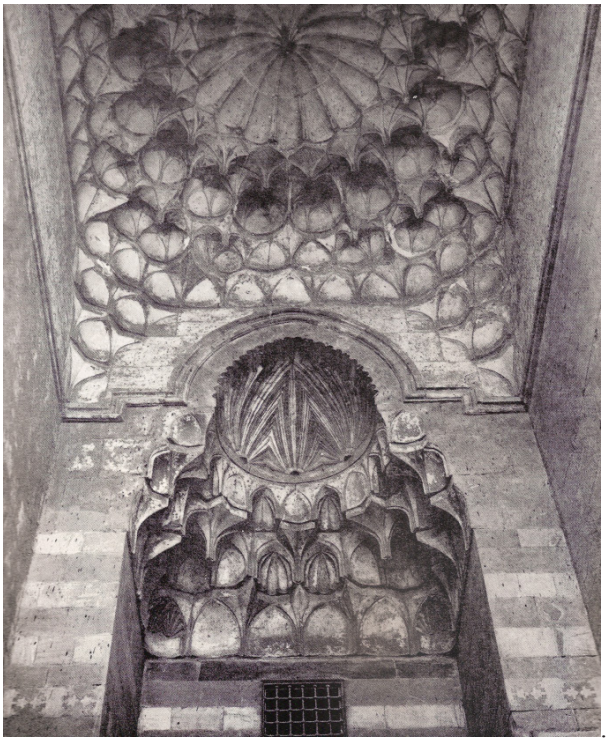


Figure 4. Stalactites. Entrée du palais Yashbak. Le Caire, Égypte

Cette deuxième *genèse terminologique* n'a pas été sans conséquence sur l'élaboration par chacun des deux spécialistes de la terminologie couvrant le champ conceptuel propre à cet objet :

Spécialiste égyptien	Spécialiste français
- *muqarnas halabi (muqarnas aleppin)* ou *muqarnas châmi (muqarnas syrien*), cet élément étant très prisé en Syrie	- *muqarnas arrondi,* la partie supérieure épousant une forme arrondie.
- *muqarnas baladi (muqarnas local)*, cet élément étant très fréquent en Égypte	- *muqarnas angulaire*, la partie supérieure épousant une forme pointue.

- *muqarnas dhu dallâya (muqarnas suspendus)*, lorsque les piédroits de plusieurs éléments juxtaposés se détachent et s'allongent pour se suspendre dans le vide.	- *stalactites*, ces éléments juxtaposés ressemblant aux concrétions calcaires qui se forment sur les voûtes des grottes.

Tableau 1 – conceptualisation par les deux spécialistes, égyptien et français, des trois types de muqarnas.

Arrivé à ce point, nous tenons à signaler que le champ conceptuel reste toujours le même dans les deux cultures, arabe et française. C'est seulement la terminologie qui en répercute les échos.

3.2.2. L'épreuve terminologique

Pour aller encore plus loin, cette différence de perception des mêmes objets impacte curieusement le discours à leur égard lorsque les spécialistes les décrivent dans les deux langues. Prenons l'exemple suivant :

> « *L'ornementation architecturale composée d'**alvéoles (muqarnas)** rangées horizontalement ou groupées verticalement avait déjà fait son apparition en Egypte dans certaines constructions fatimides.* […]. *Dans un portail monumental en pierre ou en marbre appareillé, **muqarnas** et **stalactites** se combinent parfois pour former les pendentifs de la demi-coupole (ou coquille supérieure, sinon l'encadrement d'une coupole côtelée.* » (J.C. Garcin *et al.* 1982 : 243).

Si dans le texte français, on lit : « *muqarnas et stalactites se combinent …..* », le spécialiste arabe ne peut faire usage que de *muqarnas* pour désigner les trois formes de cet élément. Dans ce cas-là, pour évoquer la même idée, il remplacerait « *muqarnas et stalactites* » par « *les muqarnas dans toutes leurs formes se combinent parfois pour....* » ; ou bien « *les différents types de muqarnas se combinent parfois pour....* » ; ou encore « *les trois types de muqarnas se combinent parfois pour....* ». Bref, il est impossible de remplacer les deux termes français *muqarnas* et *stalactites,* qui d'ailleurs couvrent trois concepts, par deux termes arabes.

Ainsi pouvons-nous déduire que, dans leur discours, axé sans aucun doute sur les concepts – les mêmes pour les deux langues – les spécialistes se trouvent dans l'obligation d'agencer les éléments de leurs discours en fonction de leur

choix pour les termes qui couvrent ces concepts. Ce choix, soulignons-le est souvent tributaire de leur perception des différents objets scientifiques.

4. L'observation terminologique

Par conséquent, un premier dépouillement des deux corpus, arabe et français, nous conduirait vers deux champs « conceptuels ? » différents, comme l'illustre le schéma suivant :

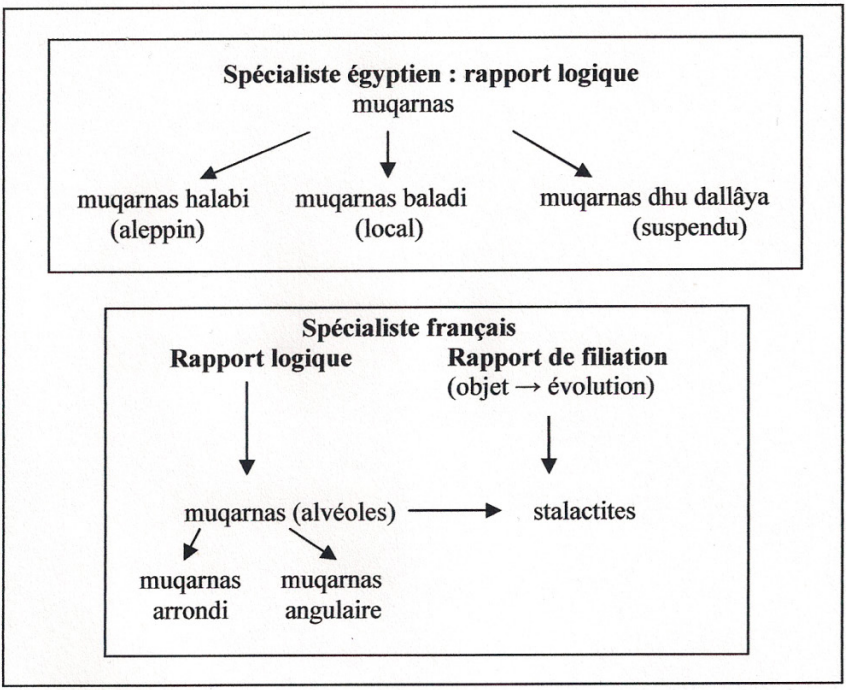


Schéma 2 : Champs de muqarnas élaboré à partir des termes

En effet, les deux champs, loin d'être incorrects, sont incomplets parce qu'élaborés à partir des termes qui couvrent les concepts, et non pas à partir des concepts eux-mêmes. Autrement dit, le dépouillement terminologique a été effectué en se cantonnant à la seule formule : signe linguistique → concept. Ce qui donnerait pour la langue arabe trois types de *muqarnas* (trois

co-hyponymes), et pour le français deux types de *muqarnas* (deux co-hyponymes) liés avec un troisième, *stalactites,* par un rapport associatif de filiation (objet → évolution).

Néanmoins, la réalité scientifique est tout autre. Car le spécialiste égyptien qui évoque dans son discours trois types de *muqarnas* est conscient que le troisième élément, *muqarnas dhu dallâya* (*muqarnas suspendu*) est une évolution des deux premiers. De même, le spécialiste français dans sa conceptualisation pour ces trois éléments admet totalement qu'ils renvoient à un objet pouvant épouser trois formes différentes. En d'autres termes, au cours du passage d'une langue à l'autre, le champ conceptuel des objets concernés n'a pas été touché. C'est plutôt le champ terminologique, basé sur le seul dépouillement textuel et constitué sans aucune analyse conceptuelle, qui a mené à ce résultat susceptible de nous éloigner de la réalité scientifique. « […], *les choses ne sont pas telles qu'elles sont dites*, dit Christophe Roche. *La conceptualisation d'un domaine ne peut-être extraite d'une seule analyse linguistique des textes.* » (2008 : 54).

Ce dernier exemple démontre une fois de plus que nous ne pouvons pas échapper à la nécessité d'une recherche terminologique axée sur le concept.

5. Conclusion

Depuis la fondation par Eugène Wüster de la science de la terminologie moderne, d'autres théories et approches dans ce domaine ont vu le jour. Si les unes viennent compléter l'approche wüstérienne, d'autres se situent aux antipodes de celle-ci, voire la rejettent totalement. Or, la remise en question des principes fondateurs de la discipline impacte depuis quelques décennies le travail du terminologue, le premier acteur à en pâtir. Cette situation tient à deux causes principales.

D'une part, ce dernier n'est pas le seul acteur, ni même l'acteur principal du domaine, celui-ci étant le spécialiste. Par conséquent, un des problèmes majeurs auxquels se heurte le terminologue est celui de la connaissance du domaine spécialisé. Pour ce faire, il devrait suivre la même logique du spécialiste et appréhender le domaine par intellection.

D'autre part, le travail terminologique est un long processus qui mérite d'être divisé en plusieurs étapes, chacune nécessitant une prise de position particulière par rapport aux éléments constitutifs de la terminologie. En d'autres termes, concevoir le travail terminologique en le divisant en plusieurs étapes permet d'avoir une vision plus précise et plus globale de notre discipline. À

travers cette nouvelle conception du domaine, le terme peut être envisagé comme un « *signe linguistique* » lors de sa création ou son intégration pour la première fois ; mais en même temps, il est considéré comme une « *étiquette* » lorsque le spécialiste en fait usage dans son discours sur l'objet scientifique. Signalons également que dans quelques situations de communication, la synonymie et la quasi-synonymie trouvent toute leur légitimité ; dans d'autres types de discours scientifique, seul le terme normalisé est toléré. De plus, cette nouvelle division du travail terminologique met en évidence l'intérêt de l'approche onomasiologique sans rejeter totalement la sémasiologie, indispensable au début de l'étape du dépouillement terminologique. Finalement, nous avons pu constater que toute étude de la terminologie d'un domaine doit tenir compte des différences entre une langue émettrice du savoir et une langue réceptrice du savoir. Cette distinction entre les deux langues – qui se manifeste dans la conceptualisation du domaine, dans la perception de ses objets, dans le choix des termes et par la suite dans l'organisation du discours – constitue la clef de voûte de tout travail en terminologie comparée. Par conséquent, elle devrait être prise en considération par le terminologue.

Au terme de cette étude, nous espérons avoir apporté quelques éclaircies sur des zones d'ombre qui couvrent le domaine de la terminologie et qui, à notre sens, font obstacle à son développement.

Références

D'Aguilar J. et Fraval A. *Glossaire progressif d'entomologie*, (en ligne) http://www7.inra.fr/opie-insectes/glossaire.htm.

Brun C. (2011). *Le grand livre de la naturopathie*, sous la direction d'Anne Ghesquière, Paris, Eyrolles.

Calbert-Challot M., Lerat P. et Roche Ch. (2009). « Quelle place accorder au corpus dans la construction d'une terminologie », in *Terminologie & Ontologie : théories et applications*, Toth, Institut Porphyre, p. 33-52.

Cornu G. (2005). *Linguistique juridique*. Éditions Montchrestien. Coll. Domat. Droit privé, 3e édition.

Depecker L. (2000). « Le signe entre signifié et concept », in Henri Béjoint et Philippe Thoiron (dir.), *Le sens en terminologie*, Lyon, Presses universitaires de Lyon, Travaux du CRTT, p. 86-121.

Diki-Kidiri M. et Atibakwa Baboya E. (2008). « Conceptualisation, dénomination et perception cuturelle », in Marcel Diki-Kidiri (dir.), *Le vocabulaire scientifique dans les langues africaines. Pour une approche culturelle de la terminologie,* p. 31-51.

Froeliger N. (2013). *Les noces de l'analogique et du numérique. De la traduction pragmatique*. Traductologiques, Paris, les belles lettres.

Garcin J.C., Maury B., Revault J. et Zakariya M. (1982). *Palais et Maisons du Caire D'Époque Mamelouke (XIIIe-XVIe siècles)*. Éditions du CNRS, Paris.

Lerat P. (2009). « La combinatoire des termes. Exemple : nectar de fruit ». *Hermès*, 42-2009, p. 211-232.

Medhat-Lecocq H. (1997). *L'architecture islamique en Égypte. Élaboration d'une terminologie et problématique de sa traduction*. Thèse de doctorat non publiée, le Caire : Université Al-Azhar. (Dictionnaire contextuel de l'architecture Islamique, en cours).

Medhat-Lecocq H. (2009). « Le travail terminographique entre terminologie normative et terminologie descriptive », décembre 2009, in *Revue des Sciences Humaines*, le Caire, Égypte, p. 157-234.

Pérouse de Montclos J.-M. (1989). *Architecture. Vocabulaire Principes d'analyse scientifique,* Imprimerie Nationale Éditions, Paris.

Roche Ch. (2008). « Faut-il revisiter les Principes terminologiques ? », in Actes de la deuxième conférence *Toth*, p. 53-72.

Wüster E. (1981). « L'Étude scientifique générale de la terminologie, zone frontalière entre la linguistique, la logique, l'ontologie, l'informatique et les Sciences des choses ». in *Textes choisis de terminologie*. p. 54-114.

Site

http://www.ecf.asso.fr/Formation-pro/Les-metiers/Batiments-et-travaux-publics/Conducteur-de-grues-mobiles

Sources des illustrations

Bourgoin J. (1889) : *Précis de l'art arabe et matériaux pour servir à l'histoire, à la théorie et à la technique des Arts de l'Orient musulman*. Vol. 1 : L'Architecture. Paris : E. Leroux. (Fig. 3)

Maury B., Raymond A., Revault J. et Zakariya M. (1983), *Palais et Maisons du Caire. Époque ottomane. (XVI^e^ – XVIII^e^ siècles),* Nouvelle édition CNRS, (Fig. 1)

Pérouse de Montclos J.-M. (1989), *Architecture. Vocabulaire Principes d'analyse scientifique,* Imprimerie Nationale Éditions, Paris, (Fig. 2)

Revault Jacques et Maury Bernard (1977). *Palais et Maisons du Caire. Du XIV^e^ au XVIII^e^ siècle.* Tome II. Planche XXV. « Palais Yashbak : détail des stalactites de pierre surmontant l'entrée du palais. (Fig. 4)

Summary

Over the past decades, founding principles of terminology have been far from unanimously agreed, even within their own field. The aim of this research is to highlight the different interpretations of these principles**,** which lay down the main outlines of our discipline. Here, we will strive to divide terminology work into several steps to better understand the different positions taken by the specialist regarding the constituent elements of the field. Following the same logic as the latter, would allow us to clarify certain misunderstandings which are hanging over terminology research for quite some time now, in particular the comparative terminology field.

Du classement alphabétique à la représentation ontologique : la terminologie des systèmes d'enseignement supérieur

Maria Teresa Zanola*, Mario Corveddu, Carol Rolla

*Università Cattolica del Sacro Cuore
Largo Gemelli 1, Milan
osservatorio.terminologie@unicatt.it
http ://centridiricerca.unicatt.it/otpl-home

Au cours des dernières années, la représentation ontologique a acquis un statut de référence en tant que modalité d'organisation, de gestion et de formalisation conceptuelle des domaines de la connaissance. Dans le cas de l'informatique et de la gestion de la connaissance elle contribue de plus en plus à un dialogue fécond parmi les spécialistes. En terminologie, l'ontologie - conçue comme la modalité de représentation conceptuelle d'un domaine - est utilisée dans les pratiques de création de ressources terminologiques monolingues et multilingues ; c'est dans ce champ d'application que s'insère la réflexion de Neches et al. (1991), selon laquelle "*An ontology defines the basic terms and relations comprising the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary"*.

Cette contribution se propose de décrire le projet et le développement d'une ontologie qui a été le résultat de la collaboration entre terminologues et spécialistes de bibliothéconomie, au sein d'un projet pilote de l'Osservatorio di Terminologie e Politiche Linguistiche (OTPL), en collaboration avec la Biblioteca d'Ateneo de l'Università Cattolica del Sacro Cuore de Milan.

L'objectif du projet est de fournir une représentation ontologique d'une ressource terminologique existante, afin de créer un produit innovant apte à simplifier la gestion des contenus, introduisant une modalité intuitive et interactive pour la recherche de l'information et offrant une représentation conceptuelle complète du domaine.

La ressource qui a constitué le point de départ pour l'élaboration de l'ontologie est le glossaire INTERHEd, un travail scientifique créé à de la première édition du Master « INTERHEd - Internazionalizzazione della Higher Education. Titoli, professioni e terminologia nei sistemi d'istruzione supe-

riore » (http ://centridiricerca.unicatt.it/otpl-prodotti-terminologici-glossario-della-higher-education), organisé par la Facoltà di Scienze linguistiche e Letterature straniere de l'Università Cattolica del Sacro Cuore de Milan en collaboration avec CIMEA (Centro d'informazione sulla mobilità e le equivalenze accademiche). Ce glossaire, ayant pour but de recueillir la variété terminologique d'un domaine complexe tel que celui des systèmes d'enseignements supérieurs, se compose d'un recueil de données textuelles - 229 unités terminologiques - sous forme de tableaux, où pour chaque terme on spécifie la catégorie grammaticale, le sigle ou l'acronyme, le synonyme, la définition, les notes et les sources. Cette structure, caractérisée par une hiérarchie informelle, permet de recueillir un nombre considérable de données, mais qui se présente inadéquate à satisfaire les nécessités d'un domaine en évolution.

Notre projet s'articule en deux parties : une première phase de construction d'un modèle de conceptualisation ontologique appliqué au glossaire originaire à travers le logiciel Protégé, et une seconde phase de représentation graphique de l'ontologie à travers un logiciel de *knowledge visualization*.

Pendant la première phase de travail, le moteur de recherche de l'ontologie a été utilisé, d'abord pour la création du réseau sémantique conceptuelle du domaine concerné, et après, pour reconstruire le réseau des relations sémantiques transversales d'un échantillon significatif de termes extraits du glossaire avec la collaboration de spécialistes du domaine.

La deuxième phase a concerné la représentation du réseau sémantique obtenu par la première phase à travers un logiciel de *knowledge visualization* développé par l'entreprise américaine *TheBrain Technologies*, avec l'objectif d'optimiser la structure graphique fournie par Protégé et de l'enrichir avec les fonctionnalités proposées par cette deuxième application.

Ces opérations ont rendu possible la création d'un schéma conceptuel ontologique doté de plusieurs fonctions de recherches des termes, capable d'identifier les noyaux de l'ontologie et ouvert à l'instauration de nouvelles relations entre les concepts considérés. Le résultat final a été donc la transformation du glossaire originaire de la forme à tableaux à structure réticulaire, où l'aménagement des contenus et la représentation ontologique contribuent à offrir un instrument remarquable pour la systématisation du savoir et l'exploitation du domaine de la connaissance concerné (Losito 2015).

La construction de la carte conceptuelle grâce à *TheBrain* représente une expérience concrète de la possibilité d'utiliser une ressource terminologique

qui, de simple support linguistique - le glossaire -, est devenue un véritable instrument d'analyse immédiate pour différents contextes d'usage.

Cette typologie d'approche ontologique, à travers l'usage d'un logiciel construit sur le langage de représentation des connaissances OWL (*Ontology Web Language*), a permis d'obtenir un milieu opérationnel dynamique qui a favorisé la réutilisation à trois niveaux :

- au niveau de projet : les phases opérationnelles de ce projet pilote peuvent être appliquées à d'autres ressources terminologiques avec les mêmes résultats ;
- au niveau de la structure : la possibilité d'expansion des champs d'application de la ressource permet de se déplacer d'un domaine précis à la conceptualisation formelle d'une discipline ;
- au niveau de contenu : une représentation ontologique qui respecte les standards de la programmation informatique permet au contenu d'être introduit dans différentes applications et d'être donc utilisé dans plusieurs domaines du savoir.

En conclusion, il est évident qu'aujourd'hui les spécialistes de la terminologie peuvent apprécier les avantages et les possibilités que l'usage d'ontologies informatiques peut apporter non seulement à une communication technique et scientifique exhaustive et rigoureuse, mais aussi à la création d'outils hybrides et innovants.

Bibliographie

Biagetti, M.T., *Le ontologie come strumenti per l'organizzazione della conoscenza in rete*, « Aida informazioni », 28 (2010), 1/2, p. 9-31.

Centro Informazioni Mobilità Equivalenze Accademiche (CIMEA), Glossario, 2013, http ://www.cimea.it/glossario.aspx ?IDC=2765.

Centro Informazioni Mobilità Equivalenze Accademiche (CIMEA), Banca dati dei sistemi universitari, 2014, http ://www.cimea.it/countryModules.aspx.

Collesi, P., Serpente A., Zanola, M.T., (dir.), *Terminologie e ontologie. Definizioni e comunicazione fra norma e uso*, Educatt, Milano, 2013.

Depecker, L., Dubois, V., De Gennes, P.G., Roche, C., *Terminologie et ontologie : description du réel*, Société française de terminologie, Paris, 2007.

Hoekstra, R., *Ontology Representation : Design Patterns and Ontologies that Make Sense*, IOS Press, Amsterdam, 2009.

Losito, L., « Il glossario reticolare : un'ontologia per la Higher Education », in D. Vellutino – M.T. Zanola (dir.), *Comunicare in Europa. lessici istituzionali e terminologie specialistiche,* Educatt, Milano, 2015, p. 149-160.

Neches, R., Fikes, R., Finin, T., Gruber, T., Patil, R., Senator, T., Swartout, W.R., *Enabling Technology for Knowledge Sharing.* AI Magazine, Fall 1991, p. 36-56.

Osservatorio di Terminologie e Politiche linguistiche, *Glossario della Higher Education* (versione italiana e inglese), 2014, http ://centridiricerca.unicatt.it/otpl-prodotti-terminologici-glossario-della-higher-education.

Poli, R. (dir.), *Theory and Applications of Ontology. Computer Applications,* Springer, Heidelberg, 2010.

Staab S., Studer R., *Handbook on Ontologies* (International Handbooks on Information Systems), Springer, 2009.

Application de ressources linguistiques à grande échelle pour compléter une ontologie de domaine (Large-scale Linguistic Resources Application for Domain Ontology Population)1

Olena Orobinska* **, Jean-Hugues Chauchat, Natalya Charonova**

* Université de Lyon, ERIC Lyon 2, 69676 Bron, France
** KhPI National University, Kharkov, 61002, Ukraine
Olena.Orobinska@univ-lyon2.fr
sharonova@kpi.kharkov.ua

Résumé. Dans cet article, nous proposons une méthode semi-automatique pour compléter une ontologie de domaine, illustrée par un exemple de taille réelle. Cette méthode hybride combine différentes techniques - statistique, linguistique, lexicale - ce qui permet de détecter les variations lexicales (les différentes entrées lexicales) des concepts initiaux du noyau de l'ontologie construite à l'aide d'experts du domaine.
Les expérimentations ont été effectuées sur le lexique du domaine de la Sécurité Radiologique (SR). On présente ici une première application qui complète un noyau d'ontologie par de nouveaux termes détectés dans les textes de normes internationales de Sécurité Radiologique. Les résultats confirment le rôle important de la modélisation linguistique initiale et de l'utilisation de ressources lexicales externes dans le processus d'instanciation.

1 Ce travail a bénéficié d'une bourse de la Région Rhône-Alpes.

1. Introduction

Pendant longtemps, les méthodes statistiques d'analyse de textes jouaient le rôle le plus important dans le processus de construction d'une ontologie ; les autres approches (linguistiques ou lexicales) avaient des fonctions auxiliaires. Ceci s'expliquait par le manque de performance des outils correspondants.

À partir du début des années 2000, on commence à reconnaitre que la résolution de certains problèmes de TAL, par exemple la détection de relations liant les concepts, la désambiguïsation lexicale (WSD – word sense disambiguation), etc., ne sera pas possible sans un changement radical de stratégie de récupération des unités sémantiques (les termes et leurs relations) pour constituer une ontologie (de domaine) à partir de ressources textuelles. Bien sûr, l'élaboration de modèles linguistiques est beaucoup plus laborieuse car ces modèles requièrent des outils additionnels telles que les analyseurs morphologiques (*TreeTagger, Brill* parmi les outils adaptés à la langue française) ; les analyseurs syntaxiques de concordances (parseurs) tels que *Cordial* où *TXM* ; les ressources lexicales extérieures telles que les dictionnaires des synonymes. Mais les résultats attendus promettent d'être plus pertinents, plus précis et plus complets que ceux obtenus par des méthodes purement statistiques. Les démarches récentes sont présentées dans la section suivante.

Pour la vérification de nos idées et les expérimentations nous avons choisi le domaine de la Sécurité Radiologique. Il y a plusieurs raisons à ce choix : le lexique du domaine est riche ; les termes sont complexes, souvent composés de plusieurs mots ; les ressources textuelles sont homogènes structurellement et stylistiquement. Ces ressources sont accessibles en plusieurs langues, ce qui autorise la comparaison d'application de la méthode sur des corpus de différentes langues.

Nous avons créé et testé deux corpus parallèles en français et en russe. La qualité de résultats obtenus est moins bonne pour le corpus russe car les ressources linguistiques en ligne sont actuellement moins performantes par rapport à ceux pour le français. Dans cet article, nous présentons et discutons les résultats sur les textes français.

La suite de l'article est organisée comme suit : la deuxième section présente l'état de l'art ; la troisième section décrit notre approche. Un exemple illustrant la méthode est présenté dans la section quatre. Finalement, nous présentons un bilan des résultats obtenus et discutons les problèmes à résoudre.

2. L'état d'art

Nous nous concentrons sur les approches et méthodes d'apprentissage d'une ontologie de domaine. La particularité de l'ontologie de domaine est ce que la terminologie du domaine fait sa base.

Traditionnellement l'apprentissage d'ontologie (OL - ontology learning) a utilisé des méthodes statistiques. Ensuite les méthodes linguistiques (syntaxiques) ont été largement appliquées en raison de la performance croissante des analyseurs morphologiques, par exemple TreeTagger, Brill [Bulteaar 2008, Schmid 2000, Brill 1992]. Au cours des dernières années, les ressources lexicales lisibles par les ordinateurs (machine readable) ont gagné du terrain. Des exemples de ces ressources sont les bases de connaissances génériques (OpenCyc[2]), thésaurus – génériques et spécialisés (UMLS[3]), dictionnaires synonymes (WordNet[4]), dictionnaires multilingues (BabelNet[5]) etc. Ce fait, à son tour, a stimulé l'émergence de nouvelles techniques lexicales.

Les méthodes d'apprentissage visent soit à créer ontologie « from scratch » ou instancier une ontologie de base existante prédéfinie à l'aide des experts du domaine.

Jusqu'à récemment on pouvait citer deux stratégies principales pour construire une ontologie. La première est dite stratégie « Layer cake » [Cimiano 2005], où chaque étape utilise les résultats des tâches précédentes. On commence par l'extraction des termes-candidats qui, ensuite, se groupent et forment la hiérarchie des relations. Les règles d'apprentissage sont aussi dérivées de ces données initiales (les termes-candidats). Les supports les plus réussis, qui sont construits sur cette approche et qui continuent de se développer, sont KAON, TEXT2ONTO [Bozsak 2002, Haase 2005].

Une autre approche propose d'appliquer d'abord le modèle linguistique permettant de découvrir les entités de concepts à travers les relations prédéfinies. La description détaillée de cette approche ainsi que les outils de mise en œuvre tels que TERMINAE et CHAMELEON sont présentés dans [Mondary et al., 2008] et [Aussenac-Gilles et al., 2000].

Un ouvrage tout récent décrit les progrès et la maturation actuelle de cette vision. Les auteurs déploient les principes les plus généraux des systèmes

2 http ://www.cyc.com/platform/opencyc
3 http ://www.nlm.nih.gov/research/umls/
4 http ://wordnet.princeton.edu/
5 http ://babelnet.org/

de traitement des textes en langue naturelle. L'ontologie globale linguistique est au cœur d'un tel système. Les auteurs ont élaboré le modèle Lemon qui applique ces idées [Cimiano et al. 2014].

Roberto Navigli et ces collègues proposent largement d'utiliser les ressources lexicales multilingues, notamment BabelNet, pour les différents aspects de traitement linguistique [Flati et Navigli 2013, Velardi et al. 2012]. En particulier, ils proposent une approche unifiée multilingue, basée sur le graphe, qui procure le liage des entités et désambiguïsation des sens de mots basée sur une identification floue de significations, combiné à l'heuristique la plus dense d'un sous-graphe. L'approche permet de sélectionner des interprétations sémantiques de haute cohérence Moro et al., 2014].

L'application de telles facilités à l'OL (ontology learning) ne se fera pas attendre longtemps.

Toutefois la classification des relations elles-mêmes est un problème qui n'est pas encore entièrement résolu car toute relation peut être exprimée linguistiquement bien différemment. Certaines relations sont actuellement acceptées : synonymie, relations hiérarchiques ou taxonomiques telles que hyperonymie, méronymie, etc. Néanmoins de nombreux ouvrages proposent leurs propres classements. Ainsi, B. Rosario et M. Hearst [Rosario, 2001] proposent 13 classes qui décrivent les relations sémantiques possibles dans un groupe nominal entre le nom central et ses dérivés. D. Moldavan et collègues proposent le système de 35 classes pour classer les relations sémantiques dans les groupes nominaux [Moldavan *et al.*, 2004] ; plus tard ce schéma a été développé par R. Girju et collègues [Girju et al., 2005]. T. Chklovski et P. Pantel introduisent cinq classes pour caractériser relations sémantiques entre les verbes [Chklovski and Pantel, 2004].

Parmi les algorithmes visant à déceler les relations sémantiques on peut mentionner les deux suivantes :

L'analyse distributive (DA - Distributional Analysis) permet de mettre les termes en relation avec les types de relations différentes [Morlane-Hondère et Fabre, 2012]. Mais on est alors contraint de faire la distinction entre les relations syntagmatiques (comme les collocations où les relations nom-verbe) et les relations paradigmatiques (comme la synonymie, hyperonymie) [Fabre et Bourigault, 2006].

L'exploitation de patrons lexico-syntaxiques permet d'avoir la bonne précision. La contrainte principale est qu'ils sont limités par les contextes où ils

peuvent être utilisés. Habituellement, un seul type de relation peut être déterminé avec cette méthode [Périnet et Hamon, 2013].

Dans les sections suivantes nous allons présenter en détail notre méthode ainsi que les résultats expérimentaux.

3. Description de la méthode

Dans cet article nous proposons la description d'une méthode de regroupement sémantique de mots - les variations lexicales de concepts génériques dans le corpus - qui sont liés par la relation de synonymie.

La terminologie d'un domaine forme le lexique de ce domaine. Les termes spécifiques, tels que les dénominations de grandeurs physiques, les dénominations des unités de mesure, sont listés dans les glossaires, et, dans les textes, ils ne sont pas remplacés par des synonymes. Par contre, les dénominations de certains phénomènes peuvent varier assez largement à travers de corpus. Par exemple les mots « dégradation », « destruction », « détérioration », etc., ont le même sens que le concept « dommage » proposé par l'expert de domaine.

L'objectif de nos recherches est de trouver tous les mots qui ont le même sens que ceux du glossaire de départ et de les rajouter dans l'ontologie de noyau. Du fait que, généralement, les concepts supérieurs ne sont pas très nombreux [Velardi et al. 2012], nous sommes convaincus que la meilleure pratique pour amorcer le processus d'OL (apprentissage d'ontologie) est l'utilisation d'un modèle initial de domaine fourni par un expert. Dans la forme simplifiée, il peut s'agir de la liste des concepts génériques, c'est-à-dire du lexique de domaine, lequel est généralement disponible en tant que tel dans le corpus. À partir de là, les techniques d'apprentissage peuvent jouer un rôle puissant à toutes les étapes de processus de la population de l'ontologie.

Pour éviter ultérieurement l'ambiguïté entre les mots « verbe », « propriété » et « prédicat », nous expliquons la différence comme suit :

« Le prédicat décrit généralement une propriété de la personne ou de la chose visée par le sujet, ou bien décrit une situation dans laquelle cette personne ou chose joue un certain rôle dans les clauses élémentaires décrivant une action ; le sujet indique normalement l'acteur, la personne ou la chose qui exécute l'action, alors que le prédicat décrit l'action. » [Huddleston et al., 2006]

Les verbes jouent un rôle clé pour décrire les actions et les autres propriétés des sujets ; en d'autres termes, les verbes révèlent explicitement les

propriétés de sujet. De cette façon, les mots « verbe » et « propriété » peuvent être considérés comme des synonymes. Nous allons utiliser « prédicat » pour indiquer les verbes qui expriment des propriétés importantes ou des relations importantes entre les concepts de l'ontologie.

Il faut noter que chaque langue naturelle a plusieurs moyens (syntaxiques et lexicaux) pour attribuer une propriété à un objet ; par exemple, les phrases « *les effets nocifs de la radio-exposition* » et « *la radio-exposition se caractérise par les effets nocifs* » rassemblent deux concepts : « *effets nocifs* » et « *radio-exposition* ». Pourtant, dans le premier cas, cela est mis en œuvre au moyen d'une préposition et d'un article défini (« *de la* ») ; dans le second cas, cela est mis en œuvre par le biais du verbe pronominal *se caractériser* au présent. Cette multitude de formes linguistiques rend difficile le traitement de textes en langue naturel. Comme il a été mentionné ci-dessus, nous nous sommes concentrés sur les verbes en tant que moyen principal de rapprochement sémantique des différents mots qui n'ont pas de ressemblance lexicale (par exemple, pas la même racine).

3.1. L'algorithme

L'idée est de récupérer toutes les possibles entrées lexicales des concepts prédéfinis qui se trouvent dans le corpus. La première étape est l'analyse syntaxique du corpus. Cette analyse est réalisée à l'aide d'un analyseur syntaxique (pour le français, nous avons utilisé *TreeTagger*). Les nouveaux candidats-termes, pour peupler l'ontologie de noyau, sont les synonymes sémantiques de chaque concept de l'ontologie de noyau. Le dictionnaire de synonymes nous donne les synonymes des concepts initiaux ; parmi tous les synonymes proposés dans le dictionnaire, nous avons sélectionné les substantifs qui sont dans le corpus. Pour résoudre le problème d'ambiguïté des mots (*WSD – word-sense disambiguation*), nous avons adapté la méthode AFC (analyse formelle de concepts (ou dans l'abréviation anglaise *FCA – formal concept analys*).

L'analyse grammaticale a été effectuée pour adapter la méthode AFC à nos objectifs : nous avons sélectionné toutes les paires « sujet-verbe » où le sujet était un des concepts de l'ontologie de noyau suivi par ce verbe.

Pour créer le contexte formel de chaque concept, nous avons pondéré les verbes liés aux sujet-concepts ; pour cela, nous avons construit un indicateur basé sur le calcul d'information mutuelle. La pondération des verbes nous a permis d'enregistrer les verbes qui sont les plus caractéristiques de chaque

concept et d'éviter les verbes trop abstraits et généraux mais aussi les verbes trop spécialisés. Notre méthode propose le peuplement d'une ontologie de domaine à travers l'algorithme décrit Fig.1.

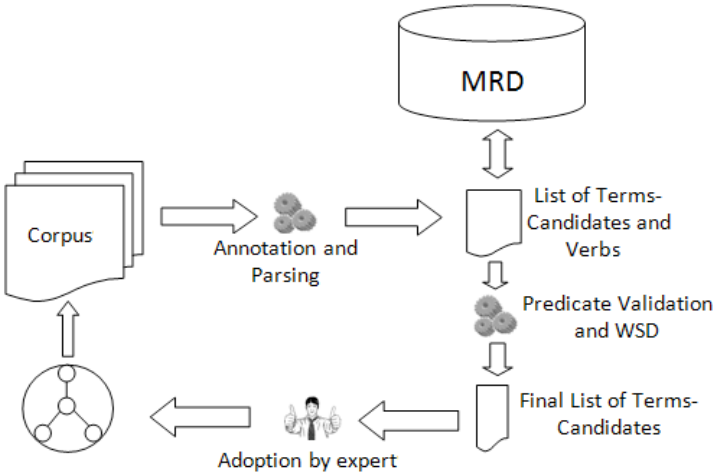


Fig.1. Schéma de peuplement de l'ontologie par les synonymes linguistiques.

En gros, on peut différencier trois étapes de ce schéma (chacune rassemblant plusieurs sous-étapes) :

Préparation des informations (données) ;

Interrogation de ressource lexicale extérieure ;

L'adoption des résultats et peuplement d'ontologie.

3.2. Phase préparatoire

Cette phase englobe les processus suivants :

- Mise en forme du corpus et modélisation primaire du domaine à l'aide des experts de domaine. À la sortie de cette étape on doit obtenir une liste des concepts génériques du domaine.
- Balisage de corpus avec TreeTagger de façon à obtenir à la sortie les textes (lignes) transformés sous la forme «…***pos/lemme pos/lemme pos/lemme***…» ; où «***pos***» correspond à une certaine partie du discours reconnue par TreeTagger, et «***lemme***» présente la forme nor-

male d'un mot : l'infinitif pour les verbes ; le masculin singulier pour les noms et adjectifs, etc.

- Analyse grammaticale superficielle des phrases afin de récupérer les informations sur tous les noms (substantifs), les verbes et leur disposition dans chaque phrases. Pour cela, nous avons écrit notre propre module en Java.
- Validation des prédicats, c'est-à-dire des verbes significatifs qui assignent à chaque concept une propriété importante (aux termes du domaine concerné) avec une probabilité au-dessus d'un certain seuil calculé. La validation de cette étape est non-supervisée ; elle se base sur les calculs d'un coefficient qui caractérise le niveau de « spécificité » de chaque verbe par rapport à chaque concept. Nous donnons plus de détails sur ces calculs dans la partie suivante de cet article.

3.3. Interrogation de ressource lexicale

- Interrogation en ligne d'un dictionnaire de synonymes pour récupérer dans le corpus tous les synonymes-candidats de chaque concept.
- Procédure non-supervisée qui permet de résoudre le problème d'ambiguïté des mots et de ne conserver, parmi tous les synonymes-candidats, que des plus pertinents. Pour cela, nous utilisons la méthode AFC - analyse formelle des concepts (ou FCA - *formal concept analysis*), adapté à nos objectifs.

3.4. L'adoption des résultats et peuplement d'ontologie

L'expert du domaine intervient pour cette phase : il doit vérifier et confirmer les résultats de l'étape précédente pour que les entités nouvelles puissent être rajoutées à l'ontologie de noyau.

3.5. Modèle mathématique

Les données initiales et intermédiaires sont :

- $W=\{w\}$ est l'ensemble des substantifs (noms) apparaissant au moins une fois dans le corpus ;
- $V=\{v\}$ est l'ensemble des verbes apparaissant au moins une fois dans le corpus ;

- $C=\{c_i \mid i=1...n,\ n \in N\}$ est la liste des concepts représentant le modèle de domaine ;
- $CV=\{c_i, v_j\}$ est l'ensemble des paires « nom – verbe » où le nom correspond à un des concepts c_i en tant que sujet dans la phrase et v_j est son prédicat.
- $DL=\{DL_i \mid i=...n,\ n \in N\}$ est le jeu des groups de synonymes qui ont été extraits de dictionnaire CRISCO pour chaque concept ;
- $DL_i=\{l_{ij}\}$ est l'ensemble des synonymes du dictionnaire pour le concept i ;
- $CL_i = W \bigcup DL_i$ et $CL_i := \{ w \subseteq W \mid \forall w : (w, c_i) \in I_{syn} \}$ est la liste de tous les synonymes potentiels pour le concept i qui on trouve dans le corpus, où $I_{syn} \subseteq DL \times W$ est la relation binaire et transitive de synonymie.

Pour évaluer les synonymes nous avons adopté lé méthode de l'Analyse Formelle de Concepts (ACF).

Pour délimiter l'ensemble des prédicats qui forment le contexte formel (*formal context*) de chaque concept, nous avons utilisé un coefficient de pondération mesurant « l'association » entre un nom (concept) et un verbe.

Ce coefficient est le produit de la valeur de l'information mutuelle :

$$I(c_i, v_j) = \sum f_{ij} * \log \frac{f_{ij}}{f_{ci} * f_{vj}}$$

et l'indice de Jaccard modifiée pour évaluer leurs similarités :

$$k_{ij} = \frac{f_{ij}}{N_i + N_j}$$

Ici, f_{ij} est la fréquence d'occurrence du concept c_i en tant que sujet de prédicat v_j ; $f(c_i)$ et $f(v_j)$ sont les fréquences des occurrences du concept c_i et du prédicat v_j respectivement.

Rappel de la méthode AFC

Définition 1. Le triple $K=(G; M; I)$ est un *contexte formel* si G et M sont des ensembles disjoints et I est une relation binaire entre G et M. Les éléments de G sont nommés ***les objets***, ceux de M sont nommés ***les attributs*** et I est ***l'incidence*** du contexte.

Définition 2. Le *concept formel* (A,B) est la paire des ensembles disjoints des objets (A) et des attributs (B) respectivement, si l'affirmation suivante est vraie : chaque objet partage tous les attributs de l'ensemble d'attributs et tous les attributs sont partagés par chaque objet de l'ensemble des objets.

Soit (*G, M, I*) un contexte formel. Nous définirons *A'* pour $A \subseteq G$: $A' := \{ m \subseteq M \mid (\forall g \in A : (g,m) \in I\}$ comme l'ensemble des attributs communs pour tous les objets A ; et *B'* pour $B \subseteq M$: $B' := \{ g \subseteq G \mid \forall m \in B : (g,m) \in I\}$ comme l'ensemble des objets qui possèdent tous les attributs B.

Étant donné la définition 2, la paire (A, B) est un concept formel (*G ; M ; I*) si et seulement si $A \subseteq G, B \subseteq M\ A' = B, B' = A$.

Les ressources utilisées

Pour tester la méthode nous avons utilisé les ressources suivantes :

- Deux corpus originaux, en partie parallèles : un corpus de textes français et un de textes russes. La majeure partie de deux corpus est fondée par les mêmes documents (au sens de la structure et le contenu), abordant la problématique du domaine. Ce sont notamment des documents officiels accessibles en ligne, tels que les normes de sûreté radiologique et les rapports de l'IAEA et des Commissions internationales et nationales de protection radiologique. Le corpus français est plus grand, il contient en plus des articles de revues spécialisées publiés depuis 1999. Ce corpus (texte brut) contient environ 1 500 000 mots et le corpus russe 600 000 mots. La plupart de ces textes étant initialement en format PDF, on a dû les convertir dans un format (texte ou HTML) utilisable pour le traitement. Dans le cadre de cet article on discute les résultats obtenus sur les données françaises.
- Dictionnaire de synonymes. Dans notre cas, après en avoir testé plusieurs (par exemple WOLF, ALEXANDRIA), nous avons choisi CRISCO[6] - le dictionnaire des synonymes en ligne élaboré à l'Université de Caen ; il fait partie du grand projet CNRTL soutenu par le CNRS depuis 2009.

6 http ://www.crisco.unicaen.fr/des/

4. Expérimentation

Nous avons effectué les expérimentations sur neuf concepts de niveau supérieur, définis avec l'expert du domaine ; ils ont formé notre liste générique initiale C={*dommage, exposition, contrôle, personnel, population, protection, rayonnement, sûreté, source*}.

Des informations intermédiaires au fil des expérimentations sont rassemblées dans les tableaux 1 et 2.

Tableau 1 Paramètres agrégés sur le corpus français

Nb des substantifs différents dans le corpus (sous forme de lemme)	*Nb de tous les verbes différents dans le corpus*	*Nb des verbes associés avec les concepts*	*Nb des prédicats*
2 464	1 094	423	211

Tableau 2 Récapitulatif des résultats des expérimentations

Concept	*Nb des verbes associés*	*Nb des prédicats*	*Nb des candidates-synonymes*	*Nb des candidates-synonymes dans le corpus*	*Nb des candidates-termes après AFC*	*Nb des entités adoptés*
dommage	28	20	37	11	7	6
exposition	92	69	34	22	5	1
contrôle	111	92	17	17	1	1
personnel	69	54				
population	42	31	9	5	5	4
protection	148	127	77	32	25	17
rayonnement	144	103	26	12	10	8
sûreté	86	51	31	12	9	5
source	159	117	35	22	9	4

La première colonne de tableau 2 contient la liste de concepts. Les valeurs qui se trouvent dans les autres colonnes du tableau 2 sont suivantes :

- ***Nb des verbes associés :*** correspond au nombre de verbes associés avec ce concept (ce chiffre présente toute variété des verbes liés avec le concept donné).

- ***Nb des prédicats :*** correspond au nombre de verbes dont le coefficient d'association est au-dessous du seuil (ils forment le contexte formel d'après la méthode AFC)
- ***Nb des candidates-synonymes :*** correspond au nombre de noms qui ont été trouvé dans le dictionnaire des synonyme CRISCO pour chaque concept initial.
- ***Nb des candidates-synonymes dans le corpus :*** correspond au nombre de ces noms (parmi les tous candidates-synonymes issus du dictionnaire) qui se trouvent dans le corpus.
- ***Nb des candidats-termes après AFC :*** correspond au nombre de mots qui, avec les prédicats, forment le contexte formel.
- ***Nb des entités adopté :*** correspond au nombre de mots qui, finalement ont été acceptés par l'expert.

Exemple qualitatif : un des concepts du domaine de SR est le mot « *dommage* ». Dans le dictionnaire des synonymes CRISCO on trouve 38 synonymes potentiels. Dans le corpus il n'y a que 11 mots de cette liste (*atteinte*, *brèche*, *dégât*, *dégradation*, *destruction*, *détérioration*, détriment, *ennui*, *intérêt*, *lésion*, *perte*). Par ailleurs, dans le corpus, le mot « dommage » paraît en tant que sujet avec 28 verbes différents. Mais après la pondération, seulement 20 de ces verbes dépassent le seuil pour former le contexte formel. Enfin, parmi les 11 candidats-termes, seulement 7 se trouvent avec les mêmes verbes : *brèche*, *dégradation*, *destruction*, *détérioration*, *détriment*, *lésion*, *perte.* Enfin le mot « *brèche* » a été exclu par l'expert de domaine.

Le graphique à la Fig.2 montre les valeurs du coefficient d'association entre le concept « dommage » et les verbes qui sont associés avec lui. Le seuil a été établi heuristiquement au niveau égal à 1,3E-4. Les verbes pour lesquels la valeur de coefficient d'association est inférieure à cette grandeur ont été conservés pour être soumis à l'expert.

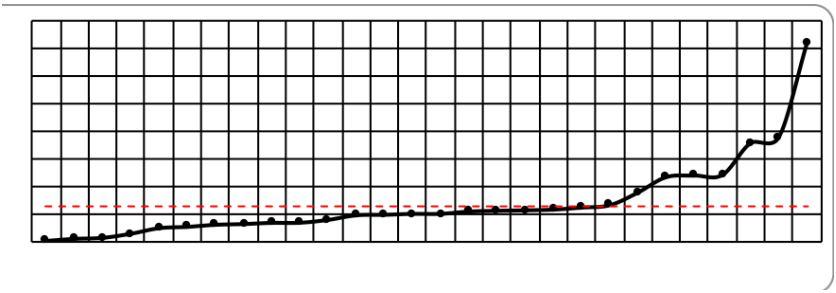


Fig.2. Rangement des verbes associés à concept « dommage » selon le coefficient d'association.

Par ailleurs, CRISCO a interprété le concept « personnel » en tant qu'adjectif, en conséquence, il était impossible d'utiliser la liste des synonymes obtenus pour le traitement dans les étapes suivantes.

Le dernier filtrage des candidats-termes, trouvés à l'étape précédente, se fait à l'aide des verbes qui forment le contexte formel de chaque concept : les candidats-termes qui partagent les mêmes verbes que le concept initial sont soumis à l'expert du domaine.

Les résultats obtenus montrent une précision supérieure à 84% dans tous les cas, sauf pour le mot « personnel » qui est, par erreur, interprété comme un adjectif dans le dictionnaire.

5. Conclusion

Nous avons présenté une méthode permettant d'enrichir une ontologie de noyau à partir d'un corpus et d'une ressource linguistique extérieure, type dictionnaire des synonymes. À notre avis, c'est la nouveauté de notre méthode : l'utilisation intensive de ressources lexicales extérieures (autres que le corpus des textes) dont le développement progressif a marqué les cinq dernières années, et qui ne sont pas encore suffisamment utilisées dans le domaine *d'ontology learning.*

Les valeurs de précision, relativement basses, s'expliquent du fait que les concepts génériques (comme *dommage,* par exemple*)* sont polysémiques. Par ailleurs, les vrais termes du domaine sont des locutions composées de plusieurs mots (groupes nominaux). Dès lors, pour réduire l'ambiguïté, il faudrait

d'affiner l'algorithme d'extraction des nouveaux termes par des patrons syntactiques et morphologiques [Orobinska et al., 2013]

Les critères de sélection des verbes significatifs jouent aussi un rôle important. Comme cela a été mentionné avant, les verbes sont souvent encore plus polyvalents que les noms ; il vaudrait donc créer les classes des verbes de même sens. Cela permettrait d'améliorer le rappel pour des termes détectés. Nos recherches se poursuivent actuellement dans cette direction.

La qualité du dictionnaire des synonymes joue un rôle crucial dans cette approche, c'est pourquoi nous aimerons souligner ici la cardinalité, la rigueur et la facilité d'utilisation de CRISCO.

Références

Amandine Périnet, Thierry Hamon, 2013 : Hybrid acquisition of semantic relations based on context normalization in distributional analysis. In Proceedings of the 10th International Conference on Terminology and Artificial Intelligence TIA 2013, pages 113-121, Paris, St Denis, 29-30 October.

Andrea Moro, Alessandro Raganato and Roberto Navigli. Entity Linking meets Word Sense Disambiguation : A Unified Approach. Transactions of the Association for Computational Linguistics (TACL), 2, 2014.

Aussenac-Gilles, N., Biébow, B. & Szulman, S. (2000) Revisiting Ontology Design : a methodology based on corpus analysis. Proceedings of the 12th *International Conference on Knowledge Engineering and Knowledge Management : EKAW'2000*, R. Dieng & O. Corby (eds), LNAI 1937, Springer, pp. 172-188.

B. Rosario and M. Hearst, 2001 : Classifying the semantic relations in noun-compounds via domain-specific lexical hierarchy. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP- 01), p. 82-90.

Bozsak E. (2002) : KAON - Towards a large scale Semantic Web. In Proceedings of the Third International Conference on E-Commerce and Web Technologies (EC-Web). Springer Lecture Notes in Computer Science.

Buitelaar P., Cimiano P., and Magnini B. (2008) Ontology Learning from Texts : An Overview./In Ontology Learning, from Text : Methods, Evaluation and Applications. / P. Buitelaar, P. Cimmiano, and B. Magnini, Eds. IOS Press. 2008.

Buitelaar Paul, Eigner Thomas. 2008. Topic Extraction from Scientific Literature for Competency Management. In Proceedings of the 3rd Expert Finder Workshop on Personal Identification and Collaborations : Knowledge Mediation and Extraction, October 2008, Karlsruhe, Germany.

Cécile Fabre and Didier Bourigault. 2006. Extraction de relations sémantiques entre noms et verbes au-delà des liens morphologiques. In TALN 2006, pages 121-129, Leuven.

Chklovski T. and Pantel P., 2004 : Verbocean : Mining the web for the fine-grained semantic verb relations. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP-04, pages 33-40, Barcelona, Spain.

Cimiano Ph., Honto A. & Staab S. (2005). Learning concept hierarchies from text corpora using formal concept analysis. In Journal of Artificial Intelligence Research (JAIR), volume 24, p. 305-339.

D. Moldovan, A. Badulescu, M. Tatu, D. Antohe, and R. Girju. 2004 : Models for the semantic classification of noun phrases. In Proceedings of the Computational Lexical Semantics Workshop at HLT-NAACL, 2004, pages 60-67, Boston, MA.

Eric Brill (1992) : A simple rule-based part of speech tagger. In Proceedings of ANLC '92, third conference on Applied natural language processing, 31 March-3 April 1992, Trento, Italy, P. 152-155.

Flati T. and Navigli R. : SPred : Large-scale Harvesting of Semantic Predicates. Proc. of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2012), Sofia, Bulgaria, August 4-9, 2013, pp. 1222-1232

François Morlane-Hondère and Cécile Fabre. 2012. Étude des manifestations de la relation de méronymie dans une ressource distributionnelle. In Proceedings of TALN'2012, pages 169–182, Grenoble, France.

Ganter B., Wille R. Formal Concept Analysis : Mathematical Foundations, 1st ed. Springer-Verlag New York, Inc., Secaucus. 1999

Helmut Schmid (2000): Unsupervised Learning of Period Disambiguation for Tokenisation. Internal Report, IMS, University of Stuttgart, May 2000

Maedche, A., Staab, S. : Mining Ontologies from Text. In : Proc. EKAW'2000, 2000

Michele Missikoff, Roberto Navigli and Paola Velardi. The Usable Ontology : An Environment for Building and Assessing a Domain Ontology. In Proceedings of the International Semantic Web Conference (ISWC), 2002, p. 39-53, Springer-Verlag DOI http ://dx.doi.org/10.1007/3-540-48005-6_6

Mondary T., Després S., Nazarenko A., Szulman S. La phase de conceptualisation : un état des lieux. In proceeding of IC 2008,19-èmes Journées Francophones d'Ingénierie des Connaissances, Nancy France.

Orobinska Olena, Chauchat Jean-Hugues, Charonova Natalya. 2013 : Enrichissement d'une ontologie de domaine par extension des relations taxonomiques à partir de corpus spécialisé. In Proceeding of 10th International Conference TIA 2013, Paris-13, October 28-30, P. 129-137, Web : http ://lipn.univ-paris13.fr ISBN :978-2-9174-9025-9

Paola Velardi, Stefano Faralli and Roberto Navigli : OntoLearn Reloaded : A Graph-Based Algorithm for Taxonomy Induction. In Computational Linguistics, 2012, Volume 39, Number 3 p. 665-707 DOI : http ://dx.doi.org/10.1162/COLI_a_00146

Peter Haase and Boris Motik : A Mapping System for the Integration of OWL-DL Ontologies In Axel Hahn, Sven Abels, Liane Haak, IHIS 05 : Proceedings of the first international workshop on Interoperability of heterogeneous information systems, P. 9-16. ACM Press, November 2005.

Philipp Cimiano, Christina Unger, John McCray. Ontology-Based Interpretation of Natural Language. Morgan & Claypool publishers. 2014. 155 pages

Philipp Cimiano, Cristina Unger et John McCrae. 2014. Ontology-based Interpretation of Natural Language (Synthesis Lectures Human language Technologies) In Morgan & Claypool Publishers March 2014, 178 pages, (doi :10.2200/S00561ED1V01Y201401HLT024)

R. Girju, D. Moldovan, M. Tatu, and D. Antohe, 2005 : On the semantics of noun compounds. In Computer Speech and Language, 19 :479-496.

Rodney Huddleston and Geoffrey K. Pullum : A Student's Introduction to English Grammar. Cambridge Univ. Press, 2006

Keywords : Domain Ontology Population, Information Extraction, NLP, Relation Recognition Rules, Instance Recognition Rules, Semantic Analysis, Radiation Protection

Mots clés : instanciation d'ontologie de domaine, extraction des informations, TAL, règles de détection des relations, règles de détection des instances, analyse sémantique, protection radiologique.

Dynamicité et formalisation. Une structuration à facettes de la Terminologie[1]

Natascia Leonardi

Université de Macerata
Dipartimento di Studi Umanistici –
Sezione di Linguistica, Letteratura e Filologia (SELLF)
Via Illuminati 4 – 62100 Macerata – Italie
natascia.leonardi@unimc.it
http://docenti.unimc.it/natascia.leonardi

Résumé : Les études et les applications de la terminologie contemporaine sont caractérisées par l'élaboration de cadres de représentation des connaissances formalisées qui visent à concilier les exigences opposées des utilisateurs humains et des agents artificiels, car ceux-ci sont souvent la double cible des Bases de Connaissances Terminologiques. Dans les principaux modèles de représentation terminologique, la structure de l'ontologie est intégrée à différents types de schémas, inspirés surtout par les théories de fonctionnement de la connaissance. On propose ici un modèle complémentaire à ceux qui existent déjà, visant à représenter efficacement la 'dynamicité' qui caractérise la terminologie. Ce modèle est conçu comme une mise en œuvre d'un système où la position théorique et sa réalisation applicative sont basées sur les mêmes principes.

Il est fondé sur une approche à facettes de l'analyse et de la représentation terminologique. Les facettes sont considérées comme une interprétation viable de la nature complexe et multidimensionnelle des Unités de Connaissances Terminologiques ainsi que comme une lecture efficace des perspectives – différentes, mais non contradictoires – sur les concepts que les spécialistes introduisent et partagent au sein d'un domaine de connaissance.

1 Ce texte est la traduction de *Dynamicity and formalization. A faceted structuring of Terminology* publié dans *AIDAinformazioni* 38/1-2, 2020, 95-115, qui est une version actualisée de la communication *Terminology and formalised knowledge representation frameworks. A proposal for a Faceted Structuring of Terminology (FaStTerm)*, présentée lors de la *Conference Toth* 2014.

1. Introduction

Dans le panorama des études terminologiques, il est possible de détecter un écart entre, d'une part, les besoins de la Terminologie en tant que discipline qui tient compte de la dynamicité et de la flexibilité des modèles linguistiques et cognitifs et, d'autre part, les exigences d'une représentation formelle rigide d'un domaine de connaissance, utilisant également un modèle formalisé, qui est principalement décrit comme une ontologie d'ingénierie. Cette analyse vise à passer en revue les méthodes de définition et de description terminologiques, ainsi que l'adéquation des modèles formalisés ayant été proposés pour organiser et représenter les propriétés des concepts et des désignations dans des référentiels destinés à la fois aux utilisateurs humains et aux ordinateurs. Ces dernières années, l'élaboration de collections terminologiques a également été de plus en plus influencée par la réalisation de systèmes de recherche d'informations linguistiques (spécialisées). Dans ce cadre, les applications informatiques peuvent également avoir un impact sur les théories terminologiques.

Cette analyse porte sur une sélection de schémas de représentation qui sont au centre des études terminologiques contemporaines. Il s'agit notamment ici de la Termontologie, de la Terminologie basée sur les Cadres (*Frame-based Terminology*) et de l'Ontoterminologie. Ces modèles peuvent être considérés comme étant représentatifs du choc entre les préoccupations descriptivistes et les modèles de formalisation sous-jacents pour la construction de référentiels de connaissances spécialisées. Ils présentent des caractéristiques communes, car ils s'appuient sur des théories caractérisantes de la terminologie et des méthodes de description, et ils sont tous orientés vers la construction de Bases de Connaissances Terminologiques (BCT). Cela implique que les cadres d'organisation et de représentation dans lesquels les données terminologiques sont distribuées peuvent être formalisés dans un schéma d'ontologie. De surcroît, les modèles de représentation qu'ils conçoivent montrent une tension entre la nécessité de rendre compte de la dynamicité caractérisant la terminologie sur son côté communicatif et la rigidité de la structure taxonomique soutenant le modèle ontologique. Le contraste qui apparaît s'explique par les besoins différents des utilisateurs humains et des applications artificielles, qui sont souvent la double cible des BCT.

La tentative de représenter la nature dynamique des Unités de Connaissances Terminologiques (UCT)[2] dans une BCT est un point clé de cette analyse. La variation est reconnue comme étant une caractéristique inhérente au langage spécialisé, à condition qu'elle traduise les choix délibérés des spécialistes afin d'exprimer des distinctions conceptuelles. Par conséquent, la dynamicité est un aspect pertinent à enregistrer dans une BCT, mais elle ne peut pas être facilement mise en correspondance avec les schémas formels. Au sein de la Terminologie, un système dynamique de définition et d'organisation constitue un avantage, car il n'impose pas une vision monolithique et univoque des concepts et des désignations, mais il permet la représentation de perspectives multiples sur les deux éléments. En conséquence, on propose ici un modèle d'analyse et de représentation terminologique qui est basé sur une interprétation à facettes des UCT. Celui-ci vise à rendre compte de tous les phénomènes de variation qui sont pertinents pour leur définition dans un domaine spécialisé et il s'avère compatible avec la formalisation requise par les applications informatiques. En fait, il semble approprié de représenter la dynamicité des domaines de connaissance spécialisés par un système qui repose sur des principes dynamiques et qui se concentre sur l'attribution aux entités analysées de tous les critères de définition et propriétés pertinents, indépendamment de leur compatibilité avec un schéma taxonomique unique. Cette approche est définie comme un processus aboutissant à une identification et à une définition claires mais non absolues et statiques des UCT.

2. Dynamicité et systèmes formalisés

Ces dernières années, l'importance de la variation terminologique a été largement reconnue comme étant une ressource plutôt qu'un simple obstacle à la description des connaissances spécialisées, comme le prouvent les nombreuses études consacrées à ce sujet, dans les approches aussi bien intra- qu'inter-linguistiques.[3] Par « dynamicité », on entend ici la variation concernant le contenu et la forme des UCT. Comme c'est le cas dans la langue générale, la variation caractérisant la terminologie peut être considérée comme sa

2 Cette désignation est utilisée ici pour se référer aux termes en tant qu'unités relationnelles dans lesquelles la matière linguistique, mentale et référentielle se fondent ; elles peuvent subir des variations (dans le temps, l'espace, les cultures, les paradigmes, etc.) et leur interaction façonne un domaine de connaissance.

3 Parmi les plus représentatives, on peut citer : Ibekwe-Sanjuan (1998) ; Freixa (2006) ; Temmerman et Van Campenhoudt (2011) ; Condamines (2010) ; Pecman (2014) ; Drouin *et al.* (2017) ; Fernández Silva (2019).

dimension naturelle, plutôt que comme un problème à résoudre par un processus de normalisation. Ainsi, ce qui sera considéré ici comme une variation *propre* n'est pas un symptôme de négligence ou d'irrégularité dans l'utilisation des termes, mais le résultat d'un choix terminologique délibéré fait par des spécialistes (ou une communauté de ceux-ci) et motivé par des besoins communicatifs et cognitifs spécifiques.

La dynamicité implique une interaction des connaissances conceptuelles, linguistiques et référentielles d'un domaine spécialisé (Cabré 2003). Les connaissances référentielles ne correspondent pas simplement à des données prétendument « factuelles », mais impliquent les connaissances situationnelles et contextuelles du domaine spécialisé, qui sont liées à des connaissances expérientielles.[4] Cela implique que la théorie et la pratique de la terminologie sont axées sur la nature des termes en tant qu'*unités à multiples facettes*, qui sont utilisées dans la communication spécialisée, comme il ressort, avec des interprétations légèrement différentes, par exemple, chez Sager (1990, 22 ; 57), Cabré (1996, 19) et Temmerman (2000, 224). Dans les UCT, la nature linguistique, conceptuelle et référentielle d'un domaine de connaissance spécialisé se fondent. Par conséquent, leur inventaire est considéré dans une perspective articulée car, d'une part, il s'agit d'un instrument de communication des connaissances de façon claire et, d'autre part, la terminologie réside également dans les textes, où elle se caractérise par une dimension conceptuelle et linguistique qui est dynamique et ne peut être facilement sacrifiée sur l'autel de l'univocité entre le concept et la désignation.

Il s'ensuit que la variation peut être considérée comme étant un indicateur d'une différenciation conceptuelle qui existe dans l'esprit des spécialistes produisant un message. Cela conduit à la coexistence de différentes variantes (tant formelles que conceptuelles) au sein de l'inventaire des lexiques spécialisés de plusieurs domaines technico-scientifiques. De ce point de vue, une représentation terminologique qui éliminerait la variation par la standardisation afin de poursuivre l'idéal d'univocité déterminerait la négation des distinctions conceptuelles pertinentes que les spécialistes font et signalent délibérément par leur usage du langage. Dans une approche descriptiviste, c'est la dimension dynamique des termes qui doit être prise en compte, afin de favoriser non seulement une production communicative claire et économique ainsi qu'une réception efficace, mais aussi une représentation et une transmis-

4 Cette approche peut s'expliquer par les concepts de cognition *experiential*, *embodied* et *situated* qui appartiennent aux vues des théories cognitivistes (Lakoff 1987 ; Barsalou 2008).

sion terminologiques qui s'appuient sur la réalité complexe d'un domaine de connaissance spécialisé.

Cette tension entre l'objectif de régularisation et l'objectif descriptif aboutit à un choc apparemment radical au sein des modèles terminologiques récents, qui dépend principalement de deux raisons de nature différente. D'un point de vue théorique, les ouvrages de référence spécialisés constituent un standard pour l'usage terminologique, même lorsqu'ils sont orientés vers la description de la nature complexe des termes. D'un point de vue méthodologique, les modèles et les dispositifs utilisés pour produire des collections terminologiques, principalement sous la forme de BCT, imposent des restrictions dues aux technologies adoptées. Par conséquent, les logiciels disponibles, le modèle d'ontologie et les paradigmes analytiques sous-jacents influencent fortement le profil final des BCT.

Le panorama contemporain de la recherche terminologique semble montrer l'émergence d'une double orientation au sein de la sphère descriptiviste, l'une principalement orientée vers les utilisateurs humains et l'autre vers les applications informatiques. La première ne s'écarte pas beaucoup de la conception de Wüster (1974, 98) de l'« informatique » comme instrument de la pratique terminographique. Cette approche utilise l'informatique pour favoriser la description et la représentation de l'information terminologique. D'autre part, la sous-branche descriptiviste de la terminologie qui est principalement orientée vers les applications informatiques, après une première tentative d'optimisation de la représentation computationnelle des termes et de leurs désignations, risque aujourd'hui de subordonner la composante communicative et dynamique de la terminologie aux exigences de l'ontologie formelle (Geentjens *et al*. 2006). En fait, l'*ingénierie ontologique* utilise les langages naturels pour l'élaboration de vocabulaires de termes et de leurs définitions, approchant ainsi la pratique prescriptiviste de la construction d'une taxonomie définitionnelle des associations univoques de concepts et de désignations standardisés.[5] L'« ontologie » mentionnée ici appartient au domaine de l'informatique, où elle peut être brièvement définie comme « a formal, explicit specification of a *shared* conceptualization » [soulignement ajouté] (Gruber 1993, 199 ; Borst 1997, 12). Plus précisément, une ontologie est « an enginee-

5 En effet, les objectifs et la portée de la Terminologie et ceux de l'ingénierie des ontologies sont différents, comme le soulignent, par exemple, Kerremans et Temmerman (2004) ; Grabar *et al*. (2012, 375-376 ; 379) ; Roche *et al*. (2009). Néanmoins, les frontières qui séparent les deux domaines deviennent de plus en plus subtiles et glissantes, comme le soulignent, par exemple, Faber et López Rodríguez (2012, 21).

ring artefact, constituted by a specific vocabulary used to describe a certain reality, plus a set of explicit assumptions regarding the intended meaning of the vocabulary words » (Guarino 1998, 4). Le « specific vocabulary » (vocabulaire spécifique) est organisé en une taxonomie et enrichi par des restrictions, ou des définitions formelles, et des termes du langage naturel. Ce vocabulaire représente le lien central entre la Terminologie et l'informatique via le modèle de l'ontologie, avec lequel le vocabulaire est souvent incorrectement identifié.

2.1. Théories de la terminologie contemporaine et modèles de formalisation

La nécessité de concevoir un schéma d'ontologie est commune aux projets terminologiques analysés ici, malgré les différents schémas complémentaires qui fournissent des informations linguistico-communicatives. Ce modèle partage avec le cadre relationnel traditionnel décrit par Wüster l'arrangement taxonomique définitionnel ainsi que la dérivation de la méthode de catégorisation d'origine aristotélicienne. Tant dans l'ontologie que dans la structure relationnelle terminologique wüsterienne, le schéma taxonomique sert à la définition des concepts, c'est-à-dire des abstractions universellement valables de classes d'entités. Les définitions découlent de l'identification des caractéristiques essentielles (propriétés et relations) des concepts. Par conséquent, le schéma taxonomique est censé organiser et représenter l'ensemble des connaissances du domaine partagé par une communauté de spécialistes.

La validité de ce modèle en tant qu'instrument de représentation des connaissances a été remise en cause directement dans la branche du cognitivisme (Fillmore 1975 ; Lakoff 1987 ; Langacker 1987 ; Taylor 1989) ainsi que dans les études anthropologiques et sociologiques orientées vers le fonctionnalisme. Cette position a mis en perspective l'application du modèle de classification traditionnel également aux théories linguistiques et terminologiques. Les approches descriptivistes de la terminologie avalisent des méthodes de définition et de représentation qui vont au-delà du schéma taxonomique, car cela limite la réalisation d'une organisation terminologique satisfaisante.

Les projets examinés ci-dessous sont des exemples représentatifs de la tentative de faire face à la dimension dynamique de la terminologie et aux exigences rigides des ontologies formalisées nécessaires aux applications informatiques. Différentes solutions à cette tension sont proposées, en fonction des approches théoriques sous-jacentes et des problèmes spécifiques sur lesquels chacune d'entre elles est axée. Afin de développer une BCT qui peut

être utilisée à la fois par les humains et les ordinateurs, ces projets s'appuient sur le modèle de base de l'ontologie formelle pour l'arrangement des données terminologiques.[6] Ce schéma est assorti et complété par différents types de structures dans lesquelles sont organisées les informations pertinentes sur la dimension dynamique des BCT.

La Termontographie (Temmerman et Kerremans 2003) a été élaborée en tant qu'aspect applicatif de l'approche Sociocognitive (Temmerman 2000). Elle combine les questions théoriques de l'analyse Sociocognitive avec la méthode de représentation de l'ingénierie ontologique (Temmerman et Kerremans 2003, 4 ; Geentjens *et al*. 2006). Dans ce cadre, les *modèles d'Unités de compréhension* s'inspirent des principes cognitivistes. Les modèles sont élaborés pour compléter les informations conceptuelles hiérarchiquement structurées avec le type de données qui peuvent rendre explicites toutes les caractéristiques qui ne correspondent pas au schéma d'ontologie, mais qui sont pertinentes pour la définition d'une unité terminologique structurée de manière prototypique (Temmerman 2000, 224). Dans les applications ultérieures, comme le soulignent Faber et López Rodríguez (2012, 21), la Termontographie semble s'être écartée de la position sociocognitive initiale, la composante ontologique formelle étant devenue dominante.

Animés par une logique similaire à celle de Temmerman, Faber et son groupe de recherche introduisent une approche de la Terminologie basée sur les Cadres (Faber 2012), où les relations conceptuelles sont organisées au sein d'*événements de domaine* qui ont une structure prototypique. Les événements de domaine se concentrent sur les actions et les processus qui instancient le domaine de connaissance analysé (Faber *et al*. 2009, 5 ; San Martín et León Araúz 2009, 13). Cette structure révèle une forte concentration sur la nature concrète (c'est-à-dire pragmatique-encyclopédique) et linguistique des termes (León Araúz *et al*. 2012). Le modèle d'ontologie reste la « colonne vertébrale technologique » de la BCT, tandis que les informations terminologiques (linguistiques, conceptuelles, référentielles-pragmatiques) sont organisées en événements de domaine. Ici, les catégories génériques d'un domaine de connaissance sont identifiées et réparties dans des modèles connexes qui fournissent un cadre de description cohérent pour les concepts spécifiques ;

6 La Théorie Communicative de la Terminologie développée par Cabré (1999) n'est pas examinée ici, car si ses fondements théoriques sont pertinents pour le développement de la Terminologie descriptiviste, elle n'est pas explicitement orientée vers la mise en œuvre d'un modèle applicatif pour la représentation terminologique dans les BCT (à l'exception partielle de Cabré 2004 et de Cabré *et al*. 2004).

ceux-ci sont liés par des relations hiérarchiques et non hiérarchiques (Faber et López Rodríguez 2012, 26-27).

Les positions prises par Temmerman et Faber visent à satisfaire, dans un modèle unique, le modèle de formalisation de l'informatique et les besoins communicatifs des utilisateurs des BCT. En fait, tant la Termontographie que la Terminologie basée sur les Cadres sont censées intégrer l'approche descendante et ascendante de l'analyse et de la représentation des termes (Temmerman et Kerremans 2003, 4 ; Geentjens *et al.* 2006 ; Faber *et al.* 2009, 6). L'ontologie agit comme une structure fonctionnelle pour l'application des matériaux terminologiques aux diverses applications TALN et informatiques en matière de gestion et de recherche de connaissances (spécialisées).

Les hypothèses cognitivistes ne sont pas partagées par le troisième modèle terminologique considéré ici, l'Ontoterminologie (Roche 2007), qui diffère des autres, tout d'abord parce qu'il peut être directement lié à la tradition de normalisation découlant de la Théorie Générale de la Terminologie et de la logique de l'ISO/TC 37 et, par conséquent, donne un rôle central à la structure ontologique et aux relations taxonomiques dans le travail terminologique (Roche 2007, 14 ; Roche *et al.* 2019). Dans cette approche, les sphères linguistiques et conceptuelles des connaissances spécialisées sont considérées comme distinctes. Puisque les termes et leurs désignations sont supposés être différents par nature, ils sont analysés et représentés selon des modalités distinctes, en faisant correspondre l'ontologie conceptuelle et définitionnelle aux descriptions linguistiques des désignations correspondantes (Roche 2012, 2628). En fait, avec l'Ontoterminologie, Roche élabore un modèle où le cadre relationnel des concepts spécialisés est le référentiel de connaissances essentielles, organisé selon les règles de l'ontologie formelle. Le cadre de l'ontologie correspond parfaitement aux exigences des définitions formelles pour les applications informatiques. Dans l'Ontoterminologie, l'utilité des informations linguistiques pour des communautés spécifiques d'utilisateurs humains est reconnue et apparaît dans des définitions linguistiques supplémentaires tirées de sources spécialisées. Par conséquent, ce schéma sépare les procédures ascendantes et descendantes, qui sont complémentaires dans la Termontologie et dans l'approche de la Terminologie basée sur les Cadres (Roche 2012, 2626).

3. Une approche à facettes de la Terminologie. Un modèle d'analyse et de représentation terminologique

La proposition décrite ici est un modèle d'interprétation et de représentation qui s'éloigne des axiomes occidentaux de la catégorisation, dans le but de fournir une approche qui unit, de façon cohérente, la manière dont les éléments de connaissance sont analysés et formellement représentés. Cette approche est provisoirement appelée « Structuration à facettes de la terminologie » (*Faceted Structuring of Terminology, FaStTerm*), car il s'agit d'un modèle d'interprétation qui utilise l'approche à facettes pour la construction d'une BCT. L'intention est d'aller au-delà des restrictions imposées par l'organisation taxonomique des propriétés et des relations des concepts mais, en même temps, de concevoir un cadre compatible avec le schéma ontologique nécessaire aux applications informatiques. Ainsi, ce modèle à facettes de représentation et de description terminologique, comme les autres considérés ici, tente de faire face à la fois à la multidimensionnalité et à la dynamicité des termes et aux restrictions imposées par la formalisation des données conceptuelles et linguistiques.

Les principes de la classification à facettes trouvent leur origine dans les Sciences de l'Information et des Bibliothèques (SIB). Elles ont été élaborées pour la première fois dans les années 1950 par Ranganathan en tant que « dynamic theory of classification » (Ranganathan 1969, 193), où le domaine global de la connaissance est considéré comme étant un continuum d'informations qui peut être « remodelé » de manière dynamique selon le point de vue à partir duquel il est considéré.[7] Cette approche est adoptée dans la FastTerm comme une source d'inspiration plutôt que comme un cadre directement applicable au domaine de la Terminologie. Voilà pourquoi les spécificités qui caractérisent cette méthode au sein des SIB ne seront pas analysées en détail ici.[8]

La classification à facettes est une approche qui peut aborder la *connaissance* dans le sens le plus large du terme, plutôt que se concentrer sur ses aspects individuels. Par conséquent, lorsqu'elle est transposée au domaine de

7 Pour un aperçu plus complet de la classification à facettes en tant qu'instrument d'organisation des connaissances dans les SIB, voir Hjørland (2013). La relation interdisciplinaire de la classification à facettes est traitée en détail chez Gnoli (2008) et Szostak *et al*. (2016).

8 Dans les SIB, l'approche à facettes n'est pas monolithique, comme le souligne par exemple Tennis (2012). Vickery (1966) est un autre représentant important de cette approche dans les SIB. En outre, dans le cadre de la théorie de Ranganathan, ce modèle a également subi des modifications au fil du temps (Tennis 2012, 302-307).

la théorie et des applications de la Terminologie, elle fournit un paradigme valable pour analyser et représenter les différentes facettes interdépendantes des UCT, c'est-à-dire que leurs aspects conceptuels, linguistiques et factuels sont considérés comme étant des composantes intégrées. Dans les SIB, ce modèle a été conçu pour optimiser l'organisation d'objets de connaissance complexes, tels que les livres, plutôt que pour des applications disciplinaires spécifiques. Ainsi, alors que les théories de l'analyse linguistique ou du fonctionnement des connaissances humaines jouent un rôle dans la délimitation des autres travaux terminologiques examinés ici, le modèle conçu pour la FaStTerm montre un manque de « spécificité disciplinaire » et est principalement orienté vers l'identification d'un système de représentation terminologique parallèle à la méthode d'analyse utilisée pour les UCT.

Les principes des sciences et de la linguistique cognitives caractérisent à la fois la Termontologie et la Terminologie basée sur les Cadres, cette dernière étant également étroitement liée à la théorie sémantique de Fillmore (1976) et à son application à l'analyse et à la représentation lexicales (Fillmore et Atkins 1998 ; Faber *et al*. 2009 ; Faber 2012). La notion de *facette* a été introduite en sémantique cognitive par Cruse (1995) pour différencier les relations sémantiques à facettes de la polysémie (du type ici défini « apparent » ; voir n. 10). Au sein de la sémantique, les facettes sont identifiées comme suit : « Facets are distinguishable components of a global whole, but they are not capable of being subsumed under a hyperonym » (Croft et Cruse 2004, 116) et sont liées à une « dynamic construal approach to categorization » ainsi qu'à un « dynamic construal of meaning » (Croft et Cruse 2004, 92-94). Il est possible de repérer des références à la notion de facettes sémantiques de Cruse dans l'approche basée sur les Cadres de Faber (Faber et López Rodríguez 2012, 26 ; Faber et Ureña 2012, 91).[9] Temmerman (2000, 67 ; 74 ; 96 ; 121-122 et *passim*) associe « facette » à la « intercategorial or internal structure », c'est-à-dire que les aspects ou attributs d'une unité de compréhension sont déterminés par ses relations inter-catégorielles avec d'autres unités, qui délimitent une perspective spécifique de l'unité en question. Cette perspective peut être associée à la lecture faite par Kageura (1997, 121) des facettes en tant que « types of characteristics ». L'approche à facettes présentée ici est compatible avec la position de Cruse – c'est-à-dire que l'accent est mis sur la valeur globale des UCT et leurs caractéristiques relationnelles. Néanmoins, l'approche FaStTerm n'est pas limitée à une interprétation sémantique des UCT. Même si la frontière

9 Pour une analyse de la relation entre les facettes et les cadres, voir Green (2017).

entre le concept et le sens n'est pas nette, les UCT sont considérées comme étant dotées d'un degré d'intension plus élevé par rapport au contenu des mots de la langue générale ; leur valeur n'est donc pas identifiée au seul sens. Ce modèle permet d'appliquer le même cadre à la complexité des UCT, conçues comme sémantiques, référentielles et conceptuelles par nature. Les facettes sont donc compatibles avec l'analyse linguistique, mais ne sont pas les instruments d'une interprétation de la Terminologie exclusivement linguistique.

Cette approche vise principalement à fournir une analyse et une représentation cohérentes des instances qui correspondent à une variation intentionnelle (formelle et/ou conceptuelle) dont le but est de satisfaire les besoins communicatifs et cognitifs des spécialistes. Dans les textes d'un domaine de connaissance, les désignations et leurs contenus ne présentent pas nécessairement un caractère statique et univoque, mais sont souvent des unités complexes à multiples facettes qui ont une vie dans le discours spécialisé (cf. exemple ci-dessous § 3.1). Avec une approche à facettes de l'analyse et de la représentation terminologiques, il pourrait être possible de rendre compte des théories concurrentes au sein d'un domaine et des relations interdisciplinaires qui contribuent à déterminer les phénomènes de ce qui peut être considéré comme une polysémie et une synonymie « apparentes » des UCT.[10] En effet, une même désignation peut correspondre à différents concepts, ou différentes désignations peuvent être associées à un même concept en fonction, par exemple, des spécialistes et/ou des courants de pensée ou des domaines interdisciplinaires dans lesquels les unités terminologiques sont introduites et utilisées.[11]

Une analyse à facettes de la terminologie permet non seulement d'établir une corrélation, mais aussi de « contextualiser » les UCT dans leur réalité spécifique de discours et/ou de (sous-)domaine de connaissance, qui est délimité par une communauté de spécialistes, située dans une zone et une période de temps spécifiques. Dans la FaStTerm, les UCT ne sont pas classées selon la perspective catégorielle traditionnelle, mais organisées selon une analyse et une représentation à facettes, qui nous permet d'« attribuer » les caracté-

10 Une distinction entre polysémie « propre » et « apparente » peut être établie en référence à l'utilisation délibérée ou accidentelle (et éventuellement stylistique) de variantes du contenu et/ou de forme dans la communication spécialisée. Ici, la polysémie et la synonymie apparentes sont mises en contraste avec la *variation propre* examinée ci-dessus (voir § 2.).

11 Un exemple clair pourrait être le terme « ontologie » tel qu'il est utilisé dans la Terminologie, dans l'informatique et dans le domaine interdisciplinaire qui est issu de leur interaction.

ristiques pertinentes aux UCT et de structurer les éléments définis dans un cadre qui supporte de multiples hiérarchies. Cela se fait par l'identification d'une série de schémas de propriétés et de relations qui représentent les paramètres de définition pertinents. Ainsi, plutôt que d'être interprétée comme une entité absolue au sein d'une hiérarchie conceptuelle complète du domaine global, une UCT peut être considérée selon la perspective « relative » qui est pertinente dans des connaissances et des contextes linguistiques spécifiques. L'organisation hiérarchique (et ontologique) passe des concepts aux propriétés et aux relations. En fait, dans une ontologie formelle, les différentes facettes du « concept » qui ressortent des théories examinées ci-dessus seraient en conflit. Pourtant, ces facettes coexistent dans la connaissance et la pratique communicative des spécialistes de ce domaine.

3.1. Une étude de cas

Afin d'illustrer les types de problèmes qui pourraient se poser dans l'analyse et dans la représentation des UCT pour la réalisation d'une BCT, nous allons maintenant examiner un exemple. Il se penche sur la question de la variation terminologique et, en particulier, sur la manière dont la polysémie apparente peut être identifiée et traitée par une analyse à facettes d'une UCT. L'exemple est tiré du domaine de la Terminologie, avec une référence particulière à l'une des théories considérées ci-dessus. Il ne s'agit pas d'une discussion sur la terminologie de l'auteur, mais elle vise à illustrer la variation factuelle des UCT ainsi que les conséquences sur la terminologie générale et sur sa représentation formelle. Le domaine de la Terminologie a été choisi pour deux raisons : tout d'abord, il représente bien un domaine (interdisciplinaire) des sciences humaines et présente un degré élevé de dynamicité ; en outre, sa terminologie se caractérise par un standard officiel, qui peut constituer un critère intéressant pour l'analyse des variations.

Dans ce domaine de connaissance, une UCT clé est un « concept », mais sa centralité n'implique pas son univocité : elle a différentes définitions, qui dépendent de plusieurs facteurs, parmi lesquels la théorie spécifique, l'école de pensée, la fonction cognitive, la langue, la culture et la période de temps pendant laquelle elle est façonnée et utilisée. Toutes ces définitions sont également pertinentes et valables dans le domaine de connaissance de la Terminologie, car elles sont basées sur des priorités inhérentes et des relations spécifiques (définies par les critères de variation), qui déterminent le profil des différentes valeurs du « concept ». Les différentes facettes de cette UCT ne

s'excluent pas mutuellement, mais coexistent dans la Terminologie, au point de pouvoir se retrouver dans un même texte spécialisé.

Dans une BCT basée sur une organisation ontologique, la présence de valeurs distinctes de « concept » ne peut pas être facilement résolue en les traitant comme des homographes, car même si elles représentent des contenus de connaissance différents et se distinguent par des types de critères (partiellement) distincts, elles ne peuvent pas être agencées dans une taxonomie définitionnelle unique. Néanmoins, toutes les perspectives dont découlent les différents contenus sont également pertinentes dans le domaine de la Terminologie et doivent être enregistrées dans une BCT.

La définition de « concept » donnée dans la documentation ISO[12] est cohérente avec la théorie de Wüster et avec l'utilisation de ce terme dans les travaux de Roche ;[13] mais elle acquiert une valeur différente dans les écrits de Temmerman et Faber. Le « concept » est conçu comme une représentation mentale des objets, une abstraction de leurs propriétés communes, qui appartient au domaine de conceptualisation de l'individu ainsi qu'au système de connaissances qui constitue le domaine partagé (ISO 704 :2000, § 5.1 ; § 5.3.1.). Il peut être considéré comme étant une entité centrale et reconnue dans la terminologie prescriptiviste des différentes époques. Toutefois, dans les approches cognitivistes et plus généralement descriptivistes de la Terminologie, la nature du « concept » et son rôle dans ce domaine de connaissance changent. La discrétion catégorielle du « concept » est complétée ou remplacée par une interprétation prototypique, comme le montre clairement la théorie de Temmerman.[14]

En effet, le terme « concept » est utilisé dans ses textes, mais la plupart du temps avec une valeur différente de celle spécifiée par la norme ISO : il *s'agit d'un type d'*« Unité de Compréhension » avec une structure non-prototypique associé au type « catégorie », qui est par contre doté d'une structure prototypique (Temmerman 2000, 224). Le système conceptuel de la Terminologie est restructuré, l'« Unité de Compréhension » jouant le rôle de notion cen-

12 Dans la norme ISO (ISO 704 :2000), les concepts sont définis comme des « mental representations of objects within a specialized context or field » (§ 5.1.) ; « the concept shall be viewed not only as a unit of thought but also as a unit of knowledge » (§ 5.3.1.).

13 La définition présente dans la théorie de Roche est tirée de la norme ISO : « The concept in terminology – an [sic] 'unit of knowledge created by a unique combination of characteristics' [ISO 1087-1] » (Roche 2012, 2627).

14 Voici un exemple de déclaration explicative et programmatique : « The main key word of traditional Terminology was concept […]. In Sociocognitive Terminology it is to be replaced *by unit of understanding* » (Temmerman 2000, 223).

trale selon Temmerman. Au contraire, les termes « catégorie » et « Unité de Compréhension » ne sont pas pertinents dans la norme ISO pour identifier le contenu des termes. En outre, dans le même volet théorique, Faber *et al*. (2009) utilisent le terme « concept » pour désigner également une entité dotée d'une structure prototypique. Mais dans l'approche basée sur les Cadres, c'est la dimension relationnelle des concepts qui est principalement mise en exergue, ainsi que leur valeur sémantique et leur appartenance à un processus ou à un événement représenté par des Cadres. Cette notion de « concept » s'écarte donc en partie de celle de Temmerman et, plus radicalement encore, de la définition de l'ISO.

Cet exemple montre à quel point la variation et la dynamicité du contenu et de la forme des UCT peuvent être importantes dans un domaine, où les caractéristiques distinctives et pertinentes d'un concept peuvent varier, selon l'approche théorique. Les différentes perspectives donnent lieu à la coexistence de différentes vues du « concept », caractérisées par des valeurs distinctes au sein d'un même (sous-)domaine de connaissances. Cela empêche une définition et une représentation linéaire et univoque des UCT dans une BCT. Par conséquent, une représentation taxonomique et une formalisation de ce type de situation au sein d'un domaine spécialisé ne seraient pas facilement réalisables.

Inversement, une approche à facettes pourrait permettre une représentation des UCT par ce qui pourrait être défini comme une « contextualisation sélective » ou une « restriction de la focalisation ». Cela est déterminé par la sélection de paramètres spécifiques (c'est-à-dire des propriétés et des relations) qui nous permettent de délimiter le concept particulier associé à une forme de terme, en le découpant du système global selon l'ensemble des caractéristiques qui sont pertinentes d'un point de vue spécifique. L'application de ce modèle nécessite de s'éloigner des systèmes traditionnels de classification et de représentation, qui limiteraient les paramètres relationnels à ceux autorisés par la structure hiérarchique des schémas ontologiques.

Dans l'exemple de « concept » considéré ici, différents cadres de propriétés peuvent être axés sur la période de temps et, pour ce qui est des branches théoriques spécifiques et des théoriciens d'une discipline, des sous-domaines peuvent être identifiés et éventuellement liés à des disciplines voisines et à l'objectif pour lequel concept a été formé, etc. De cette façon, celles qui sont généralement identifiées comme des relations non hiérarchiques dans la Terminologie peuvent être enregistrées et représentées dans le même modèle formel que les relations hiérarchiques. La vue à facettes n'empêche pas de

délimiter les relations hiérarchiques traditionnelles de genre/espèce et de partie/ensemble. Néanmoins, dans l'organisation des unités de connaissance dans une BCT à facettes, ces types de relations n'ont pas un rôle primordial par rapport aux autres.

Ce modèle d'analyse et de représentation nous permet de nous concentrer à la fois sur les différentes facettes des UCT sans déclencher de conflits entre elles, car elles peuvent se chevaucher partiellement et/ou être liées entre elles, mais sont substantiellement indépendantes les unes des autres, chacune d'entre elles étant façonnée dans un contexte de connaissance spécifique. Une analyse à facettes s'applique non seulement à la valeur « purement » conceptuelle des termes, qui est conçue comme abstraite et universellement partagée (si jamais elle peut être identifiée avec précision), mais aussi à son interaction avec les aspects linguistiques et référentiels des UCT, car elles contribuent toutes à la constitution d'une UCT dans un domaine de connaissance.

3.2. Aspects applicatifs et travaux futurs

Le défi concerne les aspects applicatifs de ce projet, car le système FaStTerm d'analyse et de représentation des unités terminologiques présuppose la possibilité d'avoir une taxonomie multiple. Cela entre en conflit avec les procédures de formalisation qui caractérisent l'ontologie formelle, car cette dernière nécessite l'utilisation d'un seul schéma de classification hiérarchique. La solution possible a été identifiée dans les études de Giunchiglia et de son équipe de recherche, qui ont commencé par le modèle dit *Faceted Lightweight Ontology* (FLO)[15] et développé en DERA (*Domain, Entity, Relation, Attribute*), « a faceted knowledge organization (KO) framework (*a classificatory structure for developing knowledge sources*) » (Giunchiglia et Dutta 2011 [soulignement dans l'original]). Ce modèle, conçu pour l'organisation des données de connaissance, leur gestion et leur récupération, peut être appliqué à tout domaine de connaissance et permettra, espérons-le, une formalisation satisfaisante d'une BCT basée sur une approche à facettes.

Ce cadre d'organisation et de formalisation des connaissances est compatible avec la FaStTerm tant d'un point de vue théorique qu'applicatif. Il s'ap-

15 « *Lightweight ontologies* are ontologies with a tree structure where each node is associated a natural language label [sic]. *Faceted lightweight ontologies* are lightweight ontologies where the labels of nodes are organized according to certain predefined patterns which capture different aspects of meaning, i.e., *facets* » [soulignement dans l'original] (Giunchiglia *et al*. 2009, 36).

puie sur les études de Giunchiglia sur le raisonnement contextuel (Giunchiglia 1993 ; Giunchiglia *et al*. 2012 ; Das et Giunchiglia 2016) et s'est développé dans le sens d'un instrument efficace et dynamique de représentation et de récupération des connaissances (Giunchiglia *et al*. 2014, 44-56 ; Das et Roy 2016 ; Das 2018). Dans la FLO, l'analyse à facettes est rattachée à une « ontologie légère » (Giunchiglia et Zaihrayeu 2007), qui garantit une représentation formelle adéquate tout en imposant moins de limites à l'organisation des éléments de connaissance. Ainsi, ce modèle, tout comme les ontologies descriptives (Giunchiglia *et al*. 2014), est plus flexible, même s'il est moins expressif que les ontologies formelles propres visant la classification. En fait, cette approche unit et améliore le potentiel d'organisation et de définition des ontologies formelles légères et les avantages de l'approche de classification à facettes – telle que définie dans les SIB et dans les études sur l'organisation des connaissances – avec la rigueur et l'expressivité des ontologies descriptives pour la représentation des connaissances. Qui plus est, le cadre DERA représente un modèle formel qui est prévu pour être pleinement intégré aux connaissances incorporées dans le langage naturel. Son application permet de faire passer la structure ontologique de l'organisation des contenus des UCT aux propriétés et aux relations qui les définissent. Comme pour les projets mentionnés ci-dessus (§ 2.1) dans cette proposition, l'ontologie est principalement interprétée comme un « dispositif » de calcul, mais ici la définition des UCT est donnée à travers des facettes structurées. Les propriétés intensionnelles peuvent être structurées en Attributs et en Relations qui sont identifiés comme spécifiques du domaine analysé. Par exemple, dans le cas du « concept » évoqué ci-dessus, « la théorie, l'école de pensée, la fonction cognitive, la langue, la culture et la période de temps » peuvent être traités comme des attributs, avec des valeurs spécifiques, du modèle DERA et contribuer à constituer une facette définissant une entité.

La phase actuelle du projet FaStTerm est centrée sur la vérification de la mesure dans laquelle la position méthodologique d'une approche à facettes de l'analyse et de la représentation terminologiques peut être appliquée efficacement à une BCT par le biais du modèle DERA. La tentative d'accomplir cette tâche est une entreprise complexe, car la FLO a été développée au sein des applications informatiques en vue de la gestion et de la récupération des connaissances (linguistiques), qui sont également orientées vers les technologies du Web Sémantique. Cependant, une BCT exige que les besoins spécifiques de la Terminologie soient satisfaits, c'est-à-dire les besoins d'une

discipline qui est axée sur l'organisation des connaissances humaines (spécialisées).

4. Conclusions

L'analyse des approches méthodologiques dans la théorie et dans la pratique de la Terminologie permet d'identifier un noyau commun dans les projets qui sont poussés par des besoins et des hypothèses différents et, par conséquent, par le développement de cadres distincts de représentation et de définition terminologiques. Cette caractéristique commune vise à surmonter la difficulté de faire correspondre un noyau ontologique avec des principes de définition et de description terminologiques, en particulier lorsqu'ils sont appliqués à la dimension dynamique des UCT. Les solutions proposées dans les projets examinés ici vont dans le sens d'une combinaison de questions contrastées ayant trait à la dynamicité et à la multidimensionnalité de la Terminologie descriptive avec les règles de formalisation rigoureuses imposées par l'ingénierie ontologique. Cela se fait en associant le cadre de l'ontologie à différents types de modèles de représentation, la plupart inspirés des théories du fonctionnement des connaissances.

Une approche alternative de l'analyse terminologique et de sa représentation formelle est présentée ici comme une tentative d'identifier une position interprétative caractérisée par la cohérence entre les questions théoriques et applicatives. La FaStTerm est présentée comme une approche à facettes de la Terminologie qui ne considère pas les UCT comme des entités monolithiques dotées de définitions uniques et fixes, mais comme des unités complexes de connaissances qui peuvent éventuellement avoir une nature multidimensionnelle. La FaStTerm n'établit pas seulement une corrélation, mais permet également une contextualisation des UCT. De cette manière, la valeur spécifique de chaque unité peut être soulignée par l'identification de paramètres pertinents (c'est-à-dire les propriétés et les relations). L'aspect innovant est que ces paramètres doivent seulement correspondre à la définition des UCT, et pas forcément aux exigences strictes de leur agencement dans un schéma taxonomique unique. En effet, une BCT devrait fournir une « carte » des connaissances effectivement partagées au sein d'une communauté de spécialistes. Cette « carte » doit donc être fidèle à la réalité protéiforme du domaine et à sa conceptualisation, avec toutes les facettes que ses concepts présentent.

La FLO et ses applications ont été identifiées comme des modèles pour développer la représentation formelle d'une analyse à facettes de la

Terminologie. La phase actuelle du projet consiste à tester si ce modèle formel est adéquat pour une représentation terminologique qui puisse répondre aux besoins des spécialistes tout en étant un outil valide pour les applications informatiques.

Références bibliographiques

Barsalou L.W. 2008. « Grounded cognition ». *Annual Review of Psychology* 59, 617-645.

Borst W.N. 1997. *Construction of Engineering Ontologies for Knowledge Sharing and Reuse* (CTIT PhD-series No. 97-14). Enschede. The Netherlands.

Cabré, M.T. 1996. « Terminology today ». In *Terminology, LSP and Translation. Studies in language engineering in honour of Juan Carlos Sager*. H. Somers (éd.). Benjamins. Amsterdam/Philadelphia, 15-33.

— 1999. *Terminology. Theory, methods and application*. Benjamins. Amsterdam/Philadelphia.

— 2003. « Theories of terminology. Their description, prescription and explanation ». *Terminology*. 9/2, 163-199.

— 2004. « TERMINTEGRAL. Una plataforma para la construcción de bases terminológicas y ontologías ». *Linguistica Antverpiensia* 3, 245-261.

Cabré M.T., Bach C., Estopà R., Feliu J., Martínez G., Vivaldi J. 2004. « The GENOMA-KB project. Towards the integration of concepts, terms, textual corpora and entities ». In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004)*, *Lisboa, Portugal*. ELRA. Paris, 87-90.

Condamines A. 2010. « Variations in terminology. Application to the management of risks related to language use in the workplace ». *Terminology*. 16/1, 30-50.

Croft W., Cruse A.D. 2004. *Cognitive linguistics*. Cambridge University Press. Cambridge (MA).

Cruse A.D. 1995. « Polysemy and related phenomena from a cognitive linguistic viewpoint ». In *Computational Lexical Semantics*. P. Saint-Dizier, E. Viegas (éds.). Cambridge University Press. Cambridge, 33-39.

Das S. 2018. *Domain Modeling Theory and Practice*. PhD thesis. University of Trento.

Das S., Giunchiglia F. 2016. « GeoEtypes. Harmonizing Diversity in Geospatial Data (Short Paper) ». In *On the Move to Meaningful Internet*

Systems : OTM 2016 Conferences. C. Debruyne, H. Panetto, R. Meersman, T. Dillon, E. Kühn (éds.). Springer. Cham, 643-653.

Das S., S. Roy. 2016. « Faceted Ontological Model for Brain Tumour Study ». *Knowledge Organization* 43/1, 3-12.

Drouin P., Francoeur A., Humbley J., A. Picton (éds). 2017. *Multiple Perspectives on Terminological Variation*. Amsterdam/Philadelphia. Benjamins.

Faber P. (éd.). 2012. *A Cognitive Linguistics View of Terminology and Specialized Language*. De Gruyter/Mouton. Berlin/Boston.

Faber P., León Araúz P., Prieto Velasco J.A. 2009. « Semantic Relations, Dynamicity, and Terminological Knowledge Bases ». *Current Issues in Language Studies* 1, 1-23.

Faber P., López Rodríguez C.I. 2012. « Terminology and Specialized Language ». In P. Faber (éd.) 2012, 9-31.

Faber P., Ureña J.M. 2012. « Specialized Language Translation ». In P. Faber (éd.). 2012, 73-92.

Fernández Silva S. 2019. « The cognitive and communicative functions of term variation in research articles. A comparative study in Psychology and Geology ». *Applied Linguistics* 40/4, 624-645.

Fillmore C.J. 1975. « An alternative to checklist theories of meaning ». In *Proceedings of the First Annual Meeting of the Berkeley Linguistic Society*. C. Cogen, H. Thompson, G. Thurgood, K. Whistler, J. Wright (éds.). University of California, Berkeley (CA), 123-131.

— 1976. « Frame semantics and the nature of language ». *Annals of the NY Academy of Sciences* 280, 20-32.

Fillmore C.J., Atkins S.B.T. 1998. « FrameNet and lexicographic relevance ». In *Proceedings of the First International Conference on Language Resources and Evalutation (LREC98). Granada, Spain*. A. Rubio, N. Gallardo, R. Castro, A. Tejada (éds.). ELRA. Paris, 417-423.

Freixa J. 2006. « Causes of denominative variation in terminology. A typology proposal ». *Terminology* 12/1, 51-77.

Geentjens S., Temmerman R., Kerremans K., De Baer P. 2006. « Sociocognitive terminology and Termontography ». In *Proceedings of the Journées d'Etudes sur le Traitement Automatique de la Langue Arabe (JETALA 2006), Rabat, Morocco*, 138-151.

Giunchiglia F. 1993. « Contextual Reasoning ». *Epistemologia* 16, 345-364.

Giunchiglia F., Dutta B., Maltese V. 2009. « Faceted Lightweight Ontologies ». In *Conceptual Modeling. Foundations and Applications. Essays in Honor*

of John Mylopoulos. A.T. Borgida, V.K. Chaudhri, P. Giorgini, E.S. Yu (éds.). Springer Verlag. Berlin/Heidelberg, 36-51.

— 2014. « From Knowledge Organization to Knowledge Representation ». *Knowledge Organization* 4/1, 44-56.

Giunchiglia F., Dutta B. 2011. « DERA. A Faceted Knowledge Organization Framework ». Soumis à l'*International Conference on Theory and Practice of Digital Libraries 2011 (TPDL'2011)*.

Giunchiglia F., Maltese V., Dutta B. 2012. « Domains and context. First steps towards managing diversity in knowledge ». *Journal of Web Semantics* 12-13, 53-63.

Giunchiglia F., Zaihrayeu I. 2007. « Lightweight Ontologies ». Technical Report DIT-07-071.

Gnoli C. (éd.). 2008. Special issue of *Axiomathes* 18/2.

Grabar N., Hamon T., Bodenreider O. 2012. « Ontologies and terminologies. Continuum or dichotomy ? ». *Applied Ontology* 7, 375-386.

Green R., 2017. « Facet analysis and semantic frames ». *Knowledge Organization* 44/6, 397-404.

Gruber T.R. 1993. « A translation approach to portable ontology specifications ». *Knowledge Acquisition* 5, 199-220.

Guarino N. 1998. « Formal Ontology and Information Systems ». In *Proceedings of the 1st International Conference on Formal Ontologies in Information Systems FOIS'98*. N. Guarino (éd.). IOS Press. Amsterdam, 3-15.

Hjørland B. 2013. « Facet analysis. The logical approach to knowledge organization ». *Information Processing & Management* 49, 545-557.

Ibekwe-Sanjuan F. 1998. « Terminological variation, a means of identifying research topics from texts ». In *Proceeding COLING-ACL '98*. C. Boitet, P. Whitelock (éds.). Université de Montréal. Canada. I, 564-570.

ISO 704 :2000. *Terminology work - Principles and methods*. International Organization for Standardization.

Kageura K. 1997. « Multifaceted/Multidimensional Concept Systems ». In *Handbook of Terminology Management*. Vol. I. S.E. Wright, G. Budin (éds.). Benjamins. Amsterdam/Philadelphia, 119-132.

Kerremans K., Temmerman R. 2004. « Towards Multilingual, Termontological Support in Ontology Engineering ». In *Proceedings of TERMINO 2004, workshop on Terminology, Ontology & Knowledge Representation. Lyon, France, January 22-23*.

Lakoff G. 1987. *Women, fire, and dangerous things. What Categories Reveal about the Mind*. The University of Chicago Press. Chicago/London.

Langacker R. 1987. *Foundations of Cognitive Grammar. Theoretical Prerequisites*. Vol. I. Stanford University Press. Stanford.

León Araúz P., Faber P., Montero Martínez S. 2012. «Specialized language semantics». In P. Faber (éd.) 2012, 95-175.

Pecman M. 2014. «Variation as a cognitive device». *Terminology* 20/1, 1-24.

Ranganathan S.R., 1969. «Colon Classification Edition 7 (1971). A preview». *Library Science with a slant to Documentation*. 6/3, 193-242.

Roche C. 2007. «Le terme et le concept. Fondements d'une ontoterminologie». In *Actes de la première conférence TOTHh 2007. Terminologie & Ontologie. Théories et Applications*. C. Roche (éd.). Institut Porphyre. Annecy, 1-22.

— 2012. «Ontoterminology. How to unify terminology and ontology into a single paradigm». In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*. N. Calzolari, K. Choukri, T. Declerck, M.U. Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, J., S. Piperidis (éds.). ELRA. Paris, 2626-2630.

Roche C., Calberg-Challot M., Damas L., Rouard Ph. 2009. «Ontoterminology. A new paradigm for terminology». In *KEOD 2009 - Proceedings of the International Conference on Knowledge Engineering and Ontology Development, Funchal - Madeira, Portugal, October 6-8, 2009*. J.L.G. Dietz (éd.). INSTICC Press. Setúbal, 321-326.

Roche C., Costa R., Carvalho S., Almeida B. 2019. «Knowledge-based terminological e-dictionaries. The EndoTerm and al-Andalus Pottery projects». *Terminology*. 25/2, 139-145.

Sager J.C. 1990. *A practical course in terminology processing*. Benjamins. Amsterdam/Philadelphia.

San Martín A., León Araúz P. 2013. «Flexible Terminological Definitions and Conceptual Frames». In *International Workshop on Definitions in Ontologies (DO 2013), Montreal - Concordia University. CEUR Workshop Proceedings*. Vol. 1061. S. Seppälä, A. Ruttenberg (éds.).

Szostak R., Gnoli C., López-Huertas M.J. (éds.) 2016. *Interdisciplinary Knowledge Organization*. Springer. Cham.

Taylor J. 1989. *Linguistic Categorization*. Clarendon Press. Oxford.

Temmerman R. 2000. *Towards new ways of terminology description. The sociocognitive approach*. Benjamins. Amsterdam/Philadelphia.

Temmerman R., Kerremans K. 2003. «Termontography. Ontology building and the Sociocognitive Approach to terminology description». In *Proceedings of CIL17*. Matfyzpress. Prague, Czech Republic.

Temmerman R., Van Campenhoudt M. (éds.) 2011. *The dynamics of terms in specialized communication. An interdisciplinary perspective*. Special issue of *Terminology* 17/1.

Tennis J.T. 2012. « Comparative modeling of Vickery's Faceted Classification and the oeuvre of S.R. Ranganathan ». In *Facets of Knowledge Organization. Proceedings of the ISKO UK Second Biennial Conference, 4th-5th July 2011, London*. A. Gilchrist, J. Vernau (éds.). Emerald Group Publishing. United Kingdom, 301-313.

Vickery B.C. 1966. *Faceted classification schemes*. Rutgers University Press. New Jersey.

Wüster E. 1974. « Die allgemeine Terminologielehre. Ein Grenzgebiet zwischen Sprachwissenschaft, Logik, Ontologie, Informatik und den Sachwissenschaften ». *Linguistics* 119/1, 61-106.

Abstract

Contemporary Terminology studies and applications are characterized by the elaboration of formalized knowledge representation frameworks which aim at reconciling the contrasting requirements of human users and computer applications, as these are often the double target of Terminological Knowledge Bases. In leading terminology representation models the ontology framework is integrated with different types of schemes, mainly inspired by theories of knowledge functioning. A model that is complementary to those already in existence is introduced here as a proposal to effectively represent the dynamicity that characterizes terminology. It is intended as an implementation of a system where the theoretical stance and its applicative realization are founded on the same principles. It is based on a faceted approach to terminological analysis and representation. Facets are considered a viable interpretation of the complex and multidimensional nature of Terminological Knowledge Units and an effective reading of the different but non-conflicting perspectives on concepts introduced and shared by specialists within a knowledge domain.

Are definitions syntactic structures or lexical frames? A formal analysis based on Spanish evidence

César Aguilar*

*Instituto de Investigaciones en Educación
Universidad Veracruzana
Xalapa, Veracruz, México
cesar.aguilar72@gmail.com
http://cesaraguilar.weebly.com

Summary. Since the foundational study of Riegel (1987) about the existence of called *énonces définitoires*, the analysis of the linguistic configuration of definitions is a necessary phase to consider in any task of terminological extraction. Therefore, many authors have proposed different methodologies these configurations, considering mainly hybrid methods that combine lexical-syntactic patterns and statistical techniques (Auger, 1997; Pearson, 1998; Rebeyrolle, 2000; Meyer, 2001; Klavans and Muresan, 2001; among others). Nevertheless, these authors have not yet tackled the question: definitions are configured only from syntactic structures, or to make the attempt to recognize lexical frames associated to particular verbs, following the typical behavior of any argumental structure? In order to answer this question, I expose a contrastive study supported for the empirical evidence generated by two Spanish technical corpora of DCs: one focused in Engineering area, and other came to Informatics area. The results showed here can be useful for explaining how the syntactic/semantic interface works in formulating a definition.

1. Introduction

The (semi-)automatic recognition of definitions in corpora has become an important task for computational lexicography and terminology. Although there is a degree of complexity due to the role that definitions play as linguistic representation of concepts (Sager, 1990), currently many authors explore

different methods and techniques for identifying definitions, considering both linguistic and statistical criteria. Particularly, the foundational study of Riegel (1987) about the existence of called *énonces définitoires* has opened the way for analyzing how a definition is configured from a grammatical point of view. Many authors have proposed different methodologies based on Riegel's ideas, considering both linguistic and statistical points of views in order to develop hybrid methods. Some relevant methodologies are:

- The automatic detection of *énonces définitoires*: Auger (1997), Rebeyrolle (2000).
- The recognition of linguistic patters associated to terms in discursive contexts: Pearson (1998).
- The identification of definitions in English medical texts: Klavans and Muresan (2001).
- The identification of knowledge-rich contexts: Meyer (2001).
- The formulation of a sens-text framework for supporting terminological extraction: L'Homme (2001).
- The extraction of definitions for question answering systems for English: Saggion (2004).
- The automatic mining of definitions: Malaisé, Zweigenbaum and Bachimont (2005).

Attending the objectives and goals of their methodologies, these authors share at least three particular ideas:

- In a natural language, concepts are represented by sentences. Thus, a definition is a linguistic structure useful for expressing this conceptual information (Sager and Ndi-Kimbi, 1995).
- If definitions are linguistic representations of concepts, then it is possible to recognize regular patterns in syntactic and lexical-semantic levels (Riegel, 1987).
- There is a use of statistical methods and computational tools for searching and extracting these regular patterns in corpora. Therefore, the results are evaluated in order to determine if such patterns represent good or bad candidates to definitions (Malaisé, Zweigenbaum & Bachimont, 2005).

In the case of terminology, given its interest for the exploitation of text corpora for obtaining automatically terms and other relevant specialized units, has

reached important advances in the detection of conceptual information associated to terms, especially the work done by Pearson (1998) and Meyer (2001). The results obtained for these authors have allowed to formulate a set of basic ideas underlying any task oriented to search conceptual information encoded in terms and definitions:

- **Implementation and use of domain-specific corpora:** in words of Riloff and Shepherd (1999), for obtaining relevant lexical knowledge it is necessary to limit the dimensions of our corpora. Thus, we must focus on text collections pertaining to a specific subject matter, in order to extract specialized terminology for in-depth understanding this subject matter. Given the vast scope of human knowledge and the practical limitations of manual knowledge engineering, it is unrealistic to expect a complete lexical knowledge base any time soon.
- **Evaluation of search systems for terminological information extraction:** perhaps the work of Cabré, Estopà and Vivaldi (2001) is the most ambitious review around to the performance of many search systems for discovering information referred to terms. These authors provide a comparison between all the systems that they considered for their evaluation, influencing on the methodological framework of term extraction (L'Homme, 2004).
- **Use of contextual information for detecting terms:** it is a research perspective developed mainly by Auger (1997), Pearson (1998), Meyer (2001) and Rebeyrolle (2000). This perspective weights the value of linguistic and paralinguistic patterns that point to possible terms, e.g. verbal phrases (*X is Y*, *X is defined as Y*, *X refers to Y*), discursive structures (*from a... point of view, in a... way, according to*...), typographical markers (bold, italic or underlined fonts), etc.

2. Extraction of Definitional Contexts

In line with these methodologies and ideas, Sierra *et al*. (2008) developed a framework for the extraction of definitional contexts (or DCs) in Spanish, considering three minimal constitutive elements: a term, a definition, and a verbal phrase whose function is to connect the term with its definition. In addition, other elements operate as constitutive units of DCs, particularly typographical markers and discursive patterns. An example is demonstrated below:

(1) [In general $_{\text{Discursive Pattern}}$], the [**paraprofessional workers** $_{\text{Term + Typographical Pattern}}$] [are defined as $_{\text{Predicative Phrase}}$] [those persons who are engaged in the provision of social care or social services, but who do not have professional training or qualifications $_{\text{Definition}}$].

According to this example, the term *paraprofessional workers* is emphasized by the use of bold font, and it appears linked with the verbal phrase *are defined as*, and the definition *those persons who are engaged…* The term, the verbal predication and the definition are the basic units that configure the syntactic sequence of a DC. Finally, this sequence is introduced by the pragmatic pattern *in general*, which is a good clue for recognizing the start of this DC. We can represent the relation between all these units in the following way:

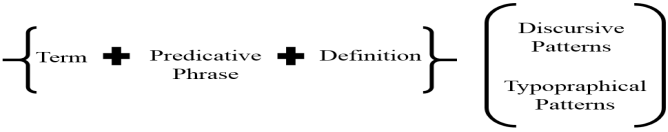


Fig. 1 – *Constitutive units of a DC structure.*

In the Figure 1, we show the canonical structure of a DC constituted by its main units. We consider that terms, predicative phrases and definitions configure the core of a DC, because these three units show a recurrent use in specialized documents. On the other hand, pragmatic patterns and typographical markers could be seen as optional units whose function is to introduce and/or indicate a potential DC in a text.

Additionally, based in the example of (1), we can distinguish a typical kind of definition according to the Aristotelian model, where the **Genus Term** is represented for the phrase *those persons*, and the **Differentia** is expressed for the sentence *who are engaged in the provision of social…*The Aristotelian model is the main pattern considered by Vossen and Copestake (1993) as well as Wilks, Slator and Guthrie (1995) for searching definitions in large-corpora. Likewise, Sager and Ndi-Kimbi (1995) observe that Aristotelian (or analytical) definitions can derivate in other formats, especially in scientific and technical corpora. Having in mind this observation, Sierra *et al*. (2008), propose a typology supported for the presence/absence of the Genus Term or Difference in the definition. Such typology is represented in the following scheme:

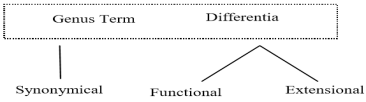


Fig. 2 – *Types of definitions derived from Aristotelian model according to Sierra* et al. *(2008 :81).*

Each type is described as follows :

- **Analytical definition :** This definition occurs when both constituents, Genus and Differentia, are present.
- **Synonymical definition :** This definition only provides a Genus term hat is conceptually equivalent to the defined term.
- **Functional definition :** Here the Genus no longer is present but the Differentia indicates the function of the entity.
- **Meronymical or extensional definition :** This definition has no Genus but includes a differentia that enumerates the parts that make up the entity.

These types of definitions, according to Sierra *et al*. (2008), as well as Aguilar and Sierra (2009), these definitions are linked to specific verbs whose function is to operate as head of predicative phrases, as we can see in the following examples :

(2) [El apartarrayos $_{\text{Term}}$] [es $_{\text{Verbal Predication}}$] [un dispositivo $_{\text{Genus Term}}$] [que protege las instalaciones contra sobretensiones de origen atmosférico $_{\text{Differentia}}$] (Eng. [The lightning conductor $_{\text{Term}}$] [**is** $_{\text{Verb}}$] [a device $_{\text{Genus Term}}$] [that protects electrical systems against surges of atmospheric origin $_{\text{Differentia}}$])

(3) [La tensión de base $_{\text{Term}}$] [se le llama también $_{\text{Verbal Predication}}$] [tensión unidad $_{\text{Genus Term}}$]. (Engl. [The base tension $_{\text{Term}}$] [**it is also called** $_{\text{Verb}}$] [unit tension $_{\text{Genus Term}}$])

(4) [La técnica de velocimetría de imágenes $_{\text{Term}}$] [permite $_{\text{Verbal Predication}}$] [medir la velocidad de un campo de flujo bi o tri dimensional $_{\text{Differentia}}$] (Engl. [The method of image velocimetry $_{\text{Term}}$] [**allows** $_{\text{Verb}}$] [to measure the speed of a flow-field in two or three dimensions $_{\text{Differentia}}$])

(5) [La zona límite $_{\text{Term}}$] [incluye $_{\text{Verbal Predication}}$] [planicies costeras, marismas, áreas de inundación, playas, dunas y corales $_{\text{Differentia}}$] (Eng. [The border zone $_{\text{Term}}$] [**includes** $_{\text{Verb}}$] [coastal plains, salt marshes, flood areas, beaches, dunes and corals $_{\text{Differentia}}$])

In this examples, we observe that, depending of the verb, is introduced a certain type of definition, according to our typology. Such relation among verbs and definitions can be explained from two points of view:

(i) A syntactic perspective, which considers the role that plays the predicative phrases as the grammatical structure that supports the discursive configuration of a DC.

(ii) A lexical-semantic perspective, where the verb maps an argument structure that configures a particular event. An example of this is the semantic description formulated for **FrameNet** (Baker, Fillmore and Lowe, 1998), in which verbs as *to consider*, *to define*, *to describe*, *to understand* and other similar configure an argument structure with — at least— four participant: (i) a *Cognizer* who performs an act of categorization, (ii) an *Item* which is the entity construed or treated by the *Cognizer* as being an instance of a particular category, (iii) a *Category* understood as a general type or class of which the Item is considered an instance, and (iv) a *Criteria*, that is, a set of properties of the Item which is used as a basis for categorization.

Therefore, in this work we expose on our work a contrastive study around these two perspectives, attending the empirical evidence generated by two Spanish technical corpora of DCs: one focused in Engineering area (Medina *et al*., 2004) and other came to Informatics area (L'Homme and Drouin, 2006). The results obtained by this contrast can give more arguments about possibility to understand how works this relation among verbs and definitions expressed in specialized domain texts.

3. Predicative phrases and definitions

Having in mind the notion of *énoncé définitoire* formulated by Riegel (1990), and applied by Auger (1997), Rebeyrolle (2000), as well as Malaisé, Zweigenbaum & Bachimont (2005) for extracting definitions from specialized corpora, Sierra *et al.* (2008), and mainly Aguilar and Sierra (2009), have

proposed that the core of a DC is the relation between terms, verbs and definitions. This relation can be represented in a **predicative phrase** (hereinafter, PrP), which is a kind of phrase structure organized around the relation *X-is-a-Subject-of/Y-is-a-predicate-of*, according to Rothstein (1983), and this relation is regulated by a syntactic rule named *rule of predicate linking*. Examples of these relations are :

(6) John *is* an intelligent professor.

(7) John *considers* his father *as* an intelligent professor.

According to Rothstein, a PrP is formed from a phrase structure where the verb is associated to its complements (e.g. : argument and/or adjuncts). This structure configures a predicative phrase, which is linked to a subject through a relation of saturation. In other words, if a predicate builds a complete structure (that is, a PrP) it needs to be linked to a specific subject.

Following Rothstein's explanation, Bowers (1993, 2001) develops a simple model to describe the syntactic configuration of these phrases, called Predicative Phrase (PrP). The PrP is mapped by a non-lexical head (that is, a functional head), and its grammatical behaviour is similar to that of phrases such as Inflexional Phrase (IP) or Complement Phrase (CP). A graphical tree representation of a PrP is :

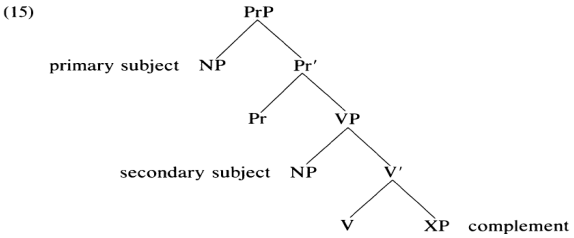


Fig. 3 – *Tree representation for PrP, according to Bowers (1993 : 596).*

Figure 3 shows the basic elements that make up a PrP. Bowers recognise a functional head with the features $^{+/-}$ predicative (Pr). This head maps two subjects, a primary subject in the position of specifier of PrP (represented by a noun phrase or NP); and a secondary subject, in the position of specifier of verbal phrase or VP (often a NP). Thus, both subjects, the VP and the PrP are linked to one or several complements, which assume many phrasal representations (e.g. NP, IP, CP, and so on).

Finally, based on his explanation, Bowers conceives two kinds of PrPs, which are:

a) If the predicate is linked to the subject mapped by the PrP, then it is a primary predication.

b) In contrast, if the predicate is associated with the subject mapped by the VP, then it is a secondary predication.

3.1. Primary and secondary predications

Based on this distinction made by Rothstein and Bowers between primary and secondary subjects, we recognize two types of predications:

- Simple or primary predication, consisting of a subject to the left of the verb, in position of Specifier of PrP, and a predicate to the right of the verb.
- Double or secondary predication, which integrates a primary subject in a pre-verbal position, a secondary subject (situated as specifier of VP), and the predicate.

We illustrate this distinction in the following examples:

(8) [A computer [is [a kind of electronic machine used to make mathematical and logical operations $_{PrP}$] $_{VP}$] $_{IP}$]).

(9) [Turing [defines a computer [as a kind of electronic device that processes a set of data $_{PrP}$] $_{VP}$] $_{IP}$]).

In (8), we observe a canonical case of primary predication, where the verb *is* inserts the predicate *a kind of electronic machine…,* linking it to the NP *a computer.* For the term extraction, the noun *computer* is equivalent to a term, while the predicate formulates the definition related to the term.

With regarding to (9), the verb *defines* maps a PrP composed by three elements: a subject represented by the NP *Turing*, an object expressed by the NP *a computer*, and the predicate *as a kind of electronic device…* This predicate introduces a set of attributes that characterizes *a computer*, following the rule of association *Y is a predicate of X*. Both the predicate and the object configure a predicate linked to the subject through the verb *defines*, generating two symmetrical predicate structures: the first predication covers the subject *Turing*, the main verb, the object and the PrP; in contrast, the second predication covers the object *a computer* with the PrP whose head is the preposition *as*, according with the explanation offered by Bowers (1993, 2001). For our

analysis, we consider these two types of predications as regular patterns that syntactically codify sequences of terms and definitions linked through PrPs.

3.2. Combinatory patterns

Taking into account the formal description of PrP proposed by Aguilar and Sierra (2009), it is possible to identify two types of patterns that structure particular sequences in DCs:

- In the case of primary predications, it codifies a sequence composed of a term, a PrP and a definition.
- In the case of secondary predication, it codifies a sequence composed of a specific author, a term, a PrP and a definition.

3.2.1. Term + PrP + Definition

This sequence is a good example of a formulation of canonical definitional patterns, because the primary predication links directly a subject represented by a term, with a specific set of attributes codified in the PrP. An example is:

(10) [El contenedor refrigerado Term] [es PrP] [una forma especializada de transporte de perecederos Definition] (Eng. [The refrigerated container Term] [is PrP] [a specialized form to transport perishable goods Definition])

Such sequence is equivalent to the structure of primary predication, then the term is situated in the position of primary subject, the verb *ser* (Eng.: *to be*) has the role of head of a PrP, and the definition is introduced through the PrP. We recognize this pattern associated not only analytic definitions, but also synonymical, functional and extensional definitions, according to the examples exposed in (2-5).

3.2.2. Author + Term + PrP + Definition

The second sequence we report here shows the sequence Author + Term + PrP + Definition. The characteristic feature of this pattern is that it explicitly points out the author (or authors) of the definition. This feature maps a semantic role, according to **FrameNet**[1] (Baker, Fillmore and Lowe, 1998), concretely the author can be conceived as a *Cognizer* that associates certain

1 For more details, see the following WEB Site: https://framenet.icsi.berkeley.edu/fndrupal/.

Categories (the definition) to a particular *Item* (or the term). This is illustrated in the following examples:

(11) [Carlos Godino $_{\text{Author}}$] [define $_{\text{PrP}}$] [la arquitectura naval $_{\text{Term}}$] [como la ciencia que se enfoca en la construcción de los buques $_{\text{Definition}}$] (Eng. [Carlos Godino $_{\text{Author}}$] [defines $_{\text{PrP}}$] [naval architecture $_{\text{Term}}$] [as the science that focuses on the construction of ships $_{\text{Definition}}$])

(12) [El artículo $_{\text{Author}}$] [describe $_{\text{PrP}}$] [la evolución de ecología del paisaje $_{\text{Term}}$] [como una ciencia integrativa y transdisciplinaria $_{\text{Definition}}$] (Eng. [The article $_{\text{Author}}$] [describes $_{\text{PrP}}$] [the evolution of landscape ecology $_{\text{Term}}$] [as an integrative and interdisciplinary science $_{\text{Definition}}$]).

(13) [Ø Podemos $_{\text{Author}}$] [considerar $_{\text{PrP}}$] [las computadoras programables modernas $_{\text{Term}}$] [como la evolución de sistemas antiguos de cálculo o de ordenación $_{\text{Definition}}$] (Eng. [We $_{\text{Author}}$] [can consider $_{\text{PrP}}$] [the modern programmable computers $_{\text{Term}}$] [as the evolution of ancient systems of calculation and management $_{\text{Definition}}$]).

Hence, the pattern followed by this sequence clearly refers to the author of a definition, as shown in (11-13). However, a syntactic behaviour observed in this pattern is its recurrent configuration in non-personal forms, i.e. impersonal and passive forms, for example:

(14) [Se conoce como $_{\text{Verbal Predication}}$] [reenganche rápido $_{\text{Term}}$] [a la operación de cierre de un interruptor después de una falla Definition] (Eng. [It is known as $_{\text{Verbal Predication}}$] [Quick re-closing $_{\text{Term}}$] [to the operation of a switch after a fault $_{\text{Definition}}$]).

(15) [Los niveles relativos de los alcances de ola $_{\text{Term}}$] [fueron descritos como $_{\text{Verbal Predication}}$] [una función del parámetro de similitud de oleaje $_{\text{Definition}}$] (Eng. [The relative levels of the wave reach $_{\text{Term}}$] [were described as $_{\text{Verbal Predication}}$] [a function of the wave similarity parameter $_{\text{Definition}}$]).

In these examples, we observe the use of non-personal verbal patterns as in (14), where the clitic *Se* is inserted (Eng. *It*) to make the sentence impersonal, or in (15), which is in the passive form. Thus, when these sequences assume a non-personal pattern, they become equivalent to primary predications, where there is not an explicit mention of the author of a definition. We summarise all these patterns in Table 1, considering some recurrent verbs in the position of head of PrP. These verbs are not exclusive, but their recurrence has been reported by Sierra *et al*. (2008):

Definition	Verbs	Associated Particles
Analytical (Primary Predication)	*referir* (to refer to) *representar* (to represent) *ser* (to be) *significar* (to signify/to mean)	*a* = to (preposition) (in the case of *referir*, it is a phrasal verb that inserts obligatory the preposition *a*)
Analytical (Secondary Predication)	*caracterizar* (to characterise) *comprender* (to include) *concebir* (to conceive) *conocer* (to know) *considerar* (to consider) *definir* (to define) *describir* (to describe) *entender* (to understand) *identificar* (to identify) *visualizar* (to visualise)	*como* = as/like (adverb) *por* = for/by (preposition)
Synonymy (Primary Predication)	*equivaler* (to be equivalent to) *llamar* (to call) *nombrar* (to name) *ser _ igual* (to be equal to) *ser _ similar* (to be similar to)	*también* = also (adverb) a = to (preposition) *igual a* = equal to (adverb phrase) *similar a* = similar to (adverb phrase)
Functional (Primary Predication)	*emplearse* (to employ + clicit *se*) *encargar* (to be in charge of) *funcionar* (to function) *ocupar* (to occupy) *permitir* (to allow) *servir* (to serve) *usar* (to use) *utilizar* (to utilise/to use)	*de* = of (preposition) *para* = for (preposition)

Definition	Verbs	Associated Particles
Extensional (Primary Predication)	*componer* (to be composed of) *comprender* (to include) *consistir* (to consist of) *constar* (to consist of) *contar* (to have) *constituir* (to constitute) *contener* (to contain) incluir (to include) integrar (to integrate)	*de* = of (preposition) *por* = for/by (preposition) *con* = with (preposition)

Tab. 1 –*Types of definitions linked to PrPs and verbs.*

4. Lexical-semantic description

Once made a formal description of the syntactic relation between terms, PrPs and definitions, we show in the following sections the results of a brief lexical-semantic analysis, having in mind the argumental structure contained in the set of verbs enlisted in the Table 1. For performing such analysis, we consult three lexical-semantic databases: FrameNet (Baker, Fillmore and Lowe, 1998), **Spanish FrameNet**[2] (Subirats, 2009), and **ADESSE**[3] (García-Miguel and Albertuz, 2005).

4.1. Categorization verbs and definitions

FrameNet grouped those verbs that, in the Table 1, are linked to secondary predications, in a particular lexical-semantic unit named *categorization verbs*. Such group is characterized according the following argumental features:

- A *Cognizer* who performs an act of categorization.
- An *Item* which is the entity construed or treated by the *Cognizer* as being an instance of a particular category.
- A *Category* understood as a general type or class of which the Item is considered an instance.

2 For more details, see the following WEB Site: http://spanishfn.org/.
3 For more details, see the following WEB Site: http://adesse.uvigo.es/.

- A *Criteria*, which expressed the properties of the Item which is used as a basis for categorization. Criteria are the general dimensions along which Items can potentially differ from one another and hence, fall into different *Categories*.

FrameNet points as a syntactic feature propter of this kind of verbs is the insertion of a prepositional phrase headed by *as* for expressing the category linked to an item, e. g. :

(16) [John $_{\text{Cognizer}}$] considers [his father $_{\text{Item}}$] [*as* an intelligent professor $_{\text{Category}}$].

(17) [Computational Linguistics $_{\text{Item}}$] is defined [*as* the study of computer systems for understanding, generating and processing natural language $_{\text{Category}}$].

The verbs considered by FrameNet in this group —in English— are : *to classify, to categorize, to conceive, to consider, to construe, to count, to deem, to define, to identify, to interpret, to peg, to perceive, to pigeonhole, to read, to regard, to render, to see, to stereotype, to translate, to understand* and *to view.*

4.2. Some lexical-semantic particularities in Spanish

Despite the description offered by FrameNet, it is necessary to consider specific differences respect to the semantic behavior of these verbs in Spanish. An example of such differences is the verb *comprender* (Eng. : *to comprehend*). FrameNet considers this verb in English as a member of the group named *Awareness Verbs*, characterized for the following argumental features :

- A *Cognizer* who performs an act of categorization.
- A *Content*, which is the object of the Cognizer's awareness.
- A *Topic*, it is the element associated to the Content.

An example of Awareness verb is :

(18) [Turing $_{\text{Cognizer}}$] [comprehends [the concept of computer $_{\text{Topic}}$] [as an electronic device that processes a set of data $_{\text{Content}}$]

However, in Spanish the verb *comprender* is also linked to extensional definitions, and this meaning is very common in specialized corpora. We can

corroborate such meaning reviewing the following set of DCs taken from Bwananet[4]:

(19) [La escritura de compiladores $_{\text{Term}}$] **comprende** [los lenguajes de programación, la arquitectura de computadores, la teoría de lenguajes, los algoritmos y la ingeniería de software $_{\text{Extensional Definition}}$] (Eng.: [*Writing compilers* $_{\text{Term}}$] ***includes*** [*programming languages, computer architecture, language theory, algorithms and software engineering* $_{\text{Extensional Definition}}$]).

(20) [La fase de desarrollo $_{\text{Term}}$] **comprende** [tres pasos distintos: diseño, generación del código (manual o automáticamente) y prueba $_{\text{Extensional Defimition}}$] (Eng.: [*Development phase* $_{\text{Term}}$] ***involves*** [*three distinct steps: design, code generation (in a manual or automatic way) and testing* $_{\text{Extensonal Definition}}$]).

(21) [La sintaxis de un lenguaje $_{\text{Term}}$] **comprende** [las reglas que determinan la estructura de todas las sentencias permitidas de ese lenguaje $_{\text{Extensonal Definition}}$]. (Eng.: [*Language syntax* $_{\text{Term}}$] ***comprises*** [*the rules that determine the structure of all permitted sentences of that language* $_{\text{Extensonal Definition}}$]).

Other verb with a similar situation is *contar* (Eng.: *to count*): in FrameNet this verb is included in the frame of categorization, but in Spanish *contar* shows another different meaning nearest to a meronymy relation (and indeed, an extensional definition). We can see it in the following examples collected again from Bwananet:

(22) [Los autómatas de pila $_{\text{Term}}$] **cuentan con** [una pila en donde pueden almacenar información para recuperar la más tarde $_{\text{Extensonal Definition}}$] (Eng.: [*Pushdown automaton* $_{\text{Term}}$] ***have*** [*a stack where they can store information to retrieve later* $_{\text{Extensional Definition}}$]).

(23) [BITNET $_{\text{Term}}$] **cuenta con** [funciones de correo electrónico y listas de correo $_{\text{Extensional Definition}}$]. (Eng.: [*BITNET* $_{\text{Term}}$] ***includes*** [*functions of email and mailing lists* $_{\text{Extensional Definition}}$].

(24) [Una cadena de caracteres $_{\text{Term}}$] **puede contar con** [más de un árbol de análisis sintáctico, así como más de un significado $_{\text{Extensional Definition}}$].

4 For more details, see the following WEB Site: http://bwananet.iula.upf.edu/.

(Eng.: *A* [*string of characters* $_{\text{Term}}$] ***can have*** [*more than one parse tree and more than one meaning* $_{\text{Extensional Definition}}$]).

4.3. Some lexical-semantic particularities in Spanish

In the case of the Spanish lexical-semantic database as Spanish FrameNet (Subirats, 2009), sometimes we observe a mismatch between the theoretical description of the argumental structure of a verb, and the empirical evidence obtained from specialized corpora. This is the case of *encontrar* (Eng.: *to find*), considered as a verb of categorization, according to the following description:

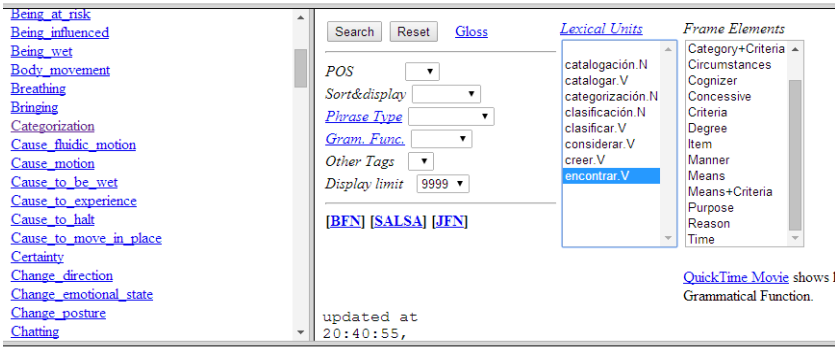


FIG. 4 – *Lexical-semantic description of* encontrar *according to Spanish FrameNet.*

Nevertheless, contrasting with the use of this verb in technical texts, it is not clear how such meaning can be derived from the categorization frame. In these examples, *encontrar* is the nearest to meaning of location:

(25) El archivo NWUSER.EXE **se encuentra en** el directorio SYSTEM de MS Windows para la instalación del software del cliente NetWare. (Eng.: *The NTUSER.EXE file* ***is located*** *in the MS Windows SYSTEM directory for installing l l NetWare client software*).

(26) La lista de grupos **se encuentra en** el centro proveedor de servicios. (Eng.: *The list of groups* ***is*** *in the provider service center*).

(27) La UC capta la instrucción que se encuentra en la posición que indica el CP (dirección 14). (Eng.: *UC captures the instruction is in the position indicated by the CP (direction 14)*).

Finally, regarding again to the description of FrameNet, there are particular cases where the contrast between English and Spanish is notorious: the verb *to create*, according to FrameNet is a member of categorization group, but comparing with the description offered for the semantic database ADESSE (García-Miguel and Albertuz, 2005), such verb shows a semantic behavior propter of verb of creation, that is, a variant of active verbs focused in the generation of concrete entities (e. g.: *to build a house*), or abstract concept (e. g.: *to create a philosophy*).

In a similar way, while FrameNet relates the argumental structure of the verb *to see* to a possible categorization verb, for Spanish FrameNet the Spanish version (*ver*) is conceived as a perception verb, equivalent to others as *escuchar* (Eng.: *to hear*), *mirar* (Eng.: *to look*), *tocar* (*to touch*), and so on.

5. Some final remarks

For concluding our work, it is worth noting that it was aimed from the beginning to answer the following question: in order to support the grammatical framework necessary for developing (semi-)automatic systems able to extract definitions, which is the better way: to consider only syntactic structures, or to make the attempt to recognize lexical frames associated to particular verbs detected in DCs?

Based in our results, we consider that:

- FrameNet is a good resource for finding the meaning of prototypical frame semantics, but it is not enough for explaining derived meaning recognizable in specialized knowledge domains.
- The predicative phrases are a useful resource for detecting candidates of definitions expressed in syntactic structures, but it is not enough for determining if such candidates are good definitions or not.
- A possible solution is to use these prototypical meanings for creating specific lexical-semantic lattices, taking into account alternative semantic models as, for example, the generative lexicon (Pustejovsky, 1991, 1995).

- It is pertinent to include this semantic information in the automatic extraction of definitions, because this information is useful to distinguish what is the semantic configuration of a good definition.

Acknowledgments

This work has been supported by the National Commission for Scientific and Technological Research (CONICYT) of Chile, Project Number : 11130565.

References

Aguilar, C., and Sierra, G. (2009): "Reconocimiento de definiciones asociadas a frases predicativas en contextos definitorios", *Revista de Procesamiento de Lenguaje Natural*, No. 43 : 151-158.

Auger, A. (1997): Repérage des énonces d'intérêt définitoire dans les bases de données textuelles, Thèse de Doctorat, Neuchâtel, Suisse, Université de Neuchâtel.

Baker, C., Fillmore, Ch., and Cronin, B. (2003): "The Structure of the Framenet Database", *International Journal of Lexicography*, 16(3): 281-296.

Bowers, John (1993): "The Syntax of Predication", *Linguistic Inquiry*, 24(4): 591-636.

Bowers, John (2001): "Predication". In Baltin, M. y Collins, C. (eds.), *The Handbook of Contemporary Syntactic Theory*, Oxford, Blackwell : 299-333.

Cabré, T., Estopá, R., and Vivaldi, J. (2001): "Automatic term detection : A review of current systems". In Bourigault, D., Jacquemin, C., and L'Homme, M.-C. (eds.), *Recent Advances in Computational Terminology*, Amsterdam/Philadelphia, John Benjamis : 53-87.

García-Miguel, J., and Albertuz, F. (2005): "Verbs, Semantic Classes and Semantic Roles in the ADESSE project". In Erk, K., Melinger, A., and Schulte im Walde, S. (eds.), *Proceedings of the Interdisciplinary Workshop on the Identification and Representation of Verb Features and Verb Classes*, Saarbrücken, Germany : february 28/may 01. Web site : http://adesse.uvigo.es.

Klavans, J., and Muresan, S. (2001): "Evaluation of the DEFINDER system for fully automatic glossary construction." In *Proceedings of the American Medical Informatics Association Symposium*, New York, ACM Press : 252-262.

L'Homme, Marie-Claude (2001): "Using Terminological Definitions to Find Conceptual Relations between Terms", Montréal, Université de Montréal, WEB site : www.ling.umontreal.ca/lhomme/.

L'Homme, M.-C., and Drouin, P. (2006): *Corpus de Informática para el español*, Groupe Éklectick, OLST-Université de Montreal : http://www.olst.umontreal.ca/.

Malaisé, V., Zweigenbaum, P., and Bachimont, B. (2005): "Mining defining contexts to help structuring differential ontologies", *Terminology* 11(1): 21-53.

Medina, A., Sierra, G., Garduño, G., Méndez, C., and Saldaña, R. (2004): "CLI : An Open Linguistic Corpus for Engineering". In De Ita, G., Fuentes, O., and Galindo, M. (eds.) *Proceedings of IX Ibero-American Workshop on Artificial Intelligence IBERAMIA 2004*, Puebla, Mexico : 203-208.

Meyer, I. (2001): "Extracting a knowledge-rich contexts for terminography : A conceptual and methodological framework". In Bourigault, D., Jaquemin, C., y L'Homme, M.-C. (eds.), *Recent Advances in Computational Terminology*, Amsterdam/Philadelphia, John Benjamins : 279-302.

Pearson, J. (1998): *Terms in Context, Philadelphia*, John Benjamins.

Rebeyrolle, Josette (2000): *Forme et fonction de la définition en discours*, Thèse de Doctorat, Toulouse, France, Université Toulouse-Le Mirail.

Riegel, M. (1987): "Définition directe et indirecte dans le langage ordinaire : les énoncés définitoires copulatifs", *Langue française*, num. 73 : 29-53.

Rothstein, S. (1983): *The Syntax Forms of Predication*, Ph. D. Thesis, Cambridge, Mass., MIT.

Sager J.C. (1990): *A Practical Course in Terminology Processing*, Amsterdam/Philadelphia : John Benjamins.

Sager, J. C., and Ndi-Kimbi, A. (1995): "The conceptual structure of terminological definitions and their linguistic realisations", *Terminology* 2(1): 87-106.

Saggion, H. (2004): "Identifying definitions in text collections for question answering". In *Proceedings of the 4th International Conference on Language Resources and Evaluation,* Lisbon, Portugal : 1927-1930.

Sierra, G., Alarcón, R., Aguilar, C., and Bach, C. (2008): "Definitional Verbal Patterns for Semantic Relation Extraction", *Terminology*, 14(1): 74–98.

Subirats, C. (2009): "Spanish Framenet : A frame-semantic analysis of the Spanish lexicon". In Boas, H. (ed.). *Multilingual FrameNets in Computational Lexicography. Methods and Applications*, Berlin/New York, Mouton de Gruyter : 135-162.

Vossen, P., and Copestake, A. (1993): “Defaults in lexical representation”. In Briscoe, T., Paiva, V., and Copestake, A. (eds.), *Inheritance, Defaults and the Lexicon*, Cambridge, Cambridge University Press : 246-274.

Wilks, Y., Slator, B., and Guthrie, L. (1996): *Electric Words*, Cambridge, Mass., MIT Press.

Articles courts

Modelage d'un domaine économique conçu pour servir de feuille de route à la traduction spécialisée, offrant traçabilité des notions

Jan Roald

École Nationale des Hautes Études Commerciales de Norvège
Adresse postale : Helleveien 30, 5045 Bergen, Norvège
Adresse électronique : jan.roald@hh.no

1. Objectif et préalables théoriques

Ce projet pilote a pour objet le modelage d'un domaine économique, celui qui est en principe à la base de l'économie prise dans son ensemble : une praxis exercée par l'entreprise, micro-acteur économique, et la finalité de celle-ci : ventes et achats de produits, matériels ou immatériels (les affaires), prenant appui sur la definition de 'commerçant' du Code de commerce.

D'inspiration wüsterienne, il part de l'hypothèse que la notion de 'terminologie' n'a de fondement ni de sens que dans la mesure où elle se définit par rapport à un univers du discours (vision du domaine transposé au plan discursif) et qu'elle se rattache au lexique, source de création des vocabulaires et donc des terminologies, seul moyen de faire passer la communication (à côté des nombreux systèmes sémiotiques en vigueur ou à créer, mais dont l'interprétation passe forcément par la langue naturelle).

Dans un premier temps, il importe donc d'identifier et de circonscrire le domaine en question.

2. Objectif pratique et critères de délimitation et de découpage du domaine

Étant donné que l'École NHH est chargée de l'accréditation des traducteurs professionnels en Norvège, il nous a semblé bon de mettre en chantier un projet destiné à concevoir un outil de gestion numérique au service de ceux qui se préparent à cet examen, ainsi que de ceux qui exercent déjà le métier

de traducteur spécialisé, sans que les applications d'un tel outil se limitent à cette finalité.

Le modèle se base sur la vision de l'entreprise comme un ensemble (variable) de fonctions, qui entretiennent entre elles des rapports assimilables à ceux qui s'établissent entre notions formant des systèmes conceptuels en fonction de rapports soit génériques, soit méronymiques, ou encore associatifs.

Le but sera de rendre les termes traçables en ce sens que les termes présents dans les textes produits au sein de la pratique économique et faisant l'objet de traduction seront renvoyés à leur "lieu d'ancrage". Ainsi par exemple, le terme de "poste client" sera renvoyé au sous-domaine "affacturage", où la notion que ce terme véhicule est domiciliée. Ce référentiel se composerait de textes produits par la profession visée, un renvoi à la législation qui régit cette activité et un équivalent en langue étrangère (le norvégien en l'espèce).

Cette démarche impose une sélection rigoureuse des textes, fondée sur une définition contraignante de la notion de 'langue de spécialité' (Roald 2010). Sous cette optique, ce qui est en jeu, n'est pas la fréquence d'une suite d'éléments telle que 'poste client' mais la notion de 'poste client' telle qu'elle se définit et se positionne au sein du réseau notionnel qui compose la fonction d'affacturage.

Ci-dessous sont présentées de simples esquisses de fonctions, à titre d'illustration :

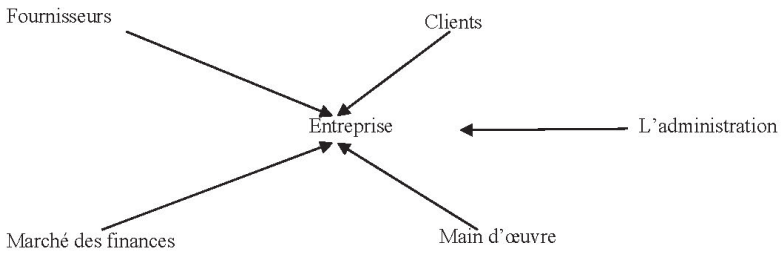


Fig. 1 – L'entreprise et ses partenaires

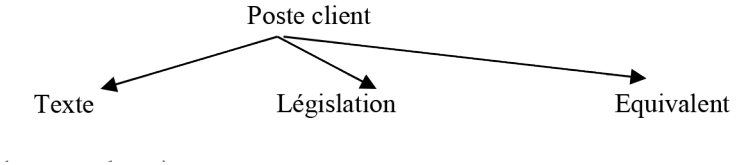


Fig. 2 – Exemple : sous-domaine

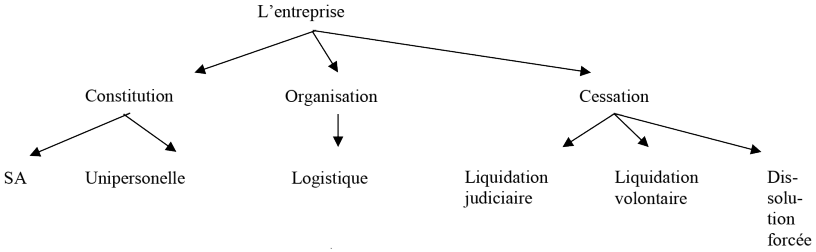


Fig. 3 – Fonctions intra-entreprise génériques

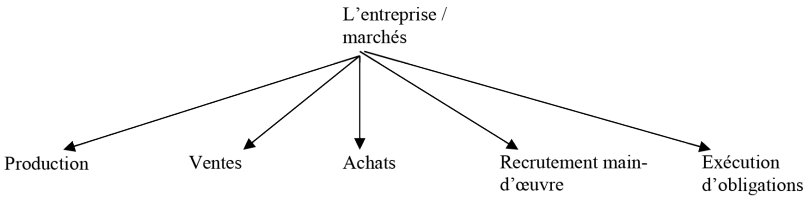


Fig. 4 – L'entreprise en rapport avec ses marchés

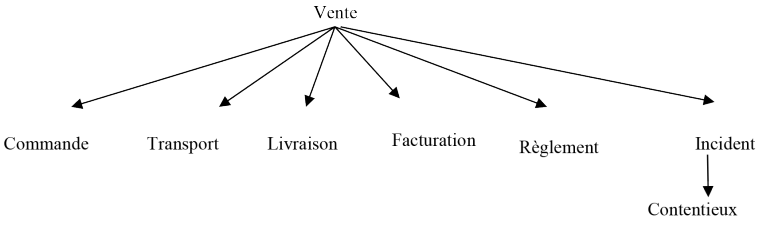


Fig. 5 – Détail d'une fonction : vente

Références

Rey, Alain (1979). *Terminologie – noms et notions*, Presses universitaires de France, Paris.

Roald, Jan (2007). *Strates discursives et textes de spécialité à traces dialogiques : Le droit et les relations du travail*. XVIth Symposium on LSP, Hamburg (Allemagne), 27-31 août 2007.

Roald, Jan (2009). *Faglig kommunikasjon*. NHH : It's Learning, pensumlitteratur, upublisert artikkel (article non publié)

Roald, Jan (2010). *Strates discursives, textes de spécialité et terminologie : Le droit et les relations du travail*. Arena Romanistica ; Issue : Professional Communication and Terminology, 7/2010

Roald, Jan & Whittaker, Sunniva (2011). *Om dannelsen av juridisk terminologi : fra kaos til lovregulering*. Nordterm, 2011

Roald, Jan & Whittaker, Sunniva (2012). *Les normes discursives dans le langage juridique : défis terminologiques et traductologiques*. La traduction juridique : Points de vue didactiques et linguistiques, Université Jean Moulin – Lyon 3. Publications du Centre d'Études Linguistiques, Nouvelle série 2012-1.

Roald, Jan & Whittaker, Sunniva (2013). *À la recherche de néologismes juridiques norvégiens et français : discussion méthodologique*. Arena Romanistica, 2013 (Jan Roald & Sunniva Whittaker)

Roald, Jan (à paraître). *En dynamisk modell for avgrensing og inndeling av et fagområde : En uomgjengelig etappe i et terminologisk arbeid og et veikart i økonomisk oversettelse*. NORDTERM, Stockholm 2013

Spang-Hanssen, Henning (1983). Kommunikation og fagsprog : I Språk i Norden. Oslo : Cappelen, 25-37

Summary

This short article presents a pilot project in the field of business terminology, aiming at creating an ontology based on the conception of a business as a series of functions. The main goal is to produce an efficient tool for professional translators.

Multilingual and Information Retrieval Systems A proposal to manage multi-word translation and ontology alignment

Johanna Monti*, Mario Monteleone**, Maria Pia di Buono**, Federica Marano**

* Dept. of Human and Social Sciences - University of Sassari – Italy
** Dept. of Political, Social and Communication Sciences - University of Salerno – Italy

1. Introduction

This paper describes a methodology for the development of an ontology-based Cross-Language Information Retrieval (CLIR) application, and shows how it is possible to achieve translations of Natural Language (NL) queries in any language by means of a knowledge-driven approach.

Our research focuses on the effective processing and translation of domain specific queries in order to overcome the current limitations of the state-of the-art applications in this field.

The growing need by users to access information on the Web in languages different from their own is fostering the research in the field of CLIR applications, which allow users to search and access information in multilingual document collections on the Web. Usually, information is searched by means of queries expressed in users' mother tongues. These queries are automatically translated in the desired foreign language and the results are translated back in users' mother tongues. This process is based on two different translation stages: query translation and document translation. The first one concerns the translation of queries expressed in users' mother tongues in the desired foreign language, whereas the second one consists in the back translation in users' mother tongues of the relevant documents found by means of the translated queries

CLIR success clearly depends on the quality of translations; therefore, faulty translations cause serious problems in retrieving the relevant information.

2. Related works

In the literature, there are several approaches to CLIR, which are either based on bilingual or multilingual Machine Readable Dictionaries (MRDs), Machine Translation (MT), parallel corpora and finally ontologies (Hull & Greffenstette, 1996; Oard & Dorr 1996; Pirkola, 1999 and more recently Oard, 2009 among others).

MRD-based and MT-based CLIR are the most popular approaches, but they present several shortcomings, particularly in domain-specific contexts, because of the inaccurate processing and translation of multi-word units (MWUs), a very frequent and productive linguistic phenomenon in languages for special purposes (LSP).

This is why the approach described in this paper is based on Lexicon-Grammar (LG), a method devised for natural language formalization, automatic textual analysis and parsing, which gives special attention to the formalization of MWUs.

3. Methodology

LG theory was conceived by the French linguist Maurice Gross during the '60s. Unlike Chomsky's transformational grammar and its various offsprings, LG assumes that linguistic formal descriptions should be based on the observation of the lexicon and the combinatory behaviours of its elements, encompassing in this way both syntax and lexicon. It has also reached important results in the domain of automatic textual analysis and parsing, with the creation of software and lingware fully oriented toward NLP, such as NooJ[1], and former software packages used in the LG framework, such as INTEX and UNITEX[2].

Thanks to the LG methodology, our research tries to overcome the shortcomings of statistical approached as in Google Translate or Bing by Microsoft concerning MWU processing in queries, where the lack of linguistic context represents a serious obstacle to disambiguation. In particular, translations of specific domains texts, as it has been widely recognized, depends on the correct interpretation and translation of the meanings of terms, which are mono-referential. Furthermore, terms are characterised by a bi-univocal rela-

1 See http://www.nooj4nlp.net/pages/nooj.html.
2 More information on the website http://www-igm.univ-mlv.fr/~unitex/.

tion with their equivalents in a foreign language, i.e. by a one-to-one coupling, which causes this relation to be exclusive and reversible.

Ontologies are used in CLIR and are considered by several scholars a promising research area to improve the effectiveness of Information Extraction (IE) techniques particularly for technical-domain queries. The aim of our research is to bring together the LG methodology, which supports our knowledge-based methodological framework, and an ontology-based approach to map both data and metadata. Besides, the feasibility of our methodology will be tested for CLIR by implementing a domain-specific early prototype system.

4. Experiment

The methodology has been set up for the Italian/English language pair in the field of Cultural Heritage (CH) but it can be easily extended to other language pairs and different domains.

It is subdivided into the following steps:

1. development of ontology-based Linguistic Resources, in order to process a linguistic analysis inside an NLP environment;
2. an iterative transformation finalized to the accomplishment of a machine-readable query;
3. the application of translation routines to obtain multilingual machine-readable queries which translate the NL queries and the associated ontology;
4. the execution of such queries against multilingual on-line repositories;
5. the translation of the retrieved results into the source language (i.e. the user's language) and their subsequent display.

It is worth underlining that the processing stages shown in steps 1, 2 and 3 are of crucial importance in reconstructing the conceptual relationships among NL query terms, ontology and machine readable language, and also in order to retrieve a meaningful value from subjects' text sequences. As for query words, the matching process to the ontology concepts is also based on domain labels that semantically tag (i.e. denote/connote) each entry of simple-word and multi-word electronic dictionaries.

Linguistic Resources and Ontology

LG presupposes that linguistic formal descriptions should be based on the examination of the lexicon and the combinatory behaviors of its elements, encompassing in this way both syntax and lexicon.

The Linguistic Resources (LRs) developed to achieve our research are simple-word and multi-word electronic dictionaries, but also Finite-state Automata together with the description of the semantic annotation and the translation routines used in query translation.

As for ontologies, the formal definition we rely upon is the one given by the International Council of Museums - Conseil International des Musees (ICOM – CIDOC) Conceptual Reference Model (CRM). CIDOC CRM is composed of a hierarchy of 90 classes (which includes subclasses and superclasses) and 148 unique properties (and sub-properties). The object-oriented semantic model and its terminology are compatible with the Resource Description Framework (RDF).

Furthermore, in our NLP environment, finite-state automata (FSA) and finite-state transducers (FSTs) are used to: (a) recognize and classify word relationships inside the query propositions entered/chosen by the user; (b) parse lexical ambiguity.

5. Electronic dictionaries

An electronic dictionary is a lexical database homogeneously structured, in which the morphologic and grammatical characteristics of lexical entries (gender, number and inflection) are formalized by means of distinctive and non-ambiguous alphanumeric tags (Vietri et al. 2004).

The following example represents an excerpt from the Italian/English compound word dictionary of Archaeological Artefacts developed in the framework of this research:

anfora di terracotta,N+NPN+FLX=C41+DOM=RA1+EN=earthenware amphora, N+AN+FLX=EC3

cerchi concentrici,N+NA+FLX=C601+DOM=RA1+EN=concentric ridges,N+AN+FLX=EC4

cottura ad alte temperature,N+NPAN+FLX =C611+DOM=RA1+EN=high fired, N+AN+FLX=EC4

fregio dorico,N+NA+FLX=C523+DOM=RA1+EN=doric frieze,N+AN+FLX=EC3

fusto a spirale,N+NPN+FLX=C7+DOM=RA1+EN=spiral stem,N+AN+FLX=EC3

fossile marino,N+NA+FLX=C501+DOM=RA1+EN=fossilised marine organism,N+AN+FLX=EC3

smalto verde rame,N+NAN+FLX=C04+DOM=RA1+EN=copper green glaze,N+AN+FLX=EC4

6. FSA/FSts

Indeed, FSA and FSTs are typically applied to locate morpho-syntactic patterns inside corpora and extract matching sequences, in order to build indexes, concordances, etc. This FST/FSA-based method, which is already available inside our NLP environment, can also be used to automatically recognize any text pattern.

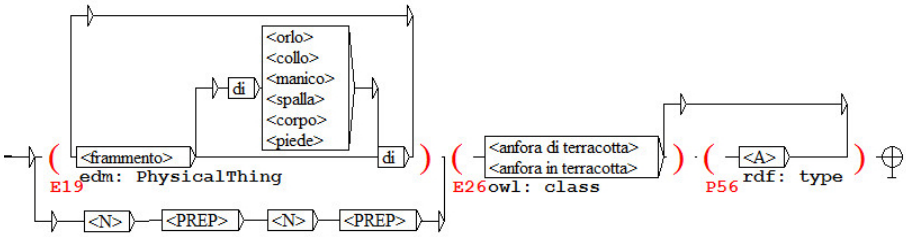


Figure 1 : Use of FSA variables for identifying classes for subject, predicate and object.

We use FSA variables for identifying ontological classes and properties for subject, object and predicate within RDF graphs, as presented in Figure 4. FSA are based on LRs, which are used during the analysis of corpora to retrieve recursive phrase structures, in which combinatorial behaviours and co-occurrence between words identify properties, also denoting a relationship. Furthermore, electronic dictionaries include all inflected verb forms allowing to process queries expressed also with passive and more generally non-declarative sentences.

In Figure 4 we also use FSA variables which apply to the sentence the following CIDOC-CRM classes and property : (i) E19 indicates “Physical Object” class ; (ii) P56 stands for “Bears Feature” property ; (iii) E26 indicates “Physical Feature” class.

Together with FSA variables we also associate POS to the Europeana

Semantic Elements (ESE) metadata format[3], currently used in Europeana, i.e. edm : PhysicalThing, owl : class, rdf : type.

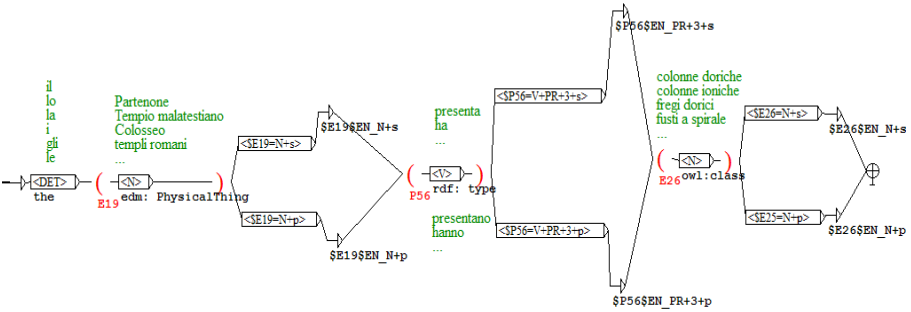


Figure 2 : Example of a translation FST.

In our model, the Translation Routines are applied independently of the mapping process of the pivot language. This allows us to preserve the semantic representation in both languages.

Indeed, identifying semantics through FSA guarantees the detection of all data and metadata expressed in any different language.

Figure 5 shows an FST in which a translation process from Italian to English is performed on the basis of a dictionary look-up, a morpho-syntactic and semantic analysis. This translation FST, in fact, identifies and annotates the different linguistic elements of declarative sentences such as “Il Partenone presenta fregi dorici”, “I templi romani hanno fusti a spirale”, etc., with their morpho-syntactic and semantic information and performs automatic translations on the basis of an LG bilingual dictionary.

7. Conclusions

In addition, with reference to this environment, an API represents the ideal solution to : (a) build an interface providing procedures callable by means of

3 http ://pro.europeana.eu/edm-documentation

external processes ; (b) drive the application for translating NL queries into Sesame RDF Query Language (SeRQL).

Bibliography

Ballesteros, L. & Croft, B. 1996. *Dictionary Methods for Cross-Lingual Information Retrieva*l. Proc. of the 7th DEXA Conference on Database and Expert Systems Applications, Zurich, Switzerland, September 1996 : 791-801.

Ballesteros, L. and Croft, W.B. 1997. *Phrasal translation and query expansion techniques for crosslanguage information retrieval*. In Proceedings of the 20th annual international ACM SIGIR conference on Research and development in information retrieval.

Ballesteros, L. & Croft, B. 1998. *Resolving Ambiguity for Cross-language Retrieval*. SIGIR'98, Melbourne, Australia, August 1998 : 64-71.

Bloomfield L. 1933. *Language*. Henry Holt, New York.

Crofts N., Doerr M., Gill T., Stead S., Stiff M. (eds.). 2008. *Definition of the CIDOC Conceptual Reference Model, Version 5.0*.

Davis, M. W., and Ogden, W. C. 1997. Free resources and advanced alignment for cross-language text retrieval. In : The Sixth Text Retrieval Conference (TREC-6). NIST, Gaithersbury, MD.

De Bueris G., Elia, A. (eds.): *Lessici elettronici e descrizioni lessicali, sintattiche, morfologiche ed ortografiche*. Plectica, Salerno (2008).

De Lacalle, Xabier Saralegi Maddalen López. 2010 *Dictionary and Monolingual Corpus-based Query Translation for Basque-English CLIR*.

Gross M. 1968. *Grammaire transformationnelle du français. – I – Syntaxe du verbe*, Larousse, Paris.

Gross M. 1975. *Méthodes en syntaxe, régime des constructions complétives*, Hermann, Paris.

Gross M. 1989. *La construction de dictionnaires électroniques*. Annales des Télécommunications, vol. 44, n° 1-2 : 4-19, CENT, Issy-les-Moulineaux/ Lannion.

Harris Z.S. 1957. *Co-occurrence and transformation in linguistic structure*. Language 33 : 293-340.

Harris Z.S. 1964. Transformations in Linguistic Structure. *Proceedings of the American Philosophical Society* 108 :5 :418-122.

Harris Z. S. 1982. *A Grammar of English on Mathematical Principles*. John Wiley and Sons, New York, USA.

Hull D. A., Gregory Grefenstette G. 1996. *Querying across languages : a dictionary-based approach to multilingual information retrieval,* Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval : 49-57.

Jianfeng Gao, Jian-Yun Nie, Endong Xun, Jian Zhang, Ming Zhou, Changning Huang. 2001. *Improving Query Translation for Cross-Language Information Retrieval using Statistical Models.* Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2001.

Knoth P., Collins T., Sklavounouy E., Zdrahal Z.. 2010. *Facilitating cross-language retrieval and machine translation by multilingual domain ontologies*

Maeda, A., Sadat, F., et al. 2000. *Query Term Disambiguation for Web Cross-Language Information Retrieval using a Search Engine*. in Proc. of the Fifth Int'l Workshop on Info. Retrieval with Asian Languages, Hong Kong, China : 173-179.

Marano F. 2012. Exploring Formal Models of Linguistic Data Structuring. Enhanced Solutions for Knowledge Management Systems Based on NLP Applications, PhD Dissertation, University of Salerno, Italy.

Monti, J. 2013. *Multi-word unit processing in Machine Translation : developing and using language resources for multi-word unit processing in Machine Translation*. PhD dissertation. University of Salerno, Italy.

Oard D. W. 2009. *Multilingual Information Access*. in Encyclopedia of Library and Information Sciences, 3rd Ed., edited by Marcia J. Bates, Editor, and Mary Niles Maack, Associate Editor, Taylor & Francis.

Pirkola, A. 1998. *The Effects of Query Structire and Dictionary Setups* in Dictionary-Based Cross-language Information Retrieval. In Croft, W.., et al., 21st Annual ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2008), Melbourne, Australia, August 24-28 :55-63.

Sadat, F., Maeda, A., et al. 2002. *A Combined Statistical Query Term Disambiguation in Cross-language Information Retrieval*. Proc. of the 13th Int'l Workshop on Database and Expert Systems Applications, Aix-en-Provence, France, September 2002 : 251-255.

Silberztein M. 1993. *Dictionnaires électroniques et analyse automatique de textes*, Masson, Paris.

Silberztein M. 2002. *NooJ Manual*. Available for download at : www.nooj4nlp.net.

Szpektor I., Dagan I., Lavie A., Shacham D., Wintner S. 2007. *Cross Lingual and Semantic Retrieval for Cultural Heritage Appreciation*. Proceedings of the ACL Workshop on Language Technology for Cultural Heritage Data, Prague, Czech Republic.

Vietri S., Elia A., D'Agostino E. 2004. *Lexicon-grammar, Electronic Dictionaries and Local Grammars in Italian*, in Laporte, E., Leclère, C., Piot, M., Silberztein M. (eds.), *Syntaxe, Lexique et Lexique-Grammaire. Volume dédié à Maurice Gross, Lingvisticae Investigationes Supplementa* 24, John Benjamins, Amsterdam/Philadelphia.

Volk, Martin, Spela Vintar, and Paul Buitelaar. 2003. *Ontologies in cross-language information retrieval.* Proceedings of WOW2003 (Workshop Ontologie-basiertes Wissensmanagement), Luzern, Switzerland.

Vossen P., A. Soroa, B. Zapirain, G. Rigau. 2012. *Cross-lingual event-mining using wordnet as a shared knowledge interface.* Proceedings of the 6th Global Wordnet Conference, C. Fellbaum, P. Vossen (Eds.), Publ. Tribun EU, Brno, Matsue, Japan, January 9-13 :382-390.

Yapomo M., Corpas G., and Mitkov R. 2012. *CLIR-and ontology-based approach for bilingual extraction of comparable documents.* The 5th Workshop on Building and Using Comparable Corpora.

Terminologie & Ontologie : Théories et Applications

Actes de la conférence

TOTh 2015

Le Bourget du Lac – 4 & 5 juin 2015

Terminologie & Ontologie : Théories et Applications

Actes de la conférence

TOTh 2015

Le Bourget du Lac – 4 & 5 juin 2015

http://toth.condillac.org

avec le soutien de :

Université Savoie Mont Blanc

École d'ingénieurs Polytech Annecy Chambéry

Ministère de la Culture - Direction Générale à la Langue Française et aux Langues de France

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

MINISTÈRE
DE LA CULTURE

Presses Universitaires Savoie Mont Blanc
Collection « Terminologica »

Comité scientifique

Président du Comité scientifique : Christophe Roche

Comité de pilotage

Rute Costa	Universidade Nova de Lisboa
Humbley John	Université Paris 7
Kockaert Hendrik	University of Leuven
Christophe Roche	Université Savoie Mont Blanc

Comité de programme 2019

Le comité de programme est constitué chaque année à partir du comité scientifique de TOTh en fonction des soumissions reçues. La composition du comité scientifique est accessible à l'adresse suivante : http://toth.condillac.org/committees

Guadelupe Aguado	Universidad Politécnica de Madrid – Spain
Amparo Alcina	Universitat Jaume I – Spain
Bruno Bachimont	Université Technologie de Compiègne – France
Jean-Paul Barthès	Université Technologie de Compiègne – France
Christopher Brewster	TNO – The Netherlands
Danielle Candel	CNRS, Université Paris Diderot – France
Sylviane Cardey	Université de Franche-Comté – France
Stéphane Chaudiron	Université de Lille 3 – France
Manuel Célio Conceição	Universidade do Algarve – Portugal
Rute Costa	Universidade NOVA de Lisboa – Portugal
Bruno Courbon	Université Laval – Canada
Lyne Da Sylva	Université de Montréal – Canada
Luc Damas	Université Savoie Mont-Blanc – France
Éric De La Clergery	INRIA – France
Dardo De Vecchi	Kedge Business School – France
Valérie Delavigne	Université Paris 3 – France
Sylvie Després	Université Paris 13 – France
Juan Carlos Diaz Vasquez	EAFIT University – Colombia
Hanne Erdman Thomsen	Copenhagen Business School – Denmark
Pamela Faber	Universidad de Granada – Spain
Christiane Fellbaum	Princeton University – USA
Iolanda Galanes	Universidade de Vigo – Spain
Christian Galinski	INFOTERM – Austria
François Gaudin	Université de Rouen – France
Teodora Ghiviriga	Alexandru Ioan Cuza University – Romania
Jean-Yves Gresser	ancien Directeur à la Banque de France – France
Ollivier Haemmerlé	Université de Toulouse – France
Gernot Hebenstreit	University of Graz – Austria
Amanda Hicks	University of Florida – USA
John Humbley	Université Paris 7 – France

Kyo Kageura	University of Tokyo – Japan
Heba Lecocq	INALCO – France
Hélène Ledouble	Université de Toulon – France
Patrick Leroyer	Aarhus University – Denmark
Georg Löckinger	University of Applied Sciences Upper Austria – Austria
António Lucas Soares	University of Porto, INESC – Portugal
Bénédicte Madinier	Dispositif d'enrichissement de la langue française – France
Candida Jaci de Sousa Melo	Universidade Federal do Rio Grande do Norte – Brazil
Jean-Guy Meunier	Université de Montréal – Canada
Christine Michaux	Université de Mons – Belgium
Fidelma Ní Ghallchobhair	Foras na Gaeilge, Irish-Language Body – Ireland
Henrik Nilsson	TNC – Sweden
Silvia Piccini	Italian National Research Council – Italy
Suzanne Pinson	Université Paris Dauphine - France
Marina Platonova	Riga Technical University – Latvia
Thierry Poibeau	CNRS Lattice – France
Maria Pozzi	el colegio de méxico – Mexico
Michele Prandi	Università degli Studi di Genova – Italy
Jean Quirion	Université d'Ottawa – Canada
Renato Reinau	Suva – Switzerland
Christophe Roche	Université Savoie Mont Blanc – France
Mathieu Roche	CIRAD – France
Laurent Romary	INRIA & HUB-ISDL – Germany
Micaela Rossi	Università degli studi di Genova – Italy
Bernadette Sharp	Staffordshire University – Great Britain
Marcus Spies	Universität München - Germany
Anne Theissen	Université de Strasbourg – France
Philippe Thoiron	Université Lyon 2 – France
Marc Van Campenhoudt	Université libre de Bruxelles – Belgium
Kara Warburton	City University of Hong Kong – China
Maria Teresa Zanola	Università Cattolica del Sacro Cuore – Italy
Fabio Massimo Zanzotto	University of Roma – Italy

Avant-propos

La Terminologie est une discipline scientifique à part entière qui puise à de nombreux domaines dont la linguistique, la théorie de la connaissance et la logique. Pour que cette diversité soit une richesse, il faut lui offrir un cadre approprié au sein duquel elle puisse s'exprimer et s'épanouir : c'est une des raisons d'être des *Conférences TOTh*.

Dans ce contexte, la formation et la transmission des connaissances jouent un rôle essentiel. La *Formation TOTh* précédant la Conférence se déroule sur deux années consécutives dédiées pour l'une à la dimension linguistique et pour l'autre à la dimension conceptuelle de la terminologie, deux dimensions étroitement liées.

A la présentation de travaux sélectionnés par un Comité de programme international, la *Conférence TOTh* inclut une *Conférence invitée* et une *Disputatio*. La première, donnée par une personnalité reconnue dans son domaine vise l'ouverture à d'autres approches de la langue et de la connaissance. La seconde, à travers une lecture commentée effectuée par un membre du comité scientifique, renoue avec une forme d'enseignement et de recherche héritée de la scolastique.

Cette année notre collègue Jean Pruvost, Professeur à l'Université de Cergy-Pontoise, a ouvert la Conférence TOTh par un exposé sur le thème des « *Dictionnaires et encyclopédies : un phénix au don d'ubiquité* ». John Humbley, Professeur à l'Université Paris Diderot, nous a parlé, dans le prolongement de la conférence d'ouverture et dans le cadre de la *Disputatio*, de « *L'encyclopédisme du dix-huitième siècle : prémices de la terminologie et de la langue de spécialité* ».

La conférence s'est terminée par le compte-rendu de la Journée d'Etude TOTh 2014 « Multilingual thesaurus and terminology » du 5 novembre 2014 (Bruxelles, Belgique) et par la présentation de la prochaine Journée d'Etude TOTh 2015 « Standards in Terminology » qui se déroulera le 3 décembre prochain à Luxembourg.

Les 17 communications, auxquelles se sont ajoutés 4 posters, ont permis d'aborder de nombreux sujets tant théoriques que pratiques rappelant qu'il ne peut y avoir de Terminologie sans langue ni savoir de spécialité. La Conférence s'est terminée par un compte rendu de la *Journée d'étude TOTh 2014* « *Multilingual Thesaurus and Terminology* » qui s'est déroulée le 5 décembre 2014 à Bruxelles et par la présentation de la *Journée d'étude TOTh 2015* qui portera sur « *Standards in Terminology* », le 3 décembre 2015 au Luxembourg.

Je vous invite à découvrir la *Disputatio* ainsi que les 9 articles longs et l'article court retenus pour publication dans les actes de TOTh 2015.

Christophe Roche

Président du comité scientifique

Sommaire

Disputatio 263

L'encyclopédisme du dix-huitième siècle : prémisses de la terminologie et de la langue de spécialité
John Humbley 265

Articles 291

Why read when you can watch ? Video articles and knowledge representation within the medical domain
Sara Carvalho, Christophe Roche, Rute Costa 293

PLOTITERM : modélisation de la terminologie de Plotin en OWL
Piccini, Silvia 313

Réingénierie de thésaurus pour l'indexation en Sciences du Langage. Retour d'expérience
Mario Marcon 345

Besoins de dénomination et classification dans les nomenclatures botaniques : termes et concepts en action
Claudio Grimaldi 369

L'implantation terminologique 20 ans après : le cas de la télédétection aérospatiale
Adam Renwick 379

Linguistics Terminology in Greek Translational Equivalents from English to Greek
Foteini Kalamida 403

Indices lexico-syntaxiques pour la reconnaissance des termes de la forme N_Adj dans les textes intégraux
Laurence Kister, Évelyne Jacquey, Mario Marcon, Sabine Barreaux 421

False borrowings in fashion language : a corpus-driven analysis
Valeria Fiasco 435

Une démarche ontoterminologique pour la rétro-ingénierie de logiciels
Françoise Deloule, Yves Dumond, Christophe Roche
Philippe Dugerdil, Eric Harth
Isabelle Chaussod, Jean-François Falconnier
David Corgier, Florian Simon 455

Disputatio

L'encyclopédisme du dix-huitième siècle : prémisses de la terminologie et de la langue de spécialité

John Humbley

Université Paris-Diderot
John.Humbley@eila.univ-paris-diderot.fr

Résumé. Les encyclopédies de la fin du dix-huitième siècle regorgent de ce que nous appelons aujourd'hui la terminologie et la langue de spécialité. Cette évolution ne doit rien au hasard : la séparation entre langue générale et langage des arts remonte au moins à la fin du dix-septième siècle, lorsque l'Académie française opéra une distinction méthodologique entre dictionnaire de langue et dictionnaire des arts. Un siècle plus tard, en apparaîtra l'illustration la plus connue sous la forme de la publication de l'*Encyclopédie*. Si son rôle dans la constitution d'une terminologie conceptuelle est connu – on y trouve notamment l'annonce de la réflexion de Lavoisier sur la nomenclature de la chimie – la recherche contemporaine n'a pas encore problématisé l'émergence d'une conception de ce que les francophones appellent (depuis les années 1970 seulement !) la langue de spécialité. Or, la réflexion sur les moyens linguistiques qu'il convient de déployer afin d'exprimer des connaissances spécialisées, voire des pratiques des « arts », est au cœur de la méthode des encyclopédistes. Leur action fut décisive, car elle a permis l'émergence d'une langue scientifique et surtout technique non seulement en France, mais dans les pays d'Europe de l'ouest et du nord. Au terme d'une étude de quelques encyclopédies de cette époque, nous proposons quelques pistes de recherche.

1. Introduction

Tout l'intérêt de TOTh est de faire dialoguer terminologues et cogniticiens, ou, plus généralement, comme le disait Christophe Roche lors d'un des

premiers colloques[1], tous ceux qui s'y intéressent et qui sont impliqués dans la dimension linguistique ou langagière de l'ingénierie des connaissances. Compte tenu des formations très différentes des participants, toutefois, c'est un dialogue qui n'est pas facile à amorcer ni à maintenir, et c'est en cela que ces rencontres représentent un défi. Défi que certains se sentent peu préparés à relever. C'est à l'invitation de Christophe Roche que je propose une réflexion sur les origines que je crois communes entre terminologie et ingénierie des connaissances et qui par ailleurs rejoint la thématique générale annoncée par Jean Pruvost[2] sous la forme de l'héritage des dictionnaires et encyclopédies.

L'expérience de l'histoire récente de la collaboration entre terminologie et ingénierie des connaissances confirme la nécessité d'un dialogue. Que de temps perdu naguère lorsque les uns ignoraient le travail - et les acquis - des autres ! Chacun se souvient des prototypes de systèmes experts alimentés en définitions tirées de dictionnaires de la langue générale, alors qu'il existait des ressources terminologiques validées par les experts des domaines concernés mais qui restaient inexploitées, tout simplement parce que trop peu connues à l'époque. Pour les terminologues, la rencontre avec les cogniticiens a été décisive : la terminologie est sortie de ses applications exclusivement langagières – l'aménagement linguistique et la traduction technique, pour résumer rapidement – pour se mesurer à de nouveaux défis, qui faisaient ressortir des insuffisances linguistiques de sa pratique. Le fameux exemple du rond-point que Daniel Kayser avait présenté à la première journée TIA[3] à Villetaneuse en 1995 illustre bien l'ironie de la situation : c'est l'informaticien qui fait une leçon de sémantique aux terminologues. On se souvient qu'il s'agissait dans son exposé de faire ressortir les problèmes rencontrés dans la construction d'un système expert pour traiter les constats d'accidents et les demandes d'indemnisations auprès d'un assureur. Kayser soulève plusieurs questions d'ordre terminologique, celle concernant le rond-point étant la plus simple. Le schéma est tiré du constat à l'étude, qui pose le problème de la définition. Que représente en fait *rond-point* ?

1 « Les conférences TOTh ont pour objectif de rassembler industriels, chercheurs, utilisateurs et formateurs dont les préoccupations relèvent à la fois de la terminologie et de l'ontologie et, de façon plus générale, de la langue et de l'ingénierie des connaissances. » Christophe Roche (2007).

2 La conférence inaugurale de TOTh 2015, prononcée par Jean Pruvost, est intitulée *Dictionnaires et encyclopédies : un phénix au don d'ubiquité.*

3 *Terminologie et intelligence artificielle*, dont le premier colloque est paru dans *La banque des mots*. Numéro spécial 7, Paris, CILF, Otman (éd.), 1995.

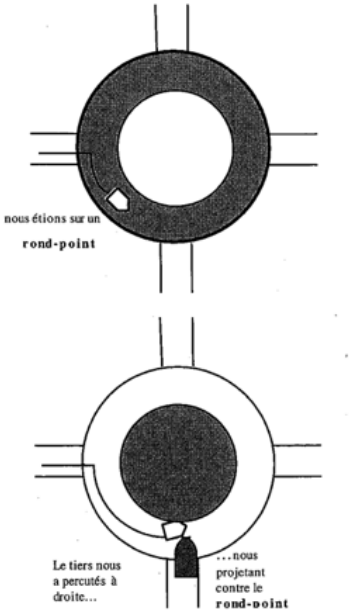


Tableau 1

Dans son constat, l'assuré emploie *rond-point* pour désigner deux référents différents, différence passée inaperçue lors du traitement humain, et qui d'ailleurs n'était pas susceptible d'analyse automatique simple. Dans le premier cas, il s'agit de l'ensemble, dans le second d'une partie. Simple question de métonymie, dirait-on. Question qui ne relève pas de la terminologie, car d'un usage général et non spécialisé. Peut-être, mais comme le dit Kayser (1995 : 19) :

> *Dans les langues de spécialité, il paraît au contraire raisonnable de supposer que U est fixé dès le départ, précisément pourrait-on dire parce qu'on se place dans le cadre d'une spécialité.*

Or, cet *a priori* simpliste mais bien commode se révèle intenable :

> *… la capacité de nomination n'est pas épuisée par la définition d'un mot ou d'un terme ; elle est à l'œuvre en permanence.* Kayser (1995 : 32).

En d'autres mots, on a besoin, pour que le système-expert puisse fonctionner, d'un outil d'analyse du fonctionnement de la langue qui permette de distinguer entre les deux ronds-points. Jusqu'où le dictionnaire – ou la base de données terminologiques voire ontoterminologique – peut-il/elle aller, compte tenu de la décontextualisation inhérente à ce genre d'outil ? Bonne question : merci aux informaticiens de l'avoir posée. La tradition se perpétue, d'ailleurs, de cette collaboration, de la part du LIPN[4] – nous avons vu à TOTh Sylvie Despres et Sylvie Szulman (2008), par exemple, et bien d'autres, et ce genre de question ne reste plus sans réponse.

Pour poursuivre le dialogue, deux voies semblent possibles : celle tournée vers l'avenir : des projets où linguistes et cogniticiens collaborent (on pense aux discussions avec Michele De Gioia et Mario Marcon (2014), qui ont besoin d'ontologiser une terminologie de la médiation) et celle qui réfléchit sur le passé et les origines communes. C'est cette deuxième voie que je me propose d'explorer.

Toute science, toute discipline a des origines multiples. Une partie du processus de la disciplinarisation est la recherche des origines et des ancêtres : les disciplines relativement récentes que sont la terminologie et l'ingénierie des connaissances n'échappent pas à cette règle, et dans cette présentation j'aimerais remonter un peu plus loin que Wüster, souvent présenté comme père de la terminologie moderne, pour retrouver une des origines, d'abord de la terminologie et en complément de l'ingénierie des connaissances. L'hypothèse est que ces origines se manifestent de manière privilégiée dans les dictionnaires et les encyclopédies du siècle des Lumières.

Plus précisément, on peut situer dans ces supports de type lexicographique une première théorisation de la langue de spécialité et de la terminologie. La méthode va être simple : d'abord isoler les traits qui caractérisent la terminologie et la langue de spécialité. Ensuite, rechercher ces mêmes traits ou leur équivalent dans les dictionnaires et encyclopédies du XVIII[e] siècle.

4 Laboratoire d'informatique de Paris Nord UMR 7030. https://lipn.univ-paris13.fr/fr/. Ce laboratoire s'engage depuis longtemps dans le dialogue cogniticiens/terminologues, ayant organisé plusieurs colloques du groupe TIA, dont le premier.

2. Terminologie et langue de spécialité : les traits définitoires

Un mot d'explication s'impose toutefois : pourquoi parler non seulement de la terminologie mais aussi des langues de spécialité ou de la langue de spécialité ? La raison tient à l'historique de l'évolution de la discipline de la terminologie et des relations étroites, voire constitutives des deux, surtout au moment où elles se constituaient en discipline. Les définitions de la terminologie (et donc de ses traits distinctifs) sont diverses et variées – elles ont fait l'objet de discussions à TOTh et peuvent être considéré comme une primitive ; celles de la langue de spécialité sont encore plus diversifiées et l'instabilité de la dénomination en est symptomatique.

En schématisant à outrance, on peut identifier deux prototypes des langues de spécialité ;

l'approche germanique, dont le but principal est de rendre compte de l'expression, de la communication de connaissances spécialisées sous toutes ses formes, langagière ou non et l'approche anglo-saxonne, qui se résume – en schématisant encore – à une technique d'enseignement de l'anglais destiné aux étudiants non anglophones (Humbley 2007). La démarche française se situerait sans doute un peu entre les deux selon la tendance : plus germanophile pour Kocourek ([1982] 1991), digne représentant de l'école de Prague, plus du côté des anglophones sous sa forme Français sur objectifs spécifiques (FOS).

Dans les pays de langue allemande la langue de spécialité reste un concept fondamental, peut-être parce qu'il n'est pas – au départ tout au moins – un terme. *Fachsprache* est un mot de tous les jours, qui s'est défini progressivement dans les années 20 et 30. *Langue de spécialité* en revanche est un écho des années 1970 de l'anglais *langues for special purposes*, première indication d'une dérive didactique. Les manuels de langue de spécialité – qui prennent en compte la dimension didactique – sont des succès de librairie dans les pays de langue allemande : le Fluck (*Fachsprachen. Einführung und Bibliographie)* en est à sa cinquième édition (Fluck [1976] 1997), celui de Roelcke ([1999] 2005), à sa troisième… On en parle ici car c'est *Fachsprache*, à la germanique, qui se rapproche le plus d'une conception systématique des questions liées à la communication spécialisée. L'apport massif des germanophones à la définition de la terminologie en tant que discipline émergeante est important, car ils la situaient au sein des langues de spécialité. Cette vision des choses, cette relation d'inclusion est présentée dans un petit manuel qui accompagnait une des premières formations de terminologie, probablement la

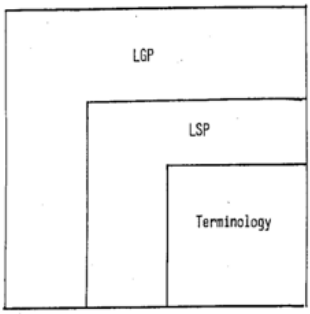
première en langue anglaise, conçue au début des années 1980 par Heribert Picht et Jennifer Draskau (1985) pour l'université de Surrey à Guildford. Ce n'est pas par hasard que l'essentiel de la terminologie anglaise (en dehors de Manchester, où travaille encore Juan Carlos Sager, longtemps professeur à l'UMIST), a été initié sinon réalisé dans cette université où travaillent en étroite collaboration la germaniste Margaret Rogers et le physicien/informaticien Khurshid Ahmad.

Pour la tradition germanique, dont Picht et Draskau sont les porte-parole, la terminologie se situe fermement dans le cadre de la langue de spécialité, ou *LSP* pour employer le sigle anglais :

- *position occupied within the framework of 'LSP' by terminology.* (Picht et Draskau 1985 :1)
- *it is equally clear that terminology is part of LSP.* (Picht et Draskau 1985 : 21)

La relation est schématisée ainsi :

This may be illustrated graphically thus:



1. The language of a given language community as a whole
2. - within which may be distinguished the area covered by LGP and LSP varieties
3. - and one section of LSP comprises special lexis - this is the central domain of terminology.

Il est non moins significatif que Picht rappelle les liens avec d'autres disciplines qui ne relèvent pas de la linguistique, à savoir la théorie de la cognition,

l'information-communication et la recherche documentaire, et bien entendu les domaines faisant partie des études de terminologie.

- *'Terminology interfaces with other, not traditionally linguistic disciplines*
  - *logic (philosophy), theory of cognition*
  - *information and documentation (classification, documentation search)*
  - *subject fields in the widest sense'*
    (Picht et Draskau 1985 : 22)

Sager (1990 : 9), à peine plus tard, souligne pour sa part le lien avec les sciences de l'information. Pour lui, il s'agit de nouvelles « directions d'étude » en quête d'identité. La langue de spécialité nous intéresse donc parce qu'elle permet sous sa forme pragoise un rapprochement avec les sciences de la cognition.

3. Éléments de définition ; terminologie, langue de spécialité

Résumons maintenant les traits que l'on peut considérer comme typiques sinon définitoires de la terminologie et de la langue de spécialité ; ce seront ces caractéristiques que nous rechercherons par la suite dans les dictionnaires et encyclopédies. Pour les besoins de la démonstration nous nous en limitons à sept, mais on pourrait en invoquer d'autres. Chaque trait sera illustré et, le cas échéant, justifié au moment de la comparaison dans la troisième partie.

1. Langue générale/langue de spécialité

Cette dichotomie est une distinction fondatrice des études de langue de spécialité, que l'on retrouve dans les manuels de langue de spécialité (Kocourek ([1982] 1991) pour le français, Sager, Dungworth et Macdonald (1980) pour l'anglais) comme de terminologie. Elle est du même ordre que celle qui existe entre lexicographie et terminographie (Wright 1991). Il s'agit bien entendu des catégories prototypiques, et on pourrait envisager des échelles ou des degrés de spécialisation.

2. Approche onomasiologique

Fidèle à sa conception de la primauté du concept par rapport à son expression en langue, la terminologie classique réserve un traitement privilégié à l'approche l'onomasiologique (Wüster 1979). Il s'agit encore une fois d'une

distinction souvent évoquée dans les premières définitions de la terminologie et de la terminographie, et centrale en lexicographie. On se souvient par ailleurs, que la thèse de Jean Pruvost (1981) porte sur les dictionnaires onomasiologiques.

Mais organiser le vocabulaire implique une structuration, voire une modélisation des connaissances, notre catégorie suivante :

3. Structuration, modélisation des connaissances

Ce critère encore se trouve au centre de la réflexion sur les langues de spécialité et surtout de la terminologie, et constitue le principal lien avec les sciences cognitives. Il est suffisamment développé dans le cadre des rencontres TOTh pour qu'on n'ait pas à l'expliciter davantage.

4. Construction de système sémiotique

Lorsqu'on parle de langue de spécialité, on sous-entend la relation qui existe avec différents codes, dont la langue naturelle, qui constituent un système sémiotique (Kocourek 1991). Les textes de spécialité sont des combinaisons de différents codes, dont l'interaction constitue un des principaux enjeux de communication. Le texte scientifique et technique écrit comporte le plus souvent tableaux, schémas, formules, icones et autres, qui sont reliés expressément et implicitement par la langue. Ces aspects sont souvent évoqués par les spécialistes des langues de spécialité, mais parfois minoré par les linguistes, qui éliminent tout ce qui n'est pas texte afin de constituer un corpus.

5. Visée de normalisation/standardisation

D'un point de vue historique, la normalisation était la première application de la terminologie, ce que les linguistes lui ont souvent reproché, en particulier dans le contexte de la socioterminologie (cf. Gambier 1991).

6. Intervention sur la langue

Critère déterminant pour la terminologie à orientation politique, mais aussi dans le cadre de la normalisation, axe privilégié par Wüster (1979).

7. Dimension internationale

Cette dimension marquait les débuts de la terminologie : on peut considérer que la terminographie internationale fut inaugurée en 1906, lorsque la Commission électrotechnique internationale (CEI), décida de faire rédiger un dictionnaire thématique plurilingue. On sait par ailleurs que Wüster était

espérantiste et que ses ambitions de promouvoir la collaboration scientifique internationale (Samain 2010) se traduisaient par ses activités de terminologie.

Ces sept critères sont des éléments centraux dans la définition de ce qui allait devenir la discipline de la terminologie, mais il aurait été possible d'en inclure d'autres, et non les moindres : j'exclus délibérément le concept ou la notion… de *notion*, que Loïc Depecker et Christophe Roche avaient problématisé en 2007. C'est une question qui aurait pris tout le temps de l'intervention ; il vaut mieux pour une présentation de type panorama de parcourir davantage de terrain et de rechercher davantage de points de vue.

3.1. La question de la transmission et des sources

Avant de partir à la recherche de ces critères dans les dictionnaires du passé, il convient de se poser deux questions préalables : une qui restera à l'état de question et l'autre plus pratique, et pour laquelle on peut indiquer des réponses.

La première concerne les modes de transition. À supposer que l'on retrouve dans les dictionnaires et encyclopédies tous les traits que l'on considère comme caractéristique de la réflexion terminologique moderne, comment savoir s'ils ont été transmis de génération en génération – représentant une espèce de tradition savante – ou s'il s'agit au contraire de redécouverte, c'est-à-dire qu'au vingtième siècle on est arrivé à des conclusions similaires par une tout autre voie ? On relève dans les écrits de certains terminologues des références sinon aux encyclopédies du moins à ceux qui les ont alimentés : Wüster (1979) parle de Lavoisier ainsi que Picht[5]. Mais la question mérite une réponse plus circonstanciée.

L'autre question est une de méthode et concerne les sources historiques. Pour les débuts, la préhistoire en ce qui nous concerne, nous nous limitons essentiellement au dictionnaire posthume d'Antoine Furetière (*Dictionnaire universel*, 1690), tout en tenant compte de la double offre de l'Académie française, leur propre dictionnaire (1694) et *Dictionnaire des Arts et des Sciences* confié à Thomas Corneille et publié la même année. Pour le dix-huitième nous n'aurons pas le temps d'examiner le « Trévoux », (*Dictionnaire Universel françois et latin vulgairement appelé Dictionnaire de Trévoux,* 1704), pourtant lien indispensable entre les dictionnaires spécialisés et l'encyclopédisme

5 Picht et Draskau 1985 : 23 "Lavoisier, who was not a linguist, clearly recognized the interdependence of concept, term, motivation of the term (i.e. its expressivity, and its communicative importance."

qui se formait, et nous nous limitons à trois publications emblématiques dans le domaine de la chimie, le *Dictionnaire de Chymie* de Macquer de 1766, l'*Encyclopédie* de Diderot et d'Alembert (1751-1779), (Daumas 1951) bien sûr et le successeur, l'*Encyclopédie méthodique* de Panckoucke (1782-1832), sous la forme du premier volume sur la chimie.(*Dictionnaire de la Chymie* de Guyton de Morveau, 1786)[6].

Mais comment les consulter ? Depuis la numérisation des fonds classiques il est désormais possible de consulter beaucoup de dictionnaires anciens sans quitter son fauteuil. On dit qu'il n'est plus possible de faire la terminologie en chambre, mais ce n'est que plus vrai pour la consultation de fonds documentaires. Deux grandes sources s'offrent à nous : le musée virtuel des dictionnaires de Cergy Pontoise, conçu et réalisé par Jean Pruvost[7] et l'ATILF[8], successeur de l'INaLF, à Nancy, où l'on peut consulter l'*Encyclopédie*, numérisée avec le concours de l'Université de Chicago. Le *Dictionnaire méthodique* de Panckoucke représente un tel défi, compte tenu de sa taille, que les différentes institutions ont repoussé sa numérisation. Il se trouve toutefois que le volume de 1786 de Guyton de Morveau est entièrement disponible sur Google Books, du moins par moment. L'Académie des sciences a mis en ligne l'Édition Numérique Collaborative et CRitique de l'Encyclopédie de Diderot, de D'Alembert et de Jaucourt (1751-1772)[9].

Lorsqu'on se lance dans la lecture de dictionnaires anciens, on a besoin de guide. Les études de Jean Pruvost représentent le point de passage obligé, à commencer par le *Que sais-je ?* (Pruvost 2002) et surtout *Les dictionnaires français, outils d'une langue et d'une culture* (2006). Mais l'essentiel, du moins pour la perspective qui nous concerne ici est fourni, par les collègues en histoire des sciences, qui exploitent depuis très longtemps le témoignage des dictionnaires et des encyclopédies. Je ne citerai ici qu'un ouvrage parmi plusieurs de Bernadette Bensaude-Vincent (1993) sur la révolution de la chimie de la fin du dix-huitième siècle, très utile pour la démonstration qui suit. Les linguistes sont relativement peu nombreux à avoir investi cette thématique. On peut citer François Gaudin (2003), qui s'intéresse à l'histoire des dictionnaires et à la dimension sociale de la terminologie, Christophe Rey (2004), spécialiste incontesté du Panckoucke, qui nous fait découvrir la réflexion

6 Nous nous limitons aux sources secondaires, à l'exclusion de documents primaires tels que les écrits de Lavoisier et de ses collaborateurs (Delaveau et Rodolphe 2006).
7 https://dictionnaires.u-cergy.fr/accueil.html.
8 http://www.atilf.fr.
9 Accessible depuis http://enccre.academie-sciences.fr/encyclopedie.

en sciences du langage contenue dans ses pages, mais surtout Maria Teresa Zanola, dont le livre de 2014 nous ouvre tant de pistes, et qui a profondément réorienté notre présentation.

4. Comparaison terminologie du vingtième siècle, lexicographie du dix-huitième

Dans les lignes qui suivent, chaque proposition énumérée plus haut est examinée d'abord sous l'angle de son expression dans le corps de doctrine de la terminologie classique, celle qui s'inspire des études de langues de spécialité telle que la tradition germanique la conçoit, puis telle qu'elle apparaît dans les dictionnaires et encyclopédies du passé.

4.1. Division langue générale/ langue de spécialité

Cette dichotomie a longtemps occupé les terminologues, surtout au moment où la discipline était en gestation. Nous avons déjà fait allusion à la profusion de dénominations en français (*langue de spécialité* – singulier ou pluriel ; *langue spécialisée* ; *LSP ; langue à usage spécifique*, etc. Kocourek 1991, Rondeau 1984), tout comme l'anglais : *language for special purposes, language for specific purposes, special language(s)*, etc., reflet à la fois du périmètre du concept qui varie selon les points de vue, mais surtout de la nature du phénomène. Pierre Lerat (1995), en préférant *langues spécialisées*, se rapproche en fait de Sager (1990), qui lui parle de *special language*, pour souligner qu'il s'agit d'un usage particulier d'une langue. Le fait que cette division entre langue générale et langue de spécialité ou langue spéciale soit violemment contestée par les socioterminologues (Gambier 1991, 1998) ne fait que souligner son importance dans le processus de la constitution de la discipline. Autrefois une ligne de fracture, elle s'estompe peu à peu, on remarque que les manuels les plus récents en parlent moins. Marie-Claude L'Homme, par exemple, ne semble pas l'employer dans son manuel de 2004 : elle parle en revanche de *textes spécialisés*, reflet d'un changement d'orientation depuis que la terminologie peut être fondée sur des corpus textuels significatifs. Il n'en reste pas moins que l'idée de langue de spécialité a constitué un des traits définitoire de la terminologie, surtout au début de son histoire. Or cette division était déjà bien présente dans la lexicographie du dix-septième siècle : l'Académie a opté pour une séparation nette et consciente au niveau de la lexicographie : le dictionnaire de langue d'un côté, et des « arts et des sciences » de l'autre.

> « C'est dans cet estat [de perfection] où la Langue Françoise se trouve aujourd'huy qu'a esté composé ce Dictionnaire ; & pour la representer dans ce mesme estat, l'Académie a jugé qu'elle ne devoit pas y mettre les vieux mots qui sont entierement hors d'usage, ni les termes des Arts & des Sciences qui entrent rarement dans le Discours ; Elle s'est retranchée à la Langue commune, telle qu'elle est dans le commerce ordinaire des honnestes gens, & telle que les Orateurs & les Poëtes l'employent. *Introduction, Dictionnaire de l'Académie française.*

Ces « termes des arts et des sciences » sont consignés dans le *Dictionnaire des Arts et des sciences* de Thomas Corneille (1694). On se souvient que les arts, dans ce contexte, comme M-T. Zanola (2014 : 41-56) le rappelle très utilement, renvoient à ce que nous appelons aujourd'hui les techniques ou la technologie, et on remarque que l'Académie les sépare déjà : c'est une sous-division qui sera très importante au siècle suivant. Mais ce que l'Académie sépare, Furetière réunit. Le sous-titre de son dictionnaire l'annonce clairement : *Dictionnaire universel Contenant généralement tous les mots françois. Tant vieux que modernes, & les Termes des sciences et des arts, divisé en trois tomes…*

On voit qu'il anticipe la division de l'Académie (à moins que ce soit l'Académie qui répond à Furetière ?), y compris la sous-division entre vocabulaire des sciences et des arts (indiquée en grand !) et qu'il parle nommément de *termes*. Voilà un des précurseurs importants de la terminologie, auquel nous reviendrons encore. Mais cette division s'est perpétuée au dix-septième siècle, et la réflexion s'est approfondie : Guyton de Morveau, dans son volume du Panckoucke, aborde cette question :

> *On est convenu dans* ***le langage ordinaire*** *d'appeler acide toute substance qui excite sur la langue une sensation vive et piquante, mais dans* ***le langage de la science,*** *cette explication ne donneroit qu'une idée bien imparfaite de ce qu'on doit entendre par le mot acide, parce que cette saveur n'est qu'un des plus foibles effets de l'action des acides, quoiqu'elle dérive nécessairement des propriétés bien plus importantes qui sont essentielles à tous les corps de cette classe.* Guyton de Morveau, 1786, Article *acide* dans *Encyclopédie méthodique*

Dans ce très long article sur *acide*, il explique la différence entre un mot de la langue générale (appelé ici *langage ordinaire*) et le terme en chimie (appelé *langage de la science*) : la même dénomination, mais des acceptions très dif-

férentes, parce que celui-ci entre dans le cadre d'une structuration manifestée par ses propriétés, qu'il énumère.

Qu'en reste-t-il aujourd'hui de cette distinction, qui remonte en réalité aux débuts de la lexicographie française ? Il est clair qu'elle était fondatrice dans le cadre de l'émergence de la terminologie en tant que discipline, mais elle est considérée aujourd'hui comme moins fondamentale, puisque la recherche s'est réorientée depuis une vingtaine d'années vers les aspects textuels (et donc discursifs) et ethnologiques. Il n'en reste pas moins que les origines de cette dichotomie se trouvent abondamment illustrées dans les dictionnaires et encyclopédies du XVII^{e} et XVIII^{e} siècles, du moins pour la France. Il serait intéressant de mener une recherche comparative sur le même type de ressource dans les autres langues de culture de l'époque, d'autant plus que c'est là où le français scientifique mais aussi l'allemand, l'anglais, l'italien (sans oublier le néerlandais, le suédois) scientifiques ont été forgés.

4.2. Approche onomasiologique

Ce deuxième trait peut s'exprimer de différentes façons : il découle, comme nous l'avons vu, du postulat de la primauté du concept par rapport à son expression en langue. Traduites en termes lexicographiques, ceci aboutit à une présentation thématique ou onomasiologique. Pour Wüster la règle numéro 1 est de commencer par le concept et il préconise une présentation « logique », comme il le dit dès la première page de son manuel : *Ausgehen von den Begriffen : Erstens : Jede Terminologiearbeit geht von den Begriffen aus.* (On commence par les concepts. Première règle : Tout travail de terminologie commence par les concepts ; Wüster 1979 : 1). Sue Ellen Wright (1991) pour sa part l'a présenté ainsi, d'un point de vue lexical, à l'association internationale de normalisation.

La réalisation idéale de cette présentation, compte tenu des possibilités de l'époque, est sans aucun doute le dictionnaire de la machine-outil (Wüster 1968), qui suit l'ordre de la CDU – classification décimale universelle. On voit ici le lien étroit entre terminologie et documentation (systèmes de documentation). Ce lien est problématisé dans le travail de Maryvonne Holzem (1999), qui souligne la parenté dans la conception et dans la méthode. Par la suite, les avancées de la terminologie textuelle ont montré la nécessité d'une étude sémasiologique systématique, et la dimension onomasiologique a été minorisée, sauf dans les projets plurilingues. Il n'en reste pas moins que cette approche a été déterminante dans la disciplinarisation de la terminologie.

Pierre-Joseph Macquer (voir aussi Zanola 2014 : 95), auteur du grand dictionnaire de la chimie du milieu du dix-huitième siècle, s'est beaucoup interrogé sur la présentation des données, et reconnaît bien les limites de l'ordre alphabétique.

> *J'avoue que j'ai eu d'abord quelque peine à me prêter à l'exécution de ce projet, parcequ'il me sembloit que toutes les parties de la Chymie étant liées entr'elles et dépendantes les unes des autres, cette Scienceétait peu propice à être traitée dans l'ordre alphabétique. [...]*
>
> *A la vérité, la disposition alphabétique paroît interrompre et déranger toute espèce de plan et de Système dans une Science, mais on peut remédier à ce défaut par des renvois qui établissent la liaison nécessaire entre tous les articles correspondans, comme on l'a fait dans cet Ouvrage…* […]
> *Dictionnaire de Chymie* (1766)

N'oublions pas que Macquer fut un grand pédagogue, et que son dictionnaire complète deux manuels qu'il avait déjà publiés (sous son nom), de chimie théorique et pratique (Viel 2004 : 263). Le dictionnaire est donc complémentaire par rapport aux manuels, et propose d'autres portes d'entrées dans la science de la chimie. L'ordre alphabétique est sans doute bête, mais il est aussi pratique. Pour son dictionnaire il se résout à garder l'ordre alphabétique mais compense en multipliant les renvois. Dans la deuxième édition de 1778 il joint trois tables, dont la première est l'ordre dans lequel il convient de lire les articles, soit la reconnaissance du besoin intellectuel d'ordonner les connaissances, et en même temps du caractère pratique de l'ordre arbitraire – alphabétique ; la deuxième une table des matières, la troisième un index des noms propres.

On peut faire valoir par ailleurs le caractère foncièrement onomasiologique des planches, qui constituent 11 tomes de l'*Encyclopédie*, et nous reviendrons plus loin à la question des illustrations. Il suffit de conclure en disant que le bien-fondé de la démarche onomasiologique et certains de ces emplois furent acceptés au dix-huitième siècle, même si la présentation alphabétique gardait encore tous ces droits.

Tout comme la dichotomie langue générale/langue de spécialité, celle de l'onomasiologie/sémasiologie n'est plus au cœur de la théorie terminologique, mais elle a fortement participé à la définition de la nouvelle discipline. Ce critère, bien présent dans les répertoires du dix-huitième siècle, est néan-

moins moins fondamental que la distinction entre langue générale et spécialisée, mais mériterait une étude plus approfondie surtout dans une perspective paneuropéenne.

4.3. Modélisation des connaissances

C'est par cette expression quelque peu anachronique, même pour la terminologie classique, que nous désignons les tentatives de matérialisations de structures conceptuelles. Nous avons vu que Wüster plaçait au centre de ses préoccupations le fait de dégager un système conceptuel, qu'il distingue dans son manuel posthume de la présentation thématique[10].

On peut dire que l'arbre du domaine, tel que nous le connaissons aujourd'hui nous vient du Canada. Dubuc (1992 : 39) en parle comme quelque chose d'acquis dès la première édition de son manuel, et il représente le plus souvent un mélange (plutôt qu'une synthèse) de systèmes conceptuels et thématiques. Il a fallu attendre l'apport des ontologies pour y voir plus clair – précisément le but de TOTh.

Or, cette structuration est déjà incorporée, certes dans un état embryonnaire, dans les articles du Furetière, surtout dans ceux qui traitent d'instruments ou d'objets fabriqués. Sous l'entrée *orgue* (voir la page du Musée virtuel des dictionnaires de Cergy-Pontoise[11]), on voit que les relations conceptuelles qu'invoque Wüster sont bien représentées : on a les relations génériques (les orgues portatives sont des sortes d'orgues), les relations méronymiques (ontologiques) : les parties de l'orgue sont déclinées, sans oublier leurs fonctions. On a aussi dans ce cas la mention d'une autre question non moins fondamentale – celle de la synonymie, qui sera développée surtout au XVIII^{e} siècle (Pruvost 2006 : 58) – pour la langue générale, mais aussi pour la langue de spécialité, où elle pose des problèmes spécifiques.

On voit que l'idée de la structuration des connaissances avait présidé à la conception même de l'*Encyclopédie*, comme ce bref extrait de l'entrée *encyclopédie* le montre bien. On peut d'ailleurs voir un parallèle avec l'idée

10 Tout ce qui ressemble à un système conceptuel, c'est-à-dire une classification de concepts, n'en est pas nécessairement un. Si les relations entre les éléments ne sont ni logiques, ni ontologiques on ne peut avoir un système conceptuel : il s'agit alors de relations thématiques. Wüster 1979 : 23-24. Il attribue ces derniers à la documentation.

11 https://dictionnaires.u-cergy.fr/auteurs/Antoine_FURETIERE/Titre2/extrait_ORGUE-1.htm.

de Wüster de pouvoir transmettre l'ensemble des connaissances grâce à leur structuration.

> *En effet, le but d'une Encyclopédie est de rassembler les connoissances éparses sur la surface de la terre ; d'en exposer le système général aux hommes avec qui nous vivons, & de le transmettre aux hommes qui viendront après nous, afin que les travaux des siecles passés n'aient pas été des travaux inutiles pour les siecles qui succéderont* […].
> Diderot, 1751-1772, article ENCYCLOPÉDIE, cité par Jacquet-Pfau 2007

Elle prend une forme proprement lexicographique et terminologique dans le premier volume Chimie du Panckoucke, rédigée principalement par Guyton de Morveau. Le très long article *acide*, (436 pages !) n'a rien à voir avec une entrée dans une base de données terminologiques contemporaine, mais il explicite un certain nombre de démarches qui permettent justement d'arriver à cette structuration.

ACIDE [...] La définition de l'acide est, à vrai dire, la clef de la Chymie, non seulement parce que ce sont les acides qui produisent les plus beaux phénomènes & en plus grand nombre, parce qu'ils sont les instruments les plus habituels de ses opérations ; mais encore parce que ce sont des agents visibles et palpables dont les effets frappent nos sens, [...] de sorte que l'on peut dire que c'est la Chymie des acides qui a ébauché la Science, & qui sert de base au système général de nos connaissances dans cette partie de l'étude de la nature…

- Renvoi aux articles : *affinité, attraction, dissolution…*
- Listes des propriétés – 20 Volume Chimie (Guyton de Morveau)

Dans l'extrait ci-dessus, on peut signaler non seulement la recherche de critères structurants (les causes qui produisent le plus d'effets et la spécification de leur rôle) mais aussi une prise de position sur les conséquences de la démarche empirique, où l'on cherche à rendre compte de manière cohérente de ce qui est observé. Sur le plan lexicographique, ces soucis sont traduits, comme dans la deuxième édition du Macquer, par le renvoi aux termes associés – constituant ainsi un réseau conceptuel visuel – et la provision de tables qui systématisent les propriétés nommées. Ce critère est sans doute le plus

important et mérite des recherches plus approfondies, car les parallèles par rapport au développement de la terminologie au XX^{e} siècle sont frappants.

4.4. Construction de systèmes sémiotiques mixtes

Le quatrième critère est celui que les linguistes ont le plus tendance à oublier : la construction de systèmes sémiotiques forcément mixtes. Les textes scientifiques et techniques ne sont pas composés exclusivement de texte, mais aussi de très nombreux autres codes. Se focaliser sur la seule composante textuelle est sans doute le prix à payer pour les progrès accomplis grâce à la priorité accordée à la terminologie textuelle, suite à la mise à disposition de corpus – de grands corpus de langue ou de corpus spécialisés plus adaptés à l'analyse terminologique. La première tâche dans la constitution de ces corpus fut d'éliminer tout ce qui n'était pas la langue – schémas, cartes, tableaux, dessins, illustrations, etc. Or, c'est le jeu de ces différents codes qui est le propre d'une langue de spécialité, comme Kocourek, fidèle à l'école de Prague, le précise dès l'introduction de son livre.

La langue de spécialité appartient, à la fois, à deux vastes ensembles signifiants qui se chevauchent : à la sémiotique de spécialité et à la langue naturelle. […]

Le monde de spécialité est saisi à l'aide de divers systèmes sémiotiques de spécialité, qui conjuguent leurs ressources pour structurer et signifier ses différents aspects (Kocourek 1991 :11). Il procède par ailleurs à une classification des signes :

- *signes* ***non linéaires*** *tridimensionnels et bidimensionnels et les signes* ***linéaires*** *au sens linguistique (les éléments de la langue ou des langages symboliques qui se suivent spatialement ou temporellement).*
- *signes* ***iconiques*** *qui sont reconnaissables de ressemblent au signifié ou au manifesté (modèles, diapositives, schémas) et signes* ***non iconiques,*** *qui ne lui ressemblent pas (unités lexicales de langue, chiffes, symboles)*
- *signes* ***directs*** *(le calcul) et signes* ***substitutifs*** *(le morse) Encyclopédie des sciences industrielles Quillet.*
  Kocourek 1991 :10-11.

Il donne le détail des différents codes employés dans la communication spécialisée, et souligne l'enjeu de la mise en relation. C'est pour cette rai-

son aussi que le *Dictionnaire de la machine outil* (Wüster 1968) comporte de nombreux schémas et autres dessins.

Or, cette mise en relation des codes est précisément un des défis auxquels l'*Encyclopédie* a trouvé une solution, en démultipliant les planches qui permettent de visualiser en particulier le travail des scientifiques mais surtout des artisans – il s'agit bien des arts dont nous parlions ; 11 volumes de planches sur 17 volumes de texte. La présence sur les planches permet de voir le détail des relations méronymiques et surtout de comprendre les liens de fonctionnalité. Il a fallu attendre les volumes plus tardifs du Panckoucke pour que les illustrations soient mieux intégrées dans le texte des articles, comme dans un dictionnaire moderne (on pense ici au dictionnaire du capitaine Paasch, que Marc van Campenhoudt (1997) a mis à notre disposition dans sa thèse et dans de nombreuses publications – celle où il compare Paasch et Schlomann est particulièrement pertinente), mais le principe de l'interdépendance des codes était déjà pleinement reconnu dans le cadre de l'*Encyclopédie*, et la réflexion portait sur les moyens de les relier : encore une piste de recherche, à la fois pour l'histoire des sciences, dont celles des connaissances, et de la lexicographie.

4.5. Monosémie comme idéal

Le cinquième critère, déjà mentionné sous la rubrique de l'organisation des connaissances, est le rôle – généralement considéré comme négatif – de la synonymie. C'est encore une question fondatrice de la démarche de Wüster. Voilà ce qu'il en dit très tôt, tiré de son expérience du travail de normalisation :

> *Bei diesen Vereinbarungen handelt es sich nicht nur zu einem kleinen Teil um die bloße Ausmerzung von Synonymen. In sehr vielen Fällen werden die Begriffe überhaupt erst geprägt (oder gemodelt) und in ein System gebracht ; Auch die entsprechenden Benennungen werden neugebildet oder nach einer bestimmten Systematik umgeformt.*
>
> *Dans le cas de ces accords, il s'agit pour beaucoup de l'élimination pure et simple de synonymes. Dans bien des cas, les concepts sont d'abord formés (ou modelés) puis incorporés dans un système. De même les dénominations correspondantes sont créées ou refaçonnées pour correspondre à une systématique donnée. (notre traduction)*
> Wüster 1953 dans Picht, Schmitz (dir.) 2001 : 13

L'essentiel du travail, c'est l'élimination de la synonymie, pour arriver à un rapport plus direct entre concept et dénomination, dans l'idéal une rela-

tion de la biunivocité, la fameuse *Eineindeutigkeit*. On sait depuis que la synonymie joue un rôle essentiel dans la communication scientifique, mais seulement dans un cadre où une certaine normalisation est acceptée et intégrée. Lorsqu'une discipline ou une technologie est en voie de constitution, ou lorsqu'une langue est appelée à assumer le rôle de vecteur de connaissances scientifiques et techniques il est indispensable de procéder à une régulation, qui peut se faire de différentes façons. La normalisation a l'avantage d'être relativement démocratique : on essaie d'aboutir à un consensus concernant les dénominations et surtout les définitions.

Au dix-huitième siècle, la synonymie est une des grandes questions qui agitent les philosophes, que ce soit dans le cadre de la langue générale ou dans le langage des arts et des sciences. Différents linguistes ont étudié le phénomène : Jean-Claude Chevalier (1971), Agnès Steuckhard (2007) et Christophe Rey (2004) en particulier. Ils mettent tous en lumière le rôle de Beauzée, l'auteur d'articles sur la langue dans l'*Encyclopédie* et dans l'*Encyclopédie méthodique*, qui parle de l'*abus des mots*, idée qui remonte à Locke. Nos collègues s'intéressent en particulier au rôle de Beauzée dans la mise au point d'une terminologie linguistique, mais sa réflexion a une grande portée et s'applique, de manière différente, au langage des arts (ou techniques), où la prolifération des ateliers aboutissait à une prolifération des termes qui désignaient les mêmes référents… ou à peu près, mais c'est une autre question.

Diderot était bien au courant du problème, qu'il situait surtout dans le domaine des arts, plutôt que dans les sciences :

> *J'ai trouvé la langue des Arts très-imparfaite par deux causes ; la disette des mots propres, & l'abondance des synonymes. Il y a des outils qui ont plusieurs noms différents ; d'autres n'ont au contraire que le nom générique, engin, machine, sans aucune addition qui les spécifie ; quelquefois la moindre petite différence suffit aux Artistes pour abandonner le nom générique & inventer des noms particuliers ; d'autres fois, un outil singulier par sa forme & son usage, ou n'a point de nom, ou porte le nom d'un autre outil avec lequel il n'a rien de commun.*
> *Encyclopédie*, Article *art*

On voit dans ce bref extrait que certains reproches peuvent être adressés également aux scientifiques. Mais on est encore frappé par la parenté de réflexion entre les pionniers de l'encyclopédie et ceux de la terminologie, encore un parallèle qui mérite une recherche approfondie, tout comme le débat

qui a caractérisé le dix-huitième siècle sur la synonymie en langue générale, dans les sciences et dans les techniques.

4.6. Intervention sur la langue

Le sixième critère, l'intervention sur la langue, découle de la question de la synonymie : comment en effet régler la question de la synonymie sinon qu'agissant sur la langue ? Nous nous permettons une entorse à notre ordre de présentation en restant chez Diderot pour l'introduction plutôt que de citer un pionnier de la terminologie.

> *Il serait à souhaiter qu'on eût plus d'égard à l'analogie des formes & des usages. Les Géomètres n'ont pas autant de noms qu'ils ont de figures ; mais dans la langue des Arts, un marteau, une tenaille, une auge, une pelle, &c. ont presque autant de dénominations qu'il y a d'Arts. La langue change en grande partie d'une manufacture à une autre. [...] Il seroit donc à souhaiter qu'un bon Logicien à qui les Arts seroient familiers entreprît des éléments de la grammaire des Arts. »*

Ici il fait la différence entre les « scientifiques » (ici les géomètres), dont la terminologie sert de modèle, et le langage des praticiens, caractérisé par la profusion et l'anarchie, pour former le vœu qu'un logicien intervienne dans le but de mettre bon ordre dans cette anarchie.

Ceci est à mettre en parallèle avec le rôle que Wüster attribue à la normalisation, qu'il limite expressément à la technologie.

> *In Wirklichkeit aber ist die technische Sprachnormung nur* ***Terminologie****normung, d.h. nur Normung der Fachausdrücke. Sie unterscheidet sich nicht wesentlich von der Vereinbarung von Fachausdrücken auf anderen Wissensgebieten, sehr wesentlich aber von der " natürlichen' d.h. ungelenkten, Entwicklung der Umgangssprache (Gemeinsprache).*
>
> *La normalisation linguistique technique n'est autre chose que la normalisation terminologique, c'est-à-dire uniquement la normalisation des expressions techniques. Elle n'est pas foncièrement différente de l'harmonisation des expressions spécialisées dans d'autres domaines du savoir, mais fondamentalement différente en revanche du développement 'naturel' ou libre de la langue courante*

(générale).
Wüster 1955 dans Picht, Schmitz (dir.) 2001 : 261

Il la distingue nettement de l'évolution naturelle de la langue. Il est significatif que Wüster parle peu de la création de nouveaux termes – *néonymie* n'est pas seulement absent en tant que mot, le concept ne figure pas chez Wüster.

Bien entendu, la meilleure expression de cette intervention sur la langue est la mise au point de la nomenclature où une nouvelle terminologie est inventée pour véhiculer la nouvelle conception de la discipline[12]. Or, celle-ci n'est présentée que partiellement dans l'*Encyclopédie méthodique* : l'article *acide* est rédigé avant la publication (et sans doute la rédaction) de la *Méthode de nomenclature chimique* (1787).de Guyton de Morveau, Lavoisier, Berthollet, Fourcroy (1787). La terminologie de Lavoisier y figure à peine.

On a l'impression que cette réalisation représente désormais le modèle d'aménagement linguistique. D'autres, comme François Gaudin (2003 : 97), considèrent en revanche que cette créativité représente une exception « …. ces enclaves que constituent les nomenclatures, dont le statut dénominatif peut être dit dérogatoire ».

Il existe un courant de recherche d'orientation linguistique sur les nomenclatures scientifiques du passé – y compris avant notre période. Les travaux de Ph. Selosse (2007) sur l'émergence des nomenclatures botaniques sont exemplaires à ce titre.

4.7. Dimension internationale

Le dernier critère retenu pour ce rapide passage en revue est sans doute le moins essentiel – la dimension internationale. Pour Wüster, espérantiste convaincu, la mission de l'entente internationale en matière scientifique et technique était primordiale. Il en parle souvent dans ses écrits, à commencer par sa thèse, qui englobe toutes les langues employées à l'époque pour les sciences. Il n'hésite pas à saluer le travail pionnier de la CEI (Commission électrotechnique internationale)

> *Die internationale Sprachnormung will die nationalen Begriffsbestimmungen und nach Möglichkeit auch die nationalen*

12 En 1782, Guyton de Morveau, chargé de rédiger le *Dictionnaire de Chimie* dans l'*Encyclopédie méthodique*, entreprit de créer une nomenclature nouvelle en donnant aux divers corps des noms propres à rappeler leur composition et leurs propriétés essentielles. […] Jacquet-Pfau 2007.

> *äusseren und inneren Wortformen aneinander angleichen. Bahnbrechend auf diesem Gebiet war die 1906 gegründete Internationale Elektrotechnische Kommission (IEC). Sie hat seit ihrer Gründung an einem sechsprachigen Definitionswörterbuch der Elektronik gearbeitet.*
>
> *La normalisation internationale vise à harmoniser les définitions nationales des concepts et, si possible, à rapprocher les formes internes et externes des mots. Une des avancées dans ce domaine est la création en 1906 de la Commission électrotechnique internationale (CEI), qui depuis sa fondation travaille à l'élaboration d'un dictionnaire de l'électronique en six langues comportant des définitions.*
> Wüster 1948 dans Picht, Schmitz (dir.), 2001 : 6

On sait que l'*Encyclopédie* était à l'origine la traduction et l'adaptation de la *Cyclopaedia* d'Ephraim Chambers (1728), qui, lui, s'était inspiré du Trévoux, témoignage encore des ramifications internationales, favorisées par la fondation des premières revues scientifiques, dont le *Journal des Sçavans* (1665), les *Philosophical Transactions of the Royal Society* (1665), *Giornale de' letterati* en 1668, pour n'en mentionner que les principales.

Comme pour les autres critères, l'idée l'internationalisation en langue de spécialité mériterait d'être approfondie, que ce soit dans le cadre des études historiques ou contemporaines. Plusieurs équipes travaillent déjà à l'échelle internationale, souvent dans une perspective de l'histoire des sciences et des technologies, où historiens et linguistes collaborent activement. Pour l'époque contemporaine, il serait important de réaffirmer la dimension plurilingue des études des langues de spécialité, qui ont tendance à se réduire au seul anglais de spécialité.

5. Conclusion

Il est difficile de parler d'une conclusion lorsque les données que l'on vient de présenter ouvrent tant de pistes de recherche. Elles n'ont pas été laissées en friche, mais chacun les a exploitées à sa guise, en négligeant le travail interdisciplinaire, à quelques brillantes exceptions près, évoquées plus haut. Nous nous sommes posé la question de la transmission aux pionniers de la terminologie de l'acquis des XVII^{e} et XVIII^{e} siècles : il est difficile de trancher, mais il est clair qu'il s'agit d'un terreau qui a nourri les réflexions qui serviront encore plus d'un siècle plus tard. De même, certains faits restent gravés

dans les mémoires, comme l'invention de la nomenclature de la chimie. Il paraît clair toutefois que ces recherches historiques constituent un terrain fertile, qui ne demandent qu'à être cultivé, mais de préférences par des équipes interdisciplinaires : linguistes (métalexicographes, terminologues…) et spécialistes de l'histoire des idées (de la science et des technologies, de l'épistémologie). Certaines initiatives généralement internationales existent déjà : citons pour mémoire, pour la France, les chercheurs qui travaillent autour de Liliane Hilaire-Pérez de l'Université Paris-Diderot et du Centre A. Koyré[13], et pour l'Italie un réseau d'universitaires réunies autour de Maria Teresa Zanola.

Bibliographie

Bensaude-Vincent, Bernadette (1993), *Lavoisier : mémoires d'une révolution*. Paris, Flammarion.

Campenhoudt, Marc van (1997), « Evaluation des terminographies multilingues : Le dictionnaire nautique du capitaine Paasch face au dictionnaire aéronautique de l'ingénieur Schlomann », dans Hermans, Ad, *Les dictionnaires spécialisés et l'analyse de la valeur,* Louvain-la-Neuve, Peeters, 75-115. http://www.termisti.org/anval.pdf.

Chevalier, Jean-Claude (1971), « Note sur la notion de synonymie chez trois grammairiens des XVII^e^ et XVIII^e^ siècles », *Langue française*, 24, 40-47.

Daumas, Maurice (1951), « La chimie dans l'*Encyclopédie* et dans l'*Encyclopédie méthodique, Revue d'histoire des sciences et de leurs applications*, 4/3-4, 334-343.

De Gioia, Michele & Mario Marcon (2014), *Mots de la médiation. Un lexique bilingue français-italien/Parole di mediazione. Un lessico bilingue francese-italiano*. Préface de Michèle Guillaume-Hofnung, Padova University Press.

Delaveau, Pierre, Rodolphe, Thérèse (2006), Quelques réflexions sur la *Méthode de nomenclature chimique* de Guyton de Morveau, Lavoisier, Berthollet, Fourcroy (1787), *Revue d'histoire de la pharmacie* 94 / 350, 252-257.

13 Citons, comme exemple de travail pluridisciplinaire réalisé dans ce cadre l'article de Hoock, Jochen (2013), « Johann Beckmann et la place du discours technologique dans la pensée économique de la deuxième moitié du XVIII^e^ siècle », qui explore la dimension linguistique de l'histoire des idées économiques et techniques. https://techniqcak.hypotheses.org/files/2013/10/Jochen-Hoock-Johann-Beckmann-et-la-place-du-discours-technologique.pdf.

Depecker, Loïc & Roche, Christophe (2007), « Entre idée et concept : vers l'ontologie », *Langages* 168, 106-114.

Despres, Sylvie et Sylvie Szulman (2008), « Réseau terminologique *versus* Ontologie », TOTh 2008 https://hal.archives-ouvertes.fr/hal-00423525/document

Dubuc, Robert ([1978] 2002), *Manuel pratique de terminologie*, Montréal, Linguatech.

Fluck, Hans-Rüdiger ([1976] 1997), *Fachsprachen. Einführung und Bibliographie*, Tübingen, Francke Verlag.

Gambier, Yves (dir.) (1998), *Discours professionnels en français,* Nordeuropäische Beiträge aus den Human- und Gesellschaftswissenschaften Bd., 16, Frankfurt am Main, Peter Lang.

Gambier, Yves (1991), " Présupposés de la terminologie : vers une remise en cause ", *Cahiers de linguistique sociale*, 18, 31-58.

Gaudin, François (2003), *Socioterminologie* : *une approche sociolinguistique de la terminologie*. Bruxelles, De Boek Duculot.

Guyton de Morveau, Lavoisier, Berthollet, Fourcroy (1787), *Méthode de nomenclature chimique*. http://gallica.bnf.fr/ark :/12148/bpt6k1050402r.r=.langFR

Holzem, Maryvonne (1999), *Terminologie et documentation*. pour une meilleure circulation des savoirs. Paris, ADBS Éditions.

Humbley, John (2007) « Réflexions autour de la définition du champ de l'anglistique de spécialité : le cas de l'Europe du Nord », *Asp*, 53-70.

Jacquet-Pfau, Christine (2007), Lexicographie et terminologie au détour du XIX^{e} siècle : *La Grande Encyclopédie, Langages* 168, 24-38.

Kayser, Daniel (1995), Terme et définition. *La banque des mots.* Numéro spécial 7, Paris, CILF, 19-34.

Kocourek, Rostislav ([1982] 1991), *La langue française de la technique et de la science. Vers une linguistique de la langue savante.* 2^{e} édition. Wiesbaden, Oscar Brandstetter Verlag

Lerat, Pierre (1995), *Les langues spécialisées*. Paris, Presses universitaires de France.

L'Homme, Marie-Claude (2004), *La terminologie : principes et techniques*. Montréal, Les Presses universitaires de Montréal.

Picht, Heribert, & Draskau, Jennifer (1985), *Terminology : An Introduction*. Guildford, University of Surrey.

Picht, Heribert & Schmitz, Klaus-Dirk (dir.) (2001), *Terminologie und Wissensordnung. Ausgewählte Schriften aus dem Gesamtwerk von Eugen Wüster.* Vienne, TermNet.

Pruvost, Jean (1981), *Recherches sur les dictionnaires onomasiologiques,* Thèse de doctorat, directeur Bernard Quemada, Villetaneuse, Université Paris XIII.

Pruvost, Jean (2002), *Les dictionnaires de la langue française*, collection *Que sais-je ?* 3622, Paris, PUF.

Pruvost, Jean (2006), *Les dictionnaires français, outils d'une langue et d'une culture*, Paris, Éditions Ophrys. Collection : L'Essentiel français.

Rey, Christophe (2004), « Charles-Joseph Panckoucke : artisan de l'encyclopédisme français », article présenté lors d'une réunion de l'équipe METADIF. https://www.u-picardie.fr/LESCLaP/rey/Reyc_panckoucke.pdf

Roche, Christophe (2007), « Avant-propos », *Actes de la conférence TOTh 1*, Annecy, Institut Porphyre, v. http://www.porphyre.org/toth/files/actes/TOTh2007_actes.pdf

Roelcke, Thorsten ([1999] 2005), *Fachsprachen*, Berlin, Erich Schmidt Verlag.

Rondeau, Guy ([1981] 1984), *Introduction à la terminologie*, Chicoutimi, Gaëtan Morin éditeur.

Sager, Juan Carlos (1990), *A Practical Course in Terminology Processing*, Amsterdam/Philadelphie, John Benjamin.

Sager, Juan Carlos, Dungworth, David et McDonald, Peter (1980), *English Special Languages, Principles and Practices in Science and Technology,* Wiesbaden, Brandstetter.

Samain, Didier (2010), Eugen Wüster : de l'espéranto à la terminologie, dans Heinz, Michaela (dir.), *Cultures et lexicographies*, Berlin : Frank & Timme, 279-296.

Selosse, Philippe (2007), « La dénomination des savoirs en français préclassique (1500-1650) : bilan du colloque organisé à Lyon en juin 2005 », dans *Lexiques scientifiques et techniques. Constitution et approche historique*, dir. Olivier Bertrand, Hiltrud Gerner et Béatrice Stumpf, Palaiseau, Editions de l'École Polytechnique, p. 35-46.

Steuckardt, Agnès (2007), « Beauzée et la rationalisation de la terminologie grammaticale » dans *Congrès international des Lumières, Sciences, Techniques et Cultures*, Montpellier. https://hal.archives-ouvertes.fr/hal-00353728/document

Viel, Claude (2004), Le *Dictionnaire de chimie* de Pierre-Joseph Macquer, premier en date des dictionnaires de chimie. Importance et éditions successives, *Revue d'histoire de la pharmacie*, 92 / 342, 261-276.

Wright, Sue Ellen (1991), « Lexicography vs. Terminology,» *ASTM Standardization News,* Vol. 19, N° 12, 40-45.

Wüster, Eugen (1968), *Dictionnaire multilingue de la machine-outil*. Londres, Science Press.

Wüster, Eugen ([1979] 1985), *Einführung in die allgemeine Terminologielehre und terminologische Lexicographie*, Copenhague, HHK.

Zanola, Maria Teresa (2014), *Arts et métiers au XVIII^e siècle. Études de terminologie diachronique*, Préface d'Alain Rey, postface de Bénédicte Madinier, Paris, L'Harmattan.

Articles

Why read when you can watch ? Video articles and knowledge representation within the medical domain

Sara Carvalho* ***, Christophe Roche** *, Rute Costa*

* NOVA CLUNL, Centro de Linguística da Universidade NOVA de Lisboa, Avenida de Berna 26-C, 1069-061 Lisboa – Portugal
** Condillac Research Group – LISTIC, Université Savoie Mont-Blanc, Campus Scientifique, 73376 Le Bourget du Lac cedex – France
*** Centro de Línguas, Literaturas e Culturas – Departamento de Línguas e Culturas, Universidade de Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro – Portugal
sara.carvalho@ua.pt, christophe.roche@univ-savoie.fr, rute.costa@fcsh.unl.pt

Abstract. Non-verbal forms of knowledge representation have become increasingly important in Terminology, both from a theoretical and methodological perspective. Examples of non-verbal knowledge representations have been abundantly present in medicine, the subject field under analysis, for centuries. However, due to the current – and impressive – technological progress in healthcare, resources that combine text, high-resolution video, as well as 2D and 3D animation, are now an integral part of modern medicine and play a pivotal role, especially in medical procedures. In this respect, this paper will focus on one of the most recent and promising of such multimedia resources, namely the video article, a peer-reviewed and indexed article which comprises both verbal and non-verbal elements (e.g. narration by experts and videos of surgeries including external and internal footage) and, therefore, verbal and non-verbal forms of knowledge representation. Such features make video articles an interesting resource to consider in the light of Terminology's double dimension approach. This will be illustrated in the paper through the analysis of a video article focusing on a surgical procedure, i.e. a laparoscopy for gynecological purposes.

1. Introduction

Specialized communication necessarily involves knowledge transfer, which requires knowledge representation that, in turn, can be either verbal, non-verbal, or mixed in nature (cf. Galinski & Picht, 1997, p. 42)[1]. As regards medicine, the subject field under study in this paper, diversity is the key feature underlying knowledge representation. Medicine is, in fact, a rather unique domain in this respect. To quote William Osler, a Canadian physician, "the practice of medicine is an art, based on science" (1904). It is, therefore, a practice-oriented field, and one where plain text still plays a key role, but also where the non-verbal has always been present, as a way to complement human observation and support the prevention, diagnosis, and treatment of diseases.

On the other hand, the development of increasingly visual and multimedia-based products that integrate both verbal and non-verbal elements has emerged as a byproduct of the exponential technological progress that has characterized healthcare in recent decades. One of the most recent – and promising – examples of such multimedia-based resources is the medical video article, a peer-reviewed and indexed scientific article fully presented in video format. By comprising both verbal and non-verbal elements (e.g. narration by experts and videos of surgeries including external and internal footage) and, therefore, verbal and non-verbal forms of knowledge representation, it is argued that video articles may constitute a pertinent resource in terminology work.

Therefore, and within the scope of a PhD project[2] which aims at the creation of a multilingual terminological resource based around the concept of <Endometriosis>[3], a chronic, inflammatory gynecological disease that is yet relatively unknown, even among the expert community (cf. Dunselman *et al.*, 2014), this paper is organized as follows: firstly, the theoretical background of this work is outlined, namely a double-dimensional approach to Terminology, which focuses on the synergies between what is conceptual and what is lin-

1 A more thorough analysis on the role of the non-verbal in terminology and knowledge representation may be found, for instance, in Galinski & Picht (1997), Picht (1999, 2011), Madsen (2016), or Prieto-Velasco (2016).

2 The PhD thesis in question, entitled "A terminological approach to knowledge organization within the scope of endometriosis: the EndoTerm project", was successfully concluded in September 2018 (cf. Carvalho, 2018).

3 Throughout this paper, concepts will be capitalized and written between single chevrons, whereas terms will be presented in lower case and between double quotation marks (cf. Roche, 2012).

guistic while also addressing the impact of such a perspective within terminology work. Terminology-related ISO standards will also be addressed in this section, in particular on how they categorize verbal, non-verbal, and multimedia information. Secondly, the topic of knowledge representation in medicine will be explored, as well as the unique features of the medical video article as a knowledge representation resource and its added value in terminological projects dedicated to healthcare. Thirdly, and based on the example of a medical video article regarding a surgical procedure, i.e. a gynecologic laparoscopy, and on the information gathered from that analysis, a concept modelling proposal for <Laparoendoscopic single-site surgery> will be put forward, followed by the concluding remarks.

2. Terminology and knowledge representation

2.1. A double-dimensional approach to Terminology

This approach, which encompasses both a linguistic and a conceptual dimension that are interrelated, has been described by Roche (2012, 2015), Costa (2013), and Santos & Costa (2015). According to Roche (2015, p. 136), Terminology is both a "science of objects and a science of terms". For Santos & Costa (2015), it is precisely this double dimension, together with the study of how both dimensions relate to each other, that grants Terminology the status of an autonomous scientific subject.

This double dimension approach implies, therefore, that the experts' conceptualization(s) of a given subject field, on the one hand, and the discourses produced by them, on the other, must be taken into account. The conceptual dimension is regarded, in this research project, as the foundational element, since, and corroborating what is stated in the ISO 704 (2009), "producing a terminology requires an understanding of the conceptualization that underpins human knowledge in a subject area" (p. 3). However, the linguistic dimension cannot be underestimated. Despite the pitfalls that often characterize natural language, such as ambiguity, polysemy, or ellipses, even in specialized communication, resorting to a specialized corpus allows a valuable insight into a given community of practice and into the way the experts view their domain.

Hence, the cornerstone of this approach lies in the complementarity of these two fundamentally different dimensions. Furthermore, understanding the relationship between the two dimensions plays a key role in terminology work, as it will contribute to define a methodology that does not compromise

the main goal of a terminological project as it is understood in this paper, which is to represent, organize and share the knowledge pertaining a certain domain, based on the way it is conceptualized by a community of experts.

As such, experts are considered indispensable to terminology work, working collaboratively with the terminologist in different steps of the project, in order to identify the fundamental concepts of that subject field, the way they relate to each other, as well as how they are represented, both verbally and non-verbally (cf. Costa *et al*., 2012). As a way to illustrate the preceding ideas and to substantiate the case study that follows, the core Terminology-related ISO standards (ISO 1087-1, 2000 and ISO 704, 2009) were analyzed to see how these different forms of concept representation are depicted.

2.2. Forms of concept representation in Terminology-related ISO standards

As argued earlier, the notion of concept is regarded as key in any terminological project under the double-dimensional approach, and the ISO 1087-1 standard defines it as a "unit of knowledge created by a unique combination of characteristics" (2000, p. 2). And how are concepts represented within Terminology as a subject field and its related work? According, again, to ISO 1087-1 (2000), either via a designation ("representation of a concept by a sign which denotes it") or a definition ("representation of a concept by a descriptive statement which serves to differentiate it from related concepts") (*ibid*., p. 6).

As regards designations, this ISO standard clearly identifies terms and appellations as "verbal designations" of concepts, further specifying that the former refers to a general concept, whereas the latter deals with an individual concept. Despite considering symbol a "type of designation", ISO 1087-1 does not proceed to define it. One's assumption that a symbol may be, as a counterpart to term and appellation, a non-verbal type of designation is corroborated, although not through a clear-cut definition, by the ISO 704 (2009), which states that a designation can represent a concept either via linguistic or non-linguistic means (p. 34), adding that a symbol's "visual representation of concepts functions independently of any given language" (p. 41).

When it comes to definitions, the ISO 1087-1 places its focus, again, solely on natural language definitions, namely the intensional (describing "the intension of a concept by stating the superordinate concept and the delimiting characteristics") and the extensional, whereby all the subordinate concepts

are enumerated under one criterion of subdivision (2000, p. 6). While the ISO 704 highlights intensional definitions as "the most explicit and precise method of concept definition" (2009, p. 44), it acknowledges other types of definitions (extensional, ostensive, lexical, precising, and stipulative). Among these additional types, the ostensive definitions can play an important role in our discussion concerning forms of concept representation, since they depict "non-lexical representations of the concept (such as a drawing, an illustration, a video, a sound clip, a computer animation, etc.)" or may even "point to an object" (p. 45). The standard adds further information in this regard, indicating that multimedia can also be used to display non-lexical concept representations and, hence, ostensive definitions (*ibid*.)[4]. Fig. 1 below summarizes the information presented so far concerning both designations and definitions, according to both ISO 1087-1 and ISO 704 standards.

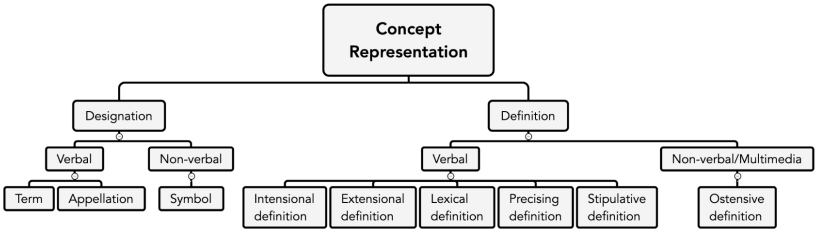


Figure 1 – Types of concept representation according to the ISO 1087-1 (2000) and ISO 704 (2009) standards.

Also in this regard, the ISO 704 (2009) standard outlines potential limitations of ostensive definitions, such as the fact that it may not always be clear what is being referred to or that it may be difficult to identify the superordinate concept, thereby recommending that these definitions "are best employed as complements to intensional definitions" (p. 45-46). We argue, however, that more technologically advanced and mixed forms of concept representation in certain subject fields – medicine being one of them – are here to stay and will become so prevalent that a discussion concerning the apparently clear-cut

4 In relation to the previous version of the ISO 704 standard, published in 2000, the 2009 version adds multimedia to the various forms of concept representation, in line with technological innovation, and also with Galinski & Picht's claim, back in 1997, of multimedia as a "dynamic and multidimensional" form of concept representation (p. 58).

notions of designation and definition will have to occur in upcoming revisions of Terminology-related standards. To further substantiate this claim, the following section will outline the role of non-verbal, but also mixed, forms of concept representation in medicine by describing a relatively new type of resource in this area: the video article.

3. The dynamics of knowledge representation in healthcare

3.1. Introduction

As referred to in the introductory section, knowledge can be represented in various ways in medicine. Fig. 2 illustrates several examples specifically related to the field of endometriosis: [1] a scientific article presenting a consensus statement about the disease; [2] the institutional website from a healthcare provider containing information about this pathology and targeted at a broader audience; [3] a monolingual glossary of endometriosis, available in the global online platform endometriosis.org[5]; [4] a diagram containing the representation of endometriosis according to SNOMED CT[6]; [5] an image depicting the different locations of endometriosis lesions; and [6] a screenshot of a video with 3D animations describing a hysterectomy, a common surgical procedure for patients with endometriosis.

5 http://endometriosis.org/glossary/

6 A comprehensive and multilingual clinical healthcare terminology that supports the representation of clinical content in Electronic Health Records (EHRs) in a consistent, comparable, reliable and computer-readable way (IHTSDO, 2018).

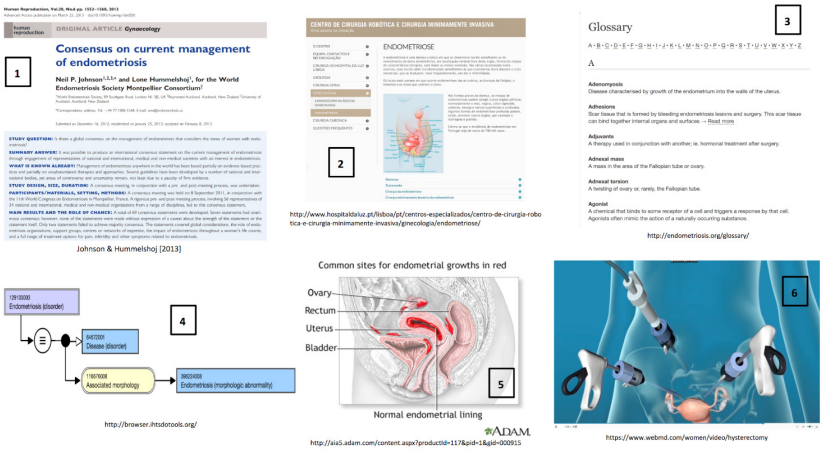


Figure 2 – Examples of knowledge representation within the scope of endometriosis.

It is clear, therefore, that on the one hand, verbal elements do play a critical role in the representation and dissemination of medical knowledge. Although specialized texts do not contain concepts *per se*, they provide, via the verbal designations of such concepts – the terms – and the systematic study of certain regularities[7], a key access point to the knowledge of a given domain. On the other hand, the relevance of a more visual approach to knowledge representation has been unequivocal in medicine: for centuries, medical illustrations have contributed significantly to the representation and dissemination of medical knowledge, and their quality, as well as degree of precision, have gone hand in hand with scientific and technological progress. An example of Henry Gray's *Anatomy of the Human Body* (1918)[8], namely an illustration of the female reproductive system, is provided in Fig. 3.

7 What Meyer (2001) describes as Knowledge-Rich Contexts and their respective Knowledge Patterns.

8 Available at: https://www.bartleby.com/107/

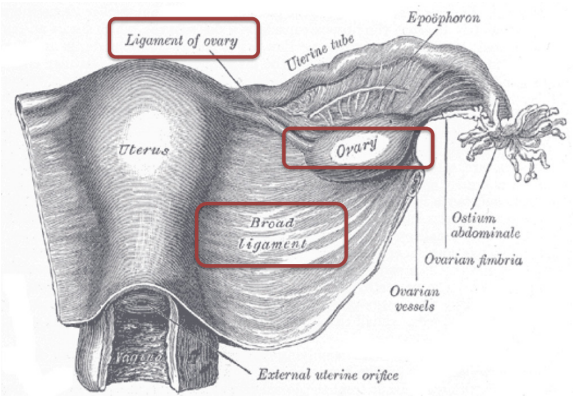


Figure 3 – Henry Gray's illustration of the female reproductive system (1918).

The illustration in question provides a rather common example in this subject field, by combining verbal and non-verbal elements in order to convey knowledge. Yet the existing terms (e.g. "ovary", "ligament of ovary", "broad ligament", etc.), albeit providing useful information, are not part of an actual text. The same happens with #5 in Fig. 2. So rather than being regarded as signs from a Saussurean perspective – in the sense that, when found in discourse, terms can give rise to the construction of meaning (a *signifié* in the Saussurean sense), i.e. they acquire value in discourse –, the terms that can be identified in Fig. 3 could be perceived as signs according to the philosopher William of Ockham, for whom a sign is "tout ce qui, étant appréhendé, fait connaître quelque chose d'autre" (1988, p. 7). In other words, they have the capacity to exist outside of discourse, pointing towards the concept and thus granting access into the specialized domain as well.

Bearing in mind the content of section 2.2 and the diagram in Fig. 1, it should also be noted that medicine is one of the subject fields where, in addition to verbal concept definitions (both intensive and extensive), non-verbal definitions can be rather prevalent as well. The image below (Fig. 4), which is part of a classification system created by the American Society for Reproductive Medicine (ASRM), represents not only the female reproductive system, but also the extent and location of endometriosis lesions and adhesions. Far from seeing a mere illustration, a subject field expert would immediately recognize it as a representation of Stage-IV (severe) endometriosis. The resource

in question could, therefore, be perceived as an example of a non-verbal, thus ostensive definition of the <Stage-IV endometriosis> concept.

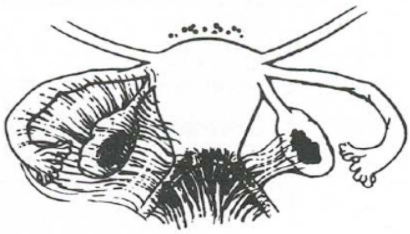


Figure 4 – Stage-IV endometriosis (ASRM, 1997).

In addition to the aforementioned illustrations and diagrams, and as stated earlier, the exponential technological progress that has characterized healthcare in recent decades has led to the development of more and more multimedia content. By combining different media types (e.g. sound, animation, text, graphics and video) and, thus, verbal and non-verbal elements, with the purpose of presenting information via a computer (cf. Bornman & Solms, 1993), multimedia is, in itself, a mixed form of knowledge representation. As will be shown in the following section, the video article is a relatively new type of such multimedia-based resources, used mainly in the context of scholarly communication.

3.2. The video article: a double-dimensional resource in medical knowledge representation

Falcone & Setúbal (2013) define the video article as a peer-reviewed scientific article containing "all the elements outlined in a structured abstract and full written manuscript but presented in video form" (p. 267). According to these authors, video articles constitute, therefore, a new paradigm in clinical research, since these videos, far from being perceived as mere "adjunct material", are now "the focal point of research", uniquely providing a "visual presentation of research" (*ibid.*).

Such video articles, which are also indexed, mainly in PubMed[9], have become increasingly popular within the biomedical field, especially among the surgeon community, which is understandable, given that most video arti-

9 Cf. https://pubmed.ncbi.nlm.nih.gov/

cles focus on surgical procedures or techniques. They can be found either in the so-called video journals[10], i.e. scientific journals which consist solely of this type of resource, or in 'traditional' journals, some of which have recently created media sections and appointed media editors, responsible for coordinating the reviews of the videos submitted for publication[11]. In the first category – that of video journals –, JOVE claims to be the first of its kind. Founded in 2006 by Moshe Pritsker, it currently publishes more than 100 new videos every month and, in 2018, it had an impact factor of 1.108.

But why the sudden and increasing interest in this type of article? According to several studies (e.g. Aarts *et al.* 2015; Baker, 2016), the underlying issues may be related to the poor reproducibility of research results: on the one hand, due to difficulties in replicating lab experiments or procedures by relying only on text-based documentation (there may be inconsistencies, biased results, etc.) and, on the other hand, most research units struggle to send researchers across the globe to receive specialized training. Video articles attempt, therefore, to solve both problems: the person watching the video is listening to and observing the actual expert in the comfort of their own research lab or operating room.

Bearing in mind the background provided in section 2, we argue that this new type of knowledge representation resource in the medical domain is double-dimensional as well since it combines, within a given field of specialized knowledge, verbal and non-verbal elements that interconnect, thereby enabling knowledge representation and, subsequently, knowledge transfer via a more comprehensive – in the sensorial perspective – dissemination strategy. In addition to the aforementioned context of scholarly communication, specialized videos such as these can be particularly effective in the teaching/learning of surgical procedures and techniques.

In short, both the conceptual and the linguistic dimensions interconnect in the video article, thus making it a pertinent resource to take into account in terminology work, especially within a double-dimensional approach. As a

10 E.g. the *Journal of Visualized Experiments* (JOVE) (https://www.jove.com/), the *Video Journal of Clinical Research* (http://www.videojournalofclinicalresearch.com/), and, more recently, the *Latest Thinking* (LT) platform (https://lt.org/).

11 Some examples of recent "Media Sections" related to gynecology and obstetrics can be found in *Fertility & Sterility* (https://www.elsevier.com/journals/fertility-and-sterility/0015-0282/guide-for-authors) and in the *Journal of Minimally Invasive Gynecology* (JMIG) (https://www.elsevier.com/journals/journal-of-minimally-invasive-gynecology/1553-4650/guide-for-authors).

new type of scholarly communication that seems to be here to stay, at least in the medical community, its inclusion in a specialized corpus in a medical terminology project may eventually become inevitable, which will, in turn, pose interesting theoretical and methodological challenges.

4. Case study : video article on single-port laparoscopy

4.1. Introduction

The case study presented here is based on a medical video article entitled "Single port laparoscopy"[12], which depicts a gynecological procedure – in this case, a hysterectomy, commonly seen as a last resort in cases of severe endometriosis – using a relatively recent type of surgery the authors call single-port laparoscopy.

After the initial acquaintance with the video, this resource was uploaded into CLAN (Child Language Analysis)[13], an open-source, cross-platform software originally developed to create and analyze transcripts in the Child Language Exchange System (CHILDES) database but that, since 2000, has broadened in scope, including corpora for aphasia, or traumatic brain injury. The purpose was to test this tool as a possible option for future work with this type of corpus, in particular CLAN's ability to link video, audio and transcript within the same work environment, making transcription activities easier by enabling the transcriber to clearly delimit the various utterances, to repeat those segments as often as necessary, or to increase/decrease the playback speed. The full transcript can also be exported into .txt if necessary, which facilitates the use of textual-corpus processing/analysis tools such as AntConc[14]. Fig. 5 displays a screenshot of one of those tests. In this specific case, the narration by the expert accompanies 3D animations, thereby illus-

12 Available at : https://www.fertstert.org/article/S0015-0282(12)00387-1/fulltext
13 http://dali.talkbank.org/clan/
14 A freeware corpus analysis toolkit used mostly for concordancing and text analysis (Anthony, 2018).

trating the interrelation between verbal and non-verbal elements referred to earlier.

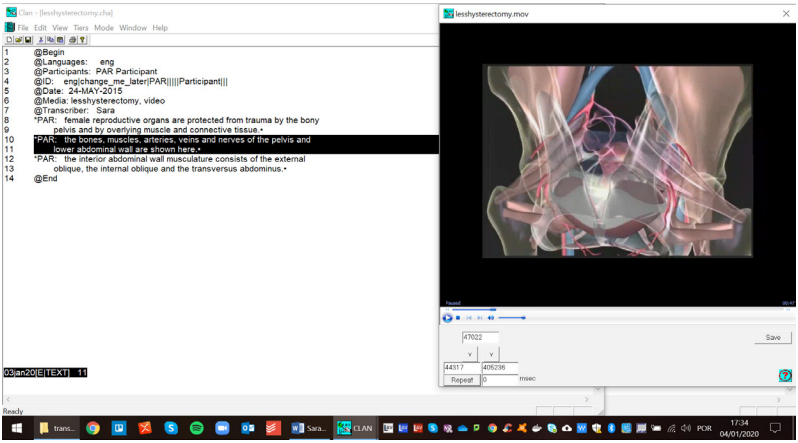


Fig. 5 – Screenshot of a partial transcription of the video using CLAN.

Throughout the video, the expert/narrator highlights the apparent lack of terminological consensus among the expert community in what concerns <Single-port laparoscopy>, listing the designations – and their respective acronyms/initialisms presented in Table 1.

Designations for single-port laparoscopy
SILS Single-incision laparoscopic surgery
SSL Single-site laparoscopy
SPA Single-port access surgery
OPUS One-port umbilical surgery
TUES Transluminal endoscopic surgery
NOTUS Natural orifice transumbilical surgery
SLAPP Single laparoscopic port procedure
LESS Laparo-endoscopic single-site surgery

Tab. 1 – *Designations for single-port laparoscopy identified in the video.*

The narrator then proceeds to state that for the purposes of that video, they would continue to refer to that procedure as "single-port laparoscopy". Interestingly, and while that consistency is maintained from an oral standpoint, the same does not happen with the written information that is shown in the video. A few minutes later, the terms "single port umbilical surgery" and "one-port umbilical surgery" appear in a segment of the video dealing with the advantages and disadvantages of this surgical technique. Through the further study of this issue, more than 20 acronyms used to designate this concept were identified in the literature and systematized in Carvalho *et al.* (2016). It was also found that a multidisciplinary medical consortium[15] gathered in 2008 to end that terminological dispersion in the area, having decided that the term "laparoendoscopic single-site surgery" (also known as LESS surgery) most accurately depicted the surgical procedure in question.

4.2. Concept modelling proposal

The following stage consisted of the creation of a concept modelling proposal for <Laparoendoscopic single-site surgery> based not only on the information provided by the video article and the preceding textual sources, but also on other videos about the same topic, as well as on the feedback from two senior expert gynecologists who are also surgeons. This proposal was developed using a software environment for concept system building called OTe Soft (OntoTerminology engine), designed by the Condillac Research Group[16], which has a clear concept orientation, although the user can also incorporate terms and, thus, integrate the linguistic dimension.

OTe Soft is structured around concepts, perceived as knowledge of a plurality of things that "help organize reality by grouping similar objects through what they have in common (Roche, 2015) (e.g. <Laparoscope>). One or more terms may be assigned to each concept, in various languages: e.g. "laparoscope" (EN); "laparoscópio" (PT); "laparoscope" (FR), etc. In addition, a concept may be qualified by attributes, which have a given value, and be

15 Called the Laparoendoscopic Single-Site Surgery Consortium for Assessment and Research (LESSCAR), that published a consensus statement with the main conclusions of that meeting (Gill *et al*., 2010).

16 http://new.condillac.org/

assigned one or more instances, also called "things", i.e. representations of elements in reality (cf. Fig. 8).

Concepts are linked to each other through concept relations : subsumption (is_a) (generic) and composition (part_of) (partitive) are presented by default. However, the tool allows the user to create new concept relations, as long as the logical principles are maintained (e.g. two concepts cannot be linked by the *instance of* relation)[17]. These relations are represented by different colors, in order to facilitate the graph's visual readability. The final "product" is called model, or semantic network, which can be exported in various formats (json, RDFS or OWL).

The following figures (6, 7 and 8) present examples of micro-concept maps built around the concept of <LESS surgery>. The first micro-map (Fig. 6) aims to position <LESS surgery> within the broader concept of <Surgery>.

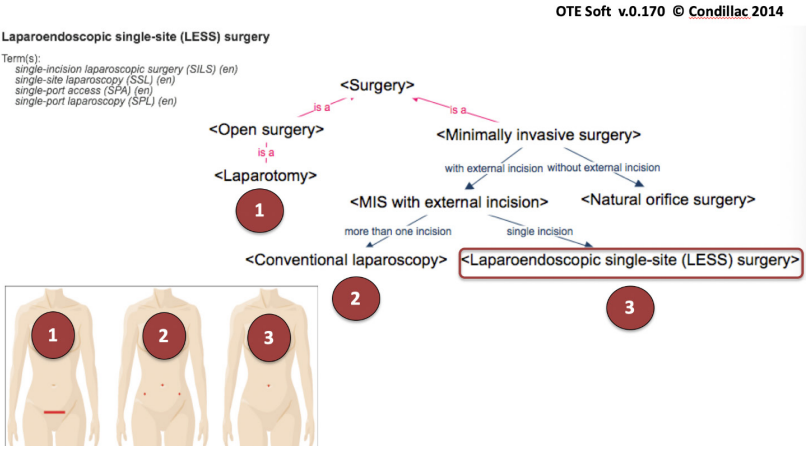


FIG. 6 – *The concept of <Laparoendoscopic single-site surgery>.*

There is a first subdivision presenting <Open surgery> and <Minimally invasive surgery> as subordinate concepts of <Surgery>. In the latter subtype, the subsequent hierarchy-based modeling was constructed through specific differentiation, bearing in mind the Aristotelian definition of *genus* +

17 One of the challenges of creating a concept-modeling proposal lies, in fact, in defining other types of concept relations that do not fall under the generic or partitive categories. The ISO standards (1087-1 :2000 and 704 : 2009) lack diversity and systematization, since they classify all the remaining relations as "non-hierarchical" (cf. Nuopponen 2011, 2014).

differentia[18]: i) with/without external incision ; ii) with one incision/with more than one incision. Besides other advantages, such as the operationalization potential provided by the exporting into computer-processable formats, this concept modeling strategy constitutes a valuable starting point for the terminologist in the construction of natural language definitions.

On the upper left side, the linguistic dimension is also visible, and it includes all the terms associated to the <LESS surgery> concept. In this case, it was decided to list some of the synonyms of the concept identified in the literature. Although the image does not show that, the user has the possibility of navigating through the concept network via concepts, terms, or relations. The three images in Fig. 6 were added afterwards, as the version of the OTe Soft tool we worked with did not allow the user to upload external resources (e.g. images, videos, diagrams, etc.)[19].

Fig. 7 explores the types of umbilical incisions that may occur in a LESS surgery, being that the single incision in the umbilicus (navel) is regarded by the expert community as the essential characteristic of the concept, i.e. the characteristic which makes the concept what it is and constitutes its essence (cf. ISO 1087-1 : 2000). In this figure, the metaphoric use attributed to the <Omega incision> should also be emphasized.

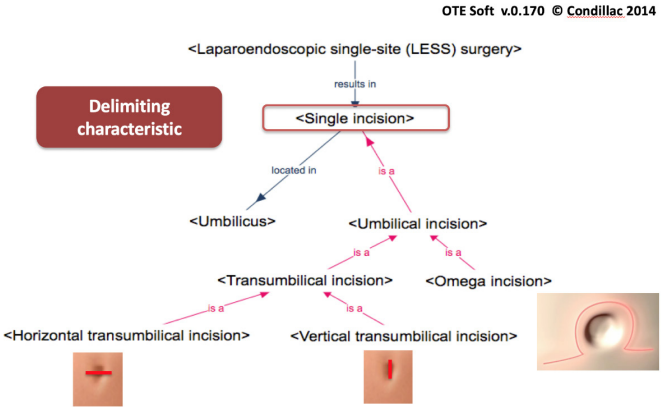


Fig. 7 – *Types of <Umbilical incision>.*

18 These, along with other Aristotelian categories, are explored in Porphyry's *Isagoge* (2003 edition).

19 This also applies to the images in the remaining figures.

Fig. 8 contains a more detailed insight on the types of laparoscopes that exist, one of which - <Flexible video laparoscope> - is currently being used to perform LESS surgeries. In this case, the *EndoEYE* is presented as an instance of this concept. There is also a basic distinction within the expert community between flexible and rigid laparoscopes, as depicted by the map.

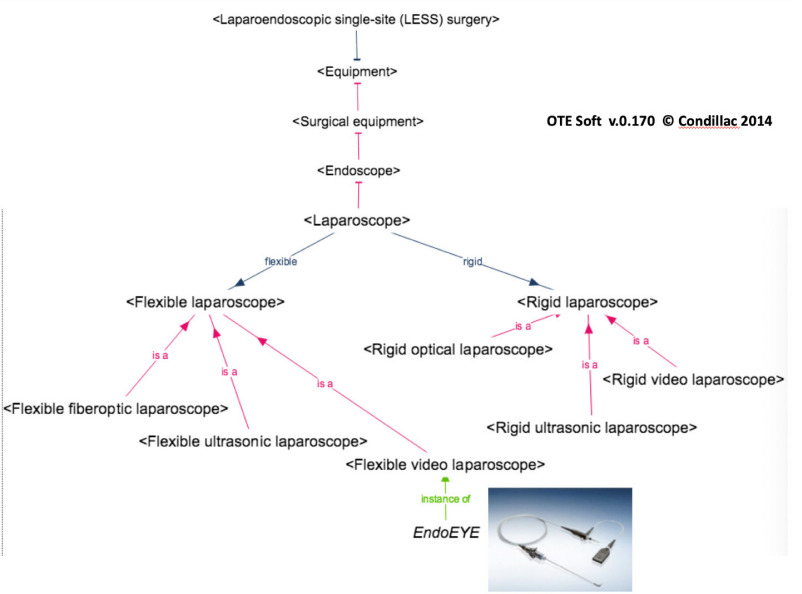


Fig. 8 – *Types of <Laparoscope>.*

5. Concluding remarks

To conclude, it is believed that the video article is a resource that terminologists working in the medical (or biomedical) domain should bear in mind, due to its double-dimensional nature and to its increasing importance in the subject field. The added value for Terminology is clear: firstly, it provides access to up-do-date knowledge, which is especially relevant in technologically advanced and cutting-edge areas such as surgery; secondly, there has been an exponential growth in video journal and e-learning surgery platforms

in life sciences, which will eventually lead to a considerable volume of data – and knowledge – that Terminology should not ignore.

The integration of such multimedia content into specialized corpora does, however, pose challenges to terminology work, for once concerning the methodology to be used, but also the tools for treating that corpus. In this particular case, the very access to the resources is also an issue, as it is, for the most part, extremely expensive/difficult due to data protection issues.

Nowadays, physicians who are training to become surgeons do so not only by reading books and scientific articles on their areas of specialty, but also by observing senior experts and participating in surgeries as assistants, or by watching video articles and live surgeries, many of which are broadcasted to the whole world. It is believed, in fact, that within the life sciences domain and perhaps in the next decade, some journals may cease to exist in its current, 'traditional' format, gradually being replaced by the video article format. This Brave New World of Medicine is challenging, but Terminology can – and will – still play a part. To quote Galinski & Picht (1997, p. 55), the increasing "power" of non-verbal – and mixed – forms of concept representation in the representation and dissemination of specialized knowledge may lead Terminology to rethink, in the very near future, their current role as 'mere' secondary, accessory material.

Acknowledgements

This research has been financed by Portuguese National Funding through the FCT – Fundação para a Ciência e Tecnologia as part of the project Centro de Linguística da Universidade Nova de Lisboa – UID/LIN/03213/2019.

References

Aarts, A. *et al.* (2015). Estimating the Reproducibility of Psychological Science. Science. 349. 10.1126/science.aac4716.

Anthony, L. (2018). AntConc (Version 3.5.7). Tokyo, Japan : Waseda University. Retrieved from http://www.laurenceanthony.net/software/antconc/

ASRM. (1997). Revised American Society for Reproductive Medicine classification of endometriosis. Fertility and Sterility, 67(5), 817-821.

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. Nature 533, p. 452-454 doi :10.1038/533452a

Bornman, H. & von Solms, S. (1993), "Hypermedia, multimedia and hypertext : definitions and overview", The Electronic Library, Vol. 11 No. 4/5, p. 259-268. https://doi.org/10.1108/eb045243

Carvalho, Luiz *et al.* (2012). Single port laparoscopy. Fertility and Sterility, Volume 97, Issue 5, e17

Carvalho, S., Costa, R., & Roche, C. (2016). LESS Can Indeed Be More : Linguistic and Conceptual Challenges in the Age of Interoperability. In B. Erdman Thomsen, Hanne ; Pareja-Lora, Antonio ; Nistrup Madsen (Ed.), Proceedings of the 12th International conference on Terminology and Knowledge Engineering (p. 157-167). Copenhagen : Copenhagen Business School. Department of International Business Communication and Politics.

Carvalho, S. (2018). A terminological approach to knowledge organization within the scope of endometriosis : the EndoTerm project. PhD thesis. Faculdade de Ciências Sociais e Humanas – Universidade NOVA de Lisboa/Communauté Université Grenoble Alpes. Available at : https://run.unl.pt/handle/10362/49745

Costa, R. (2013). Terminology and Specialised Lexicography : two complementary domains. Lexicographica, 29(1), p. 29-42.

Costa, R. *et al.* (2012). Mediation strategies between terminologists and experts. In Proceedings of GLAT 2012 - Terminologies : textes, discours et accès aux savoirs spécialisés, Genova, p. 297-308).

D'Ockham, G. (1988). Somme de Logique. (Traduit du latin par J. Biard). Trans-Europ-Repress.

Dunselman, G. *et al.* (2014). ESHRE guideline : management of women with endometriosis. Human Reproduction, 29(3), p. 400-412.

Falcone, T. & Setúbal, A. (2013). JMIG Video Articles. Journal of Minimally Invasive Gynecology, Volume 20, Issue 3, p. 267.

Galinski, C., & Picht, H. (1997). Graphic and Other Semiotic Forms of Knowledge Representation in Terminology Management. In S. E. Wright & G. Budin (Eds.), Handbook of Terminology Management Volume 1 : Basic Aspects of Terminology Management. Amsterdam : John Benjamins Publishing Company, p. 42-61.

Gill, I. *et al.* (2010). Consensus statement of the consortium for laparoendoscopic single-site surgery. Surgical Endoscopy, 24(4), p. 762-768.

IHTSDO. (2018). SNOMED CT Editorial Guide. International Health Terminology Standards Development Organisation. Retrieved from https://confluence.ihtsdotools.org/display/DOCEG/SNOMED+CT+Editorial+Guide

ISO (2000). ISO 1087-1 : Terminology Work - Vocabulary - Part 1 : Theory and Application. Geneva : International Organization for Standardization.

ISO (2009). ISO 704 : Terminology work - Principles and methods. Geneva : International Organization for Standardization.

Madsen, B. N. (2016). The Use of Linguistic and Non-linguistic Data in a Terminology and Knowledge Bank. In Proceedings of the TOTh Workshop 2013 : Verbal and nonverbal representation in terminology, p. 1-22.

Meyer, I. (2001). Extracting knowledge-rich contexts for terminography : a conceptual and methodological framework. In Didier Bourigault, Christian Jacquemin, & Marie-Claude L'Homme (Eds.), Recent Advances in Computational Terminology. Amsterdam/New York : John Benjamins Publishing Company, p. 279-302. https://doi.org/10.1075/nlp.2.15mey

Nuopponen, A. (2011). Methods of concept analysis – tools for systematic concept analysis. LSP Journal, 2(1), p. 4-15.

Nuopponen, A. (2014). Tangled Web of Concept Relations. Concept relations for ISO 1087-1 and ISO 704. In Proceedings of Terminology and Knowledge Engineering 2014. Berlin.

Osler W. (1904). Aequanimitas with other addresses to medical students, nurses and practitioners of medicine. Blakiston, Philadelphia.

Picht, H. (1999). Einige Überlegungen zur nicht-sprachlichen Repräsentation von Gegenständen und Begriffen. Synaps, 3, p. 1-50.

Picht, H. (2011). Non-verbal graphic representation of concepts. In Ivanovo School of Lexicography : Traditions and Innovations. A Festschrift in Honour of Professor Olga Karpova. Ivanovo : Ivanovo State University, p. 220-236.

Porphyry. (2003). Isagoge. (Translated by J. Barnes). Oxford : Oxford University Press.

Prieto-Velasco, J. A. (2016). Depicting specialized concepts : strategies for the visualization of terminological knowledge. In Proceedings of the TOTh Workshop 2013 : Verbal and nonverbal representation in terminology, p. 37-50.

Roche, C. (2012). Ontoterminology : How to unify terminology and ontology into a single paradigm. In Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012), Istanbul, Turkey, May 23-25, 2012. Istanbul : European Language Resources Association (ELRA), p. 2626-2630.

Roche, C. (2015). Ontological definition. In H. J. Kockaert & F. Steurs (Eds.), Handbook of Terminology - vol. 1, Amsterdam : John Benjamins Publishing Company, p. 128-152.

Santos, C., & Costa, R. (2015). Domain specificity : Semasiological and onomasiological knowledge representation. In H. J. Kockaert & F. Steurs (Eds.), Handbook of Terminology - vol. 1. Amsterdam : John Benjamins Publishing Company, p. 153-179.

Résumé

Les formes non verbales de représentation des connaissances sont devenues de plus en plus importantes en Terminologie, tant du point de vue théorique que méthodologique. Elles sont abondamment utilisées dans de nombreux domaines et en particulier en médecine et ce dès le début de cette discipline. Aujourd'hui, grâce aux progrès technologiques impressionnants réalisés dans le domaine de la santé, les ressources multimodales qui combinent textes, vidéos de haute résolution, ainsi que les animations 2D et 3D, font désormais partie de la médecine moderne où elles jouent un rôle central, en particulier dans les procédures médicales. Dans le cadre de cet article, nous nous sommes intéressés à l'une des ressources multimédias les plus récentes et les plus prometteuses, à savoir l'*article vidéo*. Un *article vidéo* est un article révisé et indexé qui comprend des éléments verbaux et non verbaux (par exemples, des narrations par des experts ou des vidéos d'opérations chirurgicales comprenant des séquences externes et internes) et, par conséquent, des formes verbales et non verbales de représentation des connaissances. Ces caractéristiques font des *articles vidéos* des ressources intéressantes à considérer dans une approche de la double dimension, linguistique et conceptuelle, de la Terminologie. L'analyse d'un *article vidéo* portant sur une intervention chirurgicale, la *laparoscopie* à des fins gynécologiques, permettra d'illustrer nos propositions.

PLOTITERM : modélisation de la terminologie de Plotin en OWL

Piccini, Silvia*

* Istituto di Linguistica Computazionale Antonio Zampolli – CNR
Via Moruzzi 1, 56124 Pisa
silvia.piccini@ilc.cnr.it

Résumé. Le présent article est dédié à la description de *Plotiterm*, une ressource termino-ontologique consacrée à la modélisation de la terminologie des *Ennéades* de Plotin et de leur traduction en langue arabe : la pseudo-*Théologie d'Aristote*.
Cette ressource a été développée dans Protégé, un éditeur d'ontologie open source supportant le langage standard OWL.
Dans *Plotiterm* le plan conceptuel et le plan terminologique sont séparés mais interconnectés, conformément aux nouveaux paradigmes et méthodologies développés durant les dernières années. Outre la composante ontologique, un ensemble de propriétés (*DataType Properties*) et de relations (*Object Properties*), subdivisées en relations lexicales et relations conceptuelles, permettent de représenter une large typologie d'informations caractérisant le contenu conceptuel et lexical des termes. Les données contenues dans *Plotiterm* peuvent être interrogées ou visualisées de façon graphique grâce aux outils offerts par Protégé ou à travers le module *Lexicon* inséré dans la plateforme de philologie numérique *G2A Web Application*.

1. Introduction

La ressource terminologique *Plotiterm* que je me propose d'illustrer dans cet article a été réalisée dans le cadre d'un projet Ideas, Advanced Grant 249431, intitulé *Greek into Arabic. Philosophical Concepts and Linguistic Bridges* et financé par le European Research Council (ERC).

Ce projet, dédié à la transmission du savoir grec dans le monde arabe, met l'accent sur l'un des textes les plus importants de la philosophie arabo-islamique des origines, la « *Théologie d'Aristote* ». Libre traduction en langue

arabe d'extraits des *Ennéades* de Plotin, cette œuvre joue un rôle fondamental dans le rencontre entre le monde musulman et la pensée philosophique grecque, car elle démontre aux élites du premier califat abbasside que la philosophie grecque était parfaitement en phase avec la doctrine islamique du تَوْحِيد , « tawhīd », c'est-à-dire de l'unicité et l'unité de Dieu.

L'objectif premier du projet était la création d'une édition critique de ce texte, édition assortie d'une traduction, de commentaires et d'index bilingues gréco-arabe et arabo-grec.

Le second objectif était représenté par la continuation du projet *Glossarium Graeco-Arabicum*, né à la Ruhr-Universität Bochum, conduit aujourd'hui par Gerhard Endress, Ruediger Arnzen et Yury Arzhanov, et proposant une base de données en ligne des termes grecs et arabes correspondants[1]. Ce glossaire peut être considéré idéalement comme la continuation du lexique papier *Greek and Arabic Lexicon* (*GALex* 1992-), édité par Gerhard Endress et Dimitri Gutas : en effet, il comprend les racines arabes de la lettre ج jusqu'à la fin de l'alphabet et contient environ 80 000 cartes scannées des termes des deux langues qui n'ont pas été publiées dans le *GALex*.

Le troisième objectif était la réalisation, par l'Istituto di Linguistica Computazionale *Antonio Zampolli* du CNR, d'une plateforme de philologie numérique permettant la visualisation, l'étude philologique et critique ainsi que l'annotation des *Ennéades* et de leur traduction arabe, et plus généralement, de textes anciens et de leurs traductions.

Au sein de cette plateforme, un module est dédié à la manipulation de *Plotiterm*, la ressource termino-ontologique consacrée à l'œuvre de Plotin et à sa traduction arabe et qui fait l'objet du présent article. Cette ressource a été développée par l'auteur de cet article en collaboration avec et sous la direction de Nilda Ruimy.

Plotiterm a été conçu afin de permettre l'étude du contenu conceptuel des *Ennéades* et pourrait constituer, selon l'auteur, une aide précieuse, tant pour l'étudiant que pour le chercheur, pour comprendre comment l'œuvre du philosophe néoplatonicien, destinée à influencer de manière déterminante la philosophie de langue arabe, a été perçue et adaptée aux catégories conceptuelles propres de la culture islamique.

1 Le glossaire est consultable à l'URL : http://telota.bbaw.de/glossga/.

Cependant, avant d'entrer dans le vif du sujet, il convient d'évoquer brièvement la figure de Plotin et les influences de son œuvre dans le monde arabe et de clarifier la notion de ressource termino-ontologique.

2. Plotin, le néoplatonisme et le monde arabe

Né à Lycopolis (Haute-Égypte) en 205 après J.C. et mort en 270 près de Minturnes, en Campanie, Plotin est considéré comme le créateur de la pensée néoplatonicienne. Sa pensée philosophique et le langage dans lequel elle s'exprime représentent en quelque sorte une synthèse admirable et originale de la pensée antique : le lexique typiquement platonicien, le langage technique de l'aristotélisme et les images de la mythologie grecque sont mis au service d'une vision nouvelle, suscitée par la spiritualité de son époque, tout en étant en compétition avec celle-ci. Les écrits de Plotin, élaborés de 254 à 270 au terme d'une longue période d'enseignement exclusivement oral, présentent une élaboration du platonisme dans sa pleine maturité. Les traités qu'il composa furent publiés par son disciple Porphyre de Tyr, qui se chargea de leur édition en y apportant toutefois d'importants changements, en ce sens qu'il donna un ordre systématique à des écrits issus de la discussion au sein de l'école. Comme il l'écrit lui-même dans sa *Vie de Plotin*, les traités ne furent pas publiés dans leur ordre chronologique mais selon un ordre systématique conceptuel, subdivisés en six groupes de neuf, d'où le titre d'*Ennéades*. Cette classification, inspirée de la mystique pythagoricienne des nombres, répondait en outre à des exigences d'ordre pédagogique : l'on décrivait ainsi une sorte de chemin initiatique que le philosophe devait suivre pour se hausser du monde sensible de la *physis*, décrit dans le premier tome, à travers les deux hypostases de l'Âme et de l'Intelligence (deuxième tome) jusqu'à la première hypostase, l'Un.

L'œuvre du philosophe présente une grande complexité liée essentiellement à trois aspects. D'une part, les pensées qui devaient être déjà parfaitement organisées dans l'esprit du maître sont confiées à une écriture instinctive[2]. Le style est riche en images et en métaphores, comme celui de son

2 Dans sa *Vie de Plotin*, Porphyre écrit à propos du style de son maître : « Ἐν δέ τῷ γράφειν σύντομος γέγονε καί πολύνους βραχύς τε καί νοήμασι πλεονάζων ἢ λέξεσι, τά πολλά ἐνθουσιῶν καί ἐκπαθῶς φράζων † καί τό συμπαθείας ἤ παραδόσεως. Ἐμμέμικται δ'ἐν τοῖς συγράμμασι καί τά Στοικά λανθάνοντα δόγματα καί τά Περιπατητικά· καταπεπύκνωται δέ καί ἡ "Μετά τά φυσικά" τοῦ Ἀριστοτέλους πραγματέια » : « Quand il écrivait, il était concis et riche d'idées, bref et plus abondant en idées qu'en mots ; la plupart du temps, il était transporté par l'inspiration et exposait avec passion † Sont mêlées dans ses ouvrages,

maître Platon. D'autre part, il n'y a pas d'exposition systématique de sa pensée. Les réflexions sur les grands thèmes partent de « problèmes du moment » (ἐκ προσκαίρων προβλημάτων), comme les définit Porphyre (*Vie de Plotin*, 5.60), posés par ses élèves. De plus, le vocabulaire de Plotin n'est pas simple à appréhender, d'autant que le philosophe a souvent coutume d'utiliser des termes sans en fournir de définition et de les employer, en outre, dans différentes acceptions.

La pensée du philosophe néoplatonicien a joué un rôle crucial dans la naissance de la philosophie arabo-islamique, en favorisant le syncrétisme entre néo-platonisme et aristotélisme qui la caractérise.

La version arabe des écrits de Plotin est connue sous le nom de pseudo-*Théologie d'Aristote*, car elle fut d'abord faussement attribuée au philosophe stagirite.

Traduite à Bagdad aux environs de l'année 840, comme on peut déduire du Prologue, la pseudo-*Théologie* se situe dans une période historique particulière pour le monde arabe, caractérisée par une grande effervescence culturelle : la conquête arabe s'étendait à la Syrie, à l'Égypte et à la Perse et entraînait une grande opération de traduction des textes grecs en langue arabe et syriaque. Ce grand mouvement de traduction devint plus intense et systématique et atteignit son *akme* sous la dynastie abbasside : au IX^{e} siècle sont présents dans les bibliothèques de Bagdad des manuscrits de Platon, d'Aristote, d'Euclide, de Ptolémée, de Galien et d'autres auteurs grecs de Byzance, d'Alexandrie et d'autres villes de langue grecque.

Parmi ces textes, se trouve également la pseudo-*Théologie d'Aristote*, traduction, ou plutôt libre paraphrase, des traités IV-V-VI des *Ennéades* de Plotin.

Parler de traduction est toujours difficile : depuis la linguistique et la sémiotique structuraliste, nous savons que la traduction conçue comme la reproduction du système de valeurs des signes qui composent l'original est chose impossible. Et c'est bien parce qu'il n'existe jamais de correspondance absolue entre les langues qu'il ne peut y avoir de « fully exact translations » (Nida 1964 : 156). Parler d'une simple traduction est d'autant plus impropre dans le cas de la pseudo-*Théologie*. En fait, la caractéristique macroscopique qui saute immédiatement à l'œil d'un lecteur même inexpérimenté, est la mise en page du texte, très différente de celle de l'original. Celle-ci fait immédia-

sans que l'on s'en aperçoive, les doctrines stoïciennes aussi bien que péripatéticiennes ; s'y trouve fréquemment employé aussi le traité d'Aristote *Métaphysique* ».

tement présager qu'il ne s'agit pas d'une simple traduction servile de Plotin, mais, pourrait-on dire, d'une nouvelle œuvre. Fréquentes et reconnaissables sont les lacunes, les interpolations de l'auteur arabe, les digressions qui n'ont aucun fondement dans le texte grec où l'auteur cherche à adapter la pensée de Plotin à ses doctrines théologiques.

Les nombreuses différences qui existent entre le texte plotinien et le texte arabe ont même conduit par le passé certains spécialistes à supposer que la pseudo-*Théologie* ne reposait pas sur les *Ennéades*, mais sur un compte rendu des leçons tenues par Plotin à Rome dans les dix ans durant lesquelles il n'a rien écrit[3].

L'œuvre de Plotin et sa version arabe ont attiré l'attention de nombreux spécialistes. En particulier la 'terminologie' de Plotin a fait l'objet de plusieurs études parmi lesquelles la plus réputée est le *Lexicon Plotinianum* de John H. Sleeman et Gilbert Pollet, ouvrage papier édité en 1980. Il existe également un volume des concordances lemmatisées de l'œuvre complète de Plotin, comprenant un index par lemme et par forme et dont la version sur CD-ROM inclut un logiciel de recherche avancée (Radice et Bombacigno 2004).

En ce qui concerne la terminologie de la pseudo-*Théologie* on pourrait énumérer nombre de travaux ponctuels ainsi que le *Greek and Arabic Lexicon* (*GALex*), édité par Gerhard Endress et Dimitri Gutas, et mentionné dans l'Introduction.

Cependant, il s'agit toujours d'ouvrages papier[4], c'est pourquoi nous avons souhaité mettre à profit les potentialités offertes par les nouvelles technologies de l'information, afin de développer pour la première fois une ressource termino-ontologique consacrée à l'œuvre plotinienne et à sa traduction arabe.

La structuration et la description formelle d'un ensemble de concepts philosophiques dénotés par les *Ennéades* de Plotin et par la pseudo-*Théologie d'Aristote* ont constitué un véritable défi. À la complexité inhérente de l'œuvre plotinienne s'ajoute, en fait, celle de sa traduction, ou plutôt, de son interprétation en langue arabe.

Dans *Plotiterm* deux âmes coexistent, c'est-à-dire deux langues différentes, évoquant un monde conceptuel qui tend à être le même mais qui souffre parfois de la pression de la langue dans laquelle il est exprimé.

3 Pour une discussion détaillée de cette question, cf. d'Ancona (2005).

4 Exception faite du *Glossarium Graeco-Arabicum* dont la continuation faisait partie des objectifs du projet.

3. Une ressource termino-ontologique : distinction entre le plan conceptuel et le plan terminologique

Pour pouvoir procéder à l'étude d'un texte et effectuer des recherches basées sur leur contenu sémantique, ou en extraire des informations, il est nécessaire que les concepts qui y sont évoqués ainsi que les termes dénotant ces concepts aient une organisation structurée, soient représentés de manière explicite et univoque par des métadonnées sémantiques, et soient définis à travers les liens qui les unissent.

C'est-à-dire que la recherche et l'extraction d'information ont besoin de ressources termino-ontologiques (RTO), constituées par deux composantes séparées mais profondément liées, une composante conceptuelle, l'ontologie et une composante lexicale, la terminologie. Dans une RTO les concepts et leur manifestation linguistique sont clairement distincts, tout en étant profondément interconnectés. Le plan terminologique se focalise sur les termes, c'est-à-dire sur des « lexical items belonging to a specialized area of usage in one or more languages » (Sager 1990 : 2).

Le terme ne se distingue du signe linguistique que par son domaine d'usage : alors que, concernant le signe linguistique, la relation biunivoque entre signifiant et signifié est fondée sur le partage par une communauté de locuteurs, dans le cas d'un terme, elle est fondée sur le partage par un groupe plus ou moins restreint d'experts.

Le contenu d'un terme, son signifié, ne doit pas être confondu avec la notion de concept. Ce dernier, en effet, en tant que représentation mentale d'un objet appartenant au monde réel, est totalement indépendant du terme, il a une existence propre en dehors du terme[5]. Cependant, comme le souligne Prandi (2011 : 6) « l'étude des signifiés est inséparable de l'étude des concepts exactement comme l'étude des concepts est inséparable de l'étude de leur mise en forme linguistique ».

La terminologie, en fait, ne peut ni ne doit oublier sa propre composante conceptuelle.

5 L'Organisation Internationale de Normalisation (ISO) a défini la notion de concept dans deux normes différentes : selon la norme ISO 704 : 1987 (2009) *Travail terminologique - Principes et méthodes*, un concept est défini comme une « construction mentale qui sert à classer les objets individuels du monde extérieur ou intérieur grâce à un processus d'abstraction plus ou moins arbitraire » ; tandis que dans la norme ISO 1087-1 : 2000 *Travaux terminologiques - Vocabulaire - Partie 1 : Théorie et application*, un concept est défini comme « une unité de connaissance créée par une combinaison unique de caractères ».

Je ne m'attarderai pas ici sur les rapports complexes entre signifié et concept, problème qui, bien que fondamental, n'entre pas dans le propos de cet article. La question importante à souligner est l'exigence théorique et méthodologique d'une ressource qui prévoit parallèlement à la dimension linguistique, terminologique, une dimension conceptuelle.

Comme nous le verrons, ceci est d'autant plus important au moment où deux systèmes linguistiques — en l'occurrence le grec et l'arabe — sont confrontés à un système conceptuel tendant à rester constant. Ceci permet en fait d'étudier la pression de la langue sur la structure des catégories extralinguistiques de la pensée, le rapport entre la langue, toujours liée à une certaine communauté de parlants et le système de concepts sous-jacents.

3.1. Bref survol de l'état de l'art

La nécessité théorique et méthodologique de distinguer le plan conceptuel du plan terminologique, bien qu'ils soient intimement liés, a donné lieu, durant les dernières années, au développement de nouveaux paradigmes et méthodologies. Citons parmi eux l''ontoterminologie', développée par Christophe Roche et par le groupe de Condillac (2007 ; 2012), basée sur une approche onomasiologique d'inspiration wüsterienne, et qui distingue et relie le système linguistique (la terminologie) et le système conceptuel (l'ontologie). Ou encore la 'termontographie', qui combine « les théories et les méthodes de l'analyse terminologique plurilingue de la théorie sociocognitive […] et les méthodes et directives de l'ingénierie ontologique » (Temmermann 2005 : 430). Cette méthodologie, élaborée par le Centre de recherches en langues spécialisées et de communication, préconise la prise en compte des deux approches : sémasiologique et onomasiologique.

Nombreuses ont été les tentatives de modélisation de ressources qui reflètent ces acquis théoriques et qui sont donc en mesure de représenter au sein d'un même modèle tant l'ontologie que le lexique ou la terminologie.

Des plateformes ont été développées comme *Terminae*, implémentée selon la méthode éponyme décrite dans Aussenac-Gilles *et al.* (2000), visant à la construction de ressources termino-ontologiques à partir de textes. Dans ce type de ressource, un niveau terminologique assure le lien entre le plan conceptuel et les textes. Dans l'exportation en OWL des ontologies construites dans *Terminae*, la structure de 'propriété d'annotation' (owl :AnnotationProperty) (cf. § 7.2.) a été utilisée pour doter le concept de propriétés (comme, par exemple, la synonymie) et d'occurrences relatives au

terme associé à ce concept (Szulman et Biébow 2004)[6]. Certains méta-modèles tentent, par contre, de mettre les deux plans au même niveau et de garantir au plan terminologique un statut indépendant.

Ainsi, Reymonet *et al*. (2007) proposent un modèle de représentation en OWL du niveau terminologique d'une RTO et de son lien avec le niveau ontologique. Les termes sont réifiés sous forme de classe OWL ; puis, au plus haut niveau de l'ontologie, la hiérarchie conceptuelle et la hiérarchie terminologique sont représentées respectivement sous les classes 'concept' et 'terme'. Le lien entre les termes et les concepts est assuré par la propriété d'objet de OWL 'dénote', orientée du terme vers le concept[7].

Contrairement à la bipartition terme-concept proposée par Reymonet, Desprès et Szulman (2008) décomposent le processus de construction d'une ontologie en trois étapes. Dans l'étape linguistique, un réseau terminologique est construit à partir des termes d'un texte associés à leurs relations lexicales. L'étape dite 'termino-ontologique' consiste en la construction du modèle conceptuel à partir du réseau terminologique ; durant cette phase, chaque acception d'un terme polysémique donne lieu à un concept terminologique distinct et les relations lexicales sont traduites en relations termino-ontologiques reliant les concepts. Enfin, l'étape ontologique est celle durant laquelle le modèle conceptuel est traduit en langage formel.

Insistant sur la nécessité d'associer de l'information linguistique aux ontologies, Buitelaar *et al.* (2009) proposent le modèle *LexInfo*, basé sur les modèles *LingInfo* et *LexOnto* ainsi que sur le standard ISO *LMF*.

Le modèle *Lemon*[8], récemment élaboré dans le cadre du projet européen Monnet[9], propose la tripartition *Lexical Entry*, *Lexical Sense* et *Domain Ontology*, tripartition qui évoque le modèle triadique classique caractéristique de certaines théories sémiotiques : 'le mot, l'idée, la chose'.

Bien qu'au stade actuel plusieurs travaux soulignent l'importance de la séparation entre l'ordre terminologique et l'ordre conceptuel, des environnements permettant de construire des RTO et donc en mesure de représenter

6 Concernant les points faibles d'une telle modélisation, cf. Reymonet (2007).

7 Reymonet et al. (2007) ont développé comme plugin de Protégé un éditeur de classes terminologiques appelé *TextViz*, caractérisé par une bonne expressivité mais non accessible au public.

8 http://lemon-model.net/.

9 Multilingual Ontologies for Networked Knowledge (2010-2013).

au sein d'un même modèle tant l'ontologie que la terminologie, font encore défaut.

C'est pourquoi, afin de développer notre base de connaissance terminologique, nous avons dû utiliser Protégé[10], un éditeur d'ontologie open source, supportant le langage déclaratif de modélisation OWL[11] et disponible en version Web. Nous avons cependant parfaitement conscience que 'Protégé-OWL' a été conçu pour la modélisation et la gestion d'ontologies mais n'est pas particulièrement adapté à représenter un niveau lexical ou terminologique.

4. Une ressource termino-ontologique de Plotin ?

Proposer une structuration et une description formelle d'un ensemble de concepts philosophiques et plus particulièrement de ceux dénotés par le vocabulaire des *Ennéades*, œuvre dont nombre de spécialistes s'accordent à reconnaître la complexité, pourrait être perçu comme une fâcheuse tentative d'enserrer une fois de plus une pensée qui semble vouloir échapper à toute systématisation rigide. Déjà emprisonnée dans le langage naturel parfois inapte à exprimer nos idées, la pensée de Plotin serait de nouveau figée dans une représentation formelle ? Il est bien certain qu'à ce jour tout au moins, aucun formalisme, aussi riche en pouvoir expressif soit-il, ne parvient à saisir et interpréter toute la complexité, l'ambiguïté inhérente à la langue et les subtilités du langage naturel. Un tel dessein pourrait donc laisser craindre une simplification et un appauvrissement des connaissances véhiculées par l'œuvre de Plotin.

À ceci il faut ajouter la composante dédiée au lexique arabe et toutes les difficultés inhérentes aux cas de traductions non littérales et d'interprétations du texte.

Nous avons néanmoins souhaité relever ce défi et entreprendre une démarche expérimentale dans ce domaine, dans l'intime conviction, étayée par les résultats de travaux précédents, qu'une telle représentation de la connaissance peut, bien au contraire, se révéler source de grande richesse.

En développant cette ressource termino-ontologique, notre objectif était de proposer un outil de travail destiné tant aux spécialistes qu'aux étudiants

10 http://protege.stanford.edu/.

11 Ainsi que la version OWL2, qui est une extension et une révision de la version originale et qui est consultable à l'URL : http://www.w3.org/TR/owl2-overview/. OWL, formalisme basé sur les logiques de description, est, depuis 2004, recommandation du Consortium World Wide Web pour la représentation et le partage d'ontologies sur le Web.

désireux d'aborder l'étude de Plotin. Au fil de leurs travaux de recherche, les utilisateurs pourront aisément augmenter la couverture conceptuelle et lexicale de cette base de connaissance. La puissance des requêtes qui peuvent y être faites ainsi que la possibilité d'inférer de nouvelles connaissances contribuera, à notre sens, à leur permettre d'approfondir le contenu conceptuel de l'œuvre, favorisant ainsi une compréhension plus profonde et une herméneutique correcte des aspects les plus complexes du domaine exploré.

5. Construction d'une ressource termino-ontologique

5.1. Structuration de la connaissance d'un domaine : l'Ontologie

Les modèles de représentation des connaissances véhiculées par les textes diffèrent selon les approches et le niveau de structuration et de formalisation. Partant d'une simple énumération alphabétique des mots, les *index*, on arrive à l'*ontologie*, qui représente le plus haut niveau de structuration de la connaissance, en passant par des niveaux intermédiaires comme la *folksonomie*, le *vocabulaire contrôlé*, la *taxonomie* et le *thesaurus*.

Le terme d'ontologie naît dans le domaine de la philosophie pour désigner la science ou étude de l'être en tant qu'être (du grec ὤν, ὄντος « étant », participe présent du verbe εἰμί « être ») dans ses caractères universels. Le Petit Robert définit ce terme comme une « partie de la métaphysique qui s'applique à l'être en tant qu'être, indépendamment de ses déterminations particulières ».

Le terme est repris ensuite dans les domaines de l'ingénierie des connaissances et de l'intelligence artificielle où la définition la plus consensuelle est celle proposée par Gruber (1993 : 199) : « une ontologie est une spécification explicite d'une conceptualisation ». De légères modifications y ont été apportées par la suite par Borst (1997), puis par Studer (1998).

Construire une ontologie, c'est donc construire un modèle abstrait, structuré, interprétable par la machine, qui définit de manière formelle, non ambiguë et consensuelle les concepts, les relations et toute autre distinction nécessaire à la modélisation des connaissances d'un domaine, afin d'en favoriser le partage.

Une ontologie se compose des concepts (ou classes) qui organisent un domaine, de leurs propriétés intrinsèques (axiomes et fonctions) et de leurs corrélations. Peuplée d'instances (ou individus) qui représentent la définition extensionnelle des concepts, elle constitue une base de connaissance. Étant

exploitable par la machine, elle offre des possibilités d'analyse, de raisonnement déductif et d'inférence de nouvelles connaissances du domaine exploré.

Grâce à l'utilisation de langages standards pour la représentation des connaissances qui favorisent la réutilisation du savoir, l'ontologie est devenue l'élément-clé du Web sémantique, dont le but est précisément de structurer les données existant sur le Web de manière à les rendre compréhensibles aux agents logiciels afin de leur permettre d'effectuer des recherches complexes et d'extraire des informations ciblées.

5.2. Modèle lexical

Convaincues de l'utilité d'une telle modélisation conceptuelle pour la structuration et la définition de la terminologie plotinienne, notre choix s'est donc porté sur un modèle lexical basé sur une ontologie.

L'élaboration d'un tel modèle constituant cependant une tâche longue, complexe et onéreuse, il nous est apparu plus sage et avantageux en matière de coût, de temps mais aussi de qualité des résultats, de nous inspirer de la structure d'un modèle existant, en l'adaptant aux exigences de notre domaine de la connaissance.

À cet effet, les deux principaux modèles lexicaux à base ontologique conçus durant les quinze dernières années pour le développement de grandes ressources lexicales informatisées ont été pris en examen. Il s'agit, d'une part, de WordNet (Fellbaum 1998) et EuroWordNet (Vossen 1999), basés sur des réseaux conceptuels ; et d'autre part, de SIMPLE (Lenci *et al.* 2000) et de Brandeis Semantic Ontology (Pustejovsky *et al.* 2006) qui conjuguent classification ontologique et valence sémantique.

Le modèle WordNet, basé sur une organisation des termes en groupes de synonymes (*synsets*) reliés entre eux par des relations, présente l'avantage d'offrir une grande cohérence dans l'établissement des liens hyperonymiques et la formulation des liens hyponymiques mais ne permet pas d'exprimer les nuances de sens existant entre des termes qui, bien qu'appartenant au même *synset*, ont une signification voisine sans être réellement interchangeables.

C'est là une des raisons pour lesquelles nous avons choisi de nous inspirer du modèle lexical SIMPLE[12]. Celui-ci a pour unité de base non pas le sens

12 Brandeis Semantic Ontology reprend largement le modèle SIMPLE, dont l'approche théorique s'inspire des principes fondamentaux de la théorie du Lexique Génératif de Pustejovsky (1995).

d'un ensemble de termes mais bien chacun des sens d'un lexème ; de plus, les relations lexico-sémantiques y jouent un rôle fondamental dans la modélisation des données lexicales.

L'approche théorique qui sous-tend la représentation de l'information sémantique est inspirée des principes fondamentaux de la théorie du Lexique Génératif (Pustejovsky 1995). La structuration conceptuelle du lexique est confiée à une ontologie linguistique mettant en relief le caractère multidimensionnel du sens lexical.

Outre la classification ontologique et un certain nombre de traits sémantiques, les unités lexicales sont caractérisées par un réseau de relations sémantiques qui sont en majeure partie le fruit d'une relecture de la Structure des Qualia (*Qualia Structure*) que Pustejovsky emprunte à Moravcsik (1975). La structure des *qualia*[13] est composée de quatre rôles permettant de modéliser l'aspect componentiel du sens d'un mot. Le rôle formel situe l'entité dénotée parmi d'autres, le rôle télique en indique la fonction, le rôle agentif l'origine et le rôle constitutif la composition. Outre les relations *qualia*, d'autres liens tels que la synonymie, la polysémie logique et la dérivation morphologique sont également définis.

Le modèle SIMPLE, récemment adapté avec succès à un domaine spécifique de la connaissance : le vocabulaire saussurien (Ruimy *et al.* 2013), offre donc un cadre riche mais cependant flexible pour la structuration du lexique.

6. Modélisation de la terminologie plotinienne

6.1. L'ontologie plotinienne

Nous inspirant de l'approche adoptée par Reymonet (2007) et dans le but de gérer aisément termes et concepts, nous avons construit l'ontologie plotinienne en modélisant le niveau conceptuel et le niveau terminologique par deux classes disjointes, CONCEPT et TERM. Notre ontologie se différencie donc de celle du modèle inspirateur SIMPLE dans lequel chaque classe ontologique regroupe un certain nombre d'unités lexicales partageant un ensemble de propriétés sémantiques.

L'ontologie plotinienne n'a pas été construite directement à partir de textes. Son point de départ a été une liste de termes représentatifs du domaine

13 Un des quatre niveaux de représentation sémantique dans la théorie du Lexique Génératif.

étudié, fournie en langue anglaise et grecque par la coordinatrice du projet, Cristina D'Ancona, spécialiste du néoplatonisme, qui a également validé en grande partie notre choix et application des relations conceptuelles et lexicales. Comme première étape dans la définition de l'ontologie, un ensemble de classes conceptuelles dénotées par ces termes a tout d'abord été identifié et organisé en hiérarchie. Ces classes ont été dans certains cas généralisées ou spécialisées à travers l'identification de concepts plus généraux ou plus spécifiques. Chaque concept identifié est représenté par une classe ; l'ensemble de ces classes étant subsumé par la superclasse CONCEPT.

La superclasse TERM s'articule, quant à elle, en deux sous-classes GREEK TERM et ARABIC TERM dont les termes grecs et arabes représentent respectivement les feuilles.

Dans notre base de connaissance, la partie conceptuelle est exprimée en anglais tandis que la partie terminologique est constituée de termes grecs ainsi que des termes arabes qui en sont la traduction. Les termes des deux langues sont interconnectés à travers les relations *translates* et *isTranslatedAs* et reliés à la partie conceptuelle à travers la relation *denotes* (cf. § 7.1.).

Comme précédemment indiqué, les sous-classes de la superclasse CONCEPT subsument à leur tour un certain nombre de classes conceptuelles, en une hiérarchie à cinq niveaux (cf. Fig. 1) :

- ENTITY regroupe des concepts tels que NATURE, LIVING_BEING et INTELLIGIBLE_WORLD ;
- PROPERTY subsume entre autres les concepts RATIONAL, INNATE et UNCHANGEABLE ;
- FACULTY réunit des concepts tels que DISCURSIVE_THOUGHT, RATIONALITY et UNDERSTANDING ;
- PROCESS compte parmi ses concepts CONVERSION, PROCESSION, CAUSE_TO_SUBSIST ;
- STATE regroupe des concepts comme ETERNITY, HARMONY et ENTELECHY ;
- CHANGE_OF_STATE subsume, BECOMING, CHANGE et COMING_TO_BE_AND_PASSING_AWAY ;
- PRINCIPLE regroupe les concepts THE_ONE, INTELLECT et SOUL.

L'étape successive dans le développement de l'ontologie a été consacrée à la définition de la structure interne des concepts : d'une part leurs propriétés

(attributs et valeurs) ; d'autre part les valeurs possibles des attributs : valeur booléenne, caractères, liste, chiffre, ainsi que leur cardinalité.

En observant les classes subsumées par la superclasse CONCEPT : ENTITY, PROPERTY, FACULTY, PROCESS, STATE, CHANGE_OF_ STATE et PRINCIPLE, les spécialistes pourraient avancer des objections ou tout au moins manifester quelque perplexité face à la structuration que nous proposons ; celle-ci, en effet, ne reflète pas les catégories selon lesquelles Plotin découpe le réel.

Plotin affronte le problème des catégories (Περί τῶν γενῶν τοῦ ὄντος, *Sur les genres de l'être*) dans un traité que Porphyre a subdivisé en trois et qui constituent les trois premiers traités de la sixième *Ennéade* VI 1-3 [42-44].

Ayant analysé et critiqué la doctrine des dix catégories d'Aristote (VI 1[42], 1-24) ainsi que celle des quatre catégories de la pensée stoïcienne (VI, 1 [42], 25-30), Plotin identifie dans le deuxième et le troisième traité celles qui, selon lui, sont les catégories, ou plutôt les genres (γένη) du monde intelligible (VI, 2) et du monde sensible (VI, 3).

En ce qui concerne le monde intelligible, en développant des idées déjà embryonnaires dans le *Sophiste* de Platon, dont il s'inspire, Plotin identifie cinq catégories : l'ÊTRE (ὄν), le MOUVEMENT (κίνησις), la STABILITÉ (στάσις), l'IDENTITÉ (ταὐτόν) et la DIFFÉRENCE (θάτερον). Dans le monde sensible, il distingue également cinq genres : l'ESSENCE ou SUBSTANCE (οὐσία), formée par la matière, la forme et le composé, la RELATION (πρός τι), la QUANTITÉ (ποσόν) et la QUALITÉ (ποιόν), le TEMPS (χρόνος) et le LIEU (τόπος), le MOUVEMENT (κίνησις).

Face à la difficulté qu'aurait pu présenter une telle articulation des concepts pour un usager qui entend s'initier à la pensée du philosophe néo-platonicien, nous avons délibérément choisi d'adopter un compromis en proposant, pour le premier niveau de structuration de la classe CONCEPT, des classes plus générales et plus intuitives, faisant l'objet d'un vaste consensus.

En revanche, notre fidélité à la pensée plotinienne transparaît dans la définition des concepts plus spécifiques du domaine, situés à partir du troisième niveau de profondeur de l'ontologie, à travers l'usage de relations et de propriétés sémantiques spécifiquement créées dans le but de modéliser finement ce domaine de la connaissance.

En développant cette ontologie, nous nous sommes efforcées d'adhérer aux principes de complétude et de cohérence et de prévoir la possibilité d'extensions ou de raffinements. Notre objectif principal a été de tendre

vers une modélisation la plus consensuelle possible des concepts et des relations constitutives propres au domaine de la connaissance exploré, en tenant compte de l'objectif applicatif et dans le but de permettre une recherche d'information efficace et exhaustive et l'acquisition de connaissances lexicales à partir de textes. Il est cependant clair que le système conceptuel ainsi conçu ne représente qu'une des modélisations possibles du domaine de la terminologie plotinienne. Le développement d'une ontologie est en effet un processus créatif dans lequel s'inscrit inévitablement un facteur de subjectivité et, en dépit de tout effort de consensus, une ontologie reflète toujours un point de vue. La version actuelle de l'ontologie pourrait, bien évidemment, être sujette à évolution, et en particulier à des raffinements visant à une définition plus précise des classes sémantiques ainsi qu'à un enrichissement par de nouveaux concepts.

6.2. Propriétés et relations sémantiques

Outre la composante ontologique du système, un ensemble de propriétés et de relations permettent de représenter une large typologie d'informations caractérisant le contenu conceptuel et lexical des termes. Les 10 propriétés sémantiques dont dispose le modèle plotinien (par exemple : *intentionality*, *will*, *space*, *time*, *active*, *movement*, *continuity*, etc.) expriment, sous forme de couples attributs/valeurs, un certain nombre de traits distinctifs transversaux à la hiérarchie des types et permettent le regroupement de certains concepts, en dépit de leur appartenance à des classes ontologiques différentes.

Les relations, quant à elles, expriment sous forme de triplets <*sujet, prédicat, objet*> une relation binaire entre un couple de concepts ou de termes. Deux ensembles distincts de relations sont en effet utilisés : les relations lexicales et les relations conceptuelles. Nous entendons par 'relations lexicales' les relations paradigmatiques que les sens des mots entretiennent dans les textes, qu'elles soient taxonomiques (hyperonymie, hyponymie), partitives (méronymie, holonymie), ou symétriques non hiérarchiques (synonymie, antonymie)[14]. Les relations lexicales doivent être différenciées des relations

14 Notons qu'il n'existe pas véritablement de consensus concernant la terminologie en usage. Le terme 'relation lexicale' est utilisé dans la littérature aussi bien comme synonyme de 'relation sémantique' qu'en opposition à ce dernier. Ainsi, Cruse (1986) parle de 'lexical relations' pour indiquer la synonymie, l'hyperonymie, etc., tandis que Murphy (2003) les nomme 'semantic relations', et désigne par 'lexical relations' toutes les relations paradigmatiques entre les mots, non seulement les relations sémantiques mais aussi les relations phonétiques, morphologiques et morphosyntaxiques.

conceptuelles, c'est-à-dire des relations que les concepts, dénotés par des termes, entretiennent[15].

Parmi elles, outre la relation de subsomption *is*-a, certaines sont empruntées à la *Extended Qualia Structure* du modèle SIMPLE, qui est elle-même le fruit d'une relecture de la structure des *qualia*. Les relations de l'*Extended Qualia* qui, dans SIMPLE, permettent de fournir une représentation structurée de la multidimensionnalité d'un sens et de saisir la nature des relations (intra- et extra-catégorielles) existant entre les unités lexicales, ont été ici utilisées afin de modéliser la complexité interne des concepts et les relations qui les lient. L'ensemble des relations dont nous disposons pour la définition des concepts de la philosophie plotinienne permet donc, en résumé, de relier un concept dénotant un objet à un concept plus générique, à ses propriétés, à son origine et à sa fonction ; une entité-événement, peut, quant à elle, être reliée à son agent, son patient, son instrument, son éventuel lieu d'action, ainsi qu'à son effet ou résultat.

Outre les relations précédemment indiquées, 17 relations ontologiques spécifiques ont été créées afin d'exprimer les liens caractérisant l'organisation conceptuelle du domaine exploré. Ainsi : *isThePrincipleOf* (dénote le lien entre un principe et celui qui lui est immédiatement inférieur), *emanatesFrom* (exprime le lien entre une réalité et le principe dont elle dérive), *isRealmOf* (dénote une collocation abstraite), *hasNaturalProperty* (indique une propriété intrinsèque), *hasParticipatedProperty* (indique une propriété partagée avec le principe supérieur), *governs* (dénote l'activité de l'âme et des âmes), *contemplates* (exprime l'attitude d'un principe envers son principe supérieur), etc.

15 Concernant les relations conceptuelles, les divergences terminologiques sont moins marquées, en dépit de différences dans leur catégorisation. Cabré (1998) distingue les 'relations ontologiques' et les 'relations logiques'. Les premières peuvent être simultanées, c'est-à-dire basées sur la proximité spatiale, ou séquentielles, i.e. basées sur la contiguïté temporelle des objets. Les relations logiques, par contre, sont basées sur la similarité entre les concepts : relation d'inclusion (rapport species-genus) et de subordination. L'Organisation internationale de normalisation ISO définit, quant à elle, deux types de relations conceptuelles : les relations hiérarchiques, subdivisées en relations génériques (inclusion) et partitives, et les relations associatives, non hiérarchiques, comme les relations séquentielles, temporelles et causales.

7. Implémentation du modèle

7.1. Les classes ontologiques

Les concepts identifiés lors de la première étape de constitution de l'ontologie ont donc été traduits en 'classes nommées' de OWL (owl :Class) en respectant les contraintes liées au formalisme adopté[16]. Comme indiqué au paragraphe 6.1., le système conceptuel élaboré est constitué d'un premier niveau de structuration comprenant deux superclasses, chacune d'entre elles accueillant un certain nombre de sous-classes. Chaque sous-classe de la classe CONCEPT subsume, à son tour, un certain nombre de classes. Parmi ces dernières, la classe SOUL, par exemple, subsume deux autres niveaux de classes conceptuelles (cf. Fig. 1) :

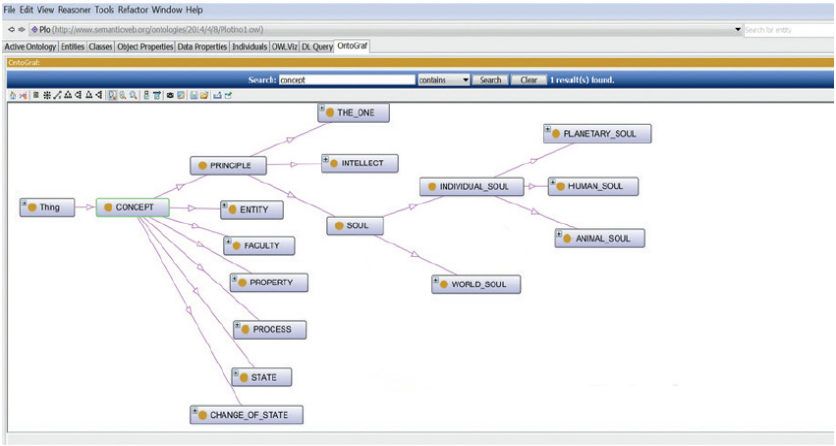


Figure 1 – *Les cinq niveaux de profondeur de la classe CONCEPT.*

Conformément à l'approche adoptée, chaque classe ontologique subsumée par les classes CONCEPT et TERM compte un unique membre, qui est en réalité une instanciation fictive du concept ou du terme. Il convient en effet de noter qu'étant donnée l'impossibilité de lier une classe et un individu dans le langage OWL, il a été nécessaire de créer des instances fictives pour chaque

16 Il est ainsi important de noter que « Chaque individu du monde OWL est membre de la classe owl :Thing. Chaque classe définie par l'utilisateur est donc implicitement une sous-classe de owl : Thing » (http://www.yoyodesign.org/doc/w3c/owl-guide-20040210/#SimpleClasses).

concept et pour chaque terme et de les relier à travers les propriétés d'objet *denotes* et *isDenotedBy*.

7.2. Les propriétés

Dans OWL, les propriétés des classes et des individus sont de deux types.

D'une part, les 'propriétés de types de données' (*DataType Properties*), qui relient des individus à des valeurs de données (valeur booléenne, caractères, liste, chiffre) et qui traduisent les propriétés sémantiques caractéristiques du domaine.

D'autre part, les 'propriétés d'objet' (*Object Properties*), relations binaires qui relient des individus à d'autres individus et traduisent donc nos relations conceptuelles et lexicales. Ces relations peuvent être hiérarchisées et avoir des sous-relations qui sont des spécialisations de la relation mère, comme c'est le cas de *hasProperty* et de ses deux sous-relations : *hasNaturalProperty* et *hasParticipatedProperty*. Les propriétés d'objet peuvent, d'autre part, être soumises à restrictions : il est possible de spécifier les classes autorisées à constituer leur 'domaine' et leur 'codomaine', leurs caractéristiques (inverse : *hasPart*/*isPartOf* ; fonctionnalité : une seule valeur admise ; transitivité ; symétrie ; antisymétrie ; réflexivité ; irréflexivité, etc.) ainsi que leur cardinalité (nombre minimal, maximal ou exact de valeurs distinctes autorisées).

Les 'propriétés d'annotation' (*Annotation Properties*) permettent, quant à elles, d'encoder les métadonnées des classes, des propriétés et des individus (versions, commentaires, définitions, exemples). Elles ont été utilisées à ce jour pour fournir une définition et un ou plusieurs exemples de chaque propriété de types de données et propriété d'objet ainsi que pour indiquer, par rapport au concept traité, tous les autres termes anglais le dénotant.

8. Définition des termes et des concepts

Un exemple : description du terme τὸ_ἕν et du concept qu'il dénote, THE_ONE (Tab. 1)[17].

Les quatre premières lignes du Tableau 1 concernent la description du terme τὸ_ἕν (l'Un) : ses relations de synonymie, sa traduction en langue

17 Cf. Pigler (2003).

arabe, الواحد, ainsi que son lien au concept qu'il dénote (THE_ONE) et dont la description suit celle du terme.

Concernant le concept THE_ONE, nous exprimons le fait que l'Un n'est ni οὐσία, ni τέλος grâce à la notion d'assertion négative 'assertions négatives de propriété d'objet' (*negative object property assertions*), introduite dans OWL 2, et qui permet d'exprimer le fait qu'un individu n'est pas lié par une propriété donnée à un autre individu ou à une valeur de donnée.

Suit un ensemble de propriétés exprimées par des relations (*Object Properties*) qui définissent le concept THE_ONE en le reliant à d'autres concepts de l'ontologie.

Les deux derniers traits (*Data Properties*) : *will:yes* et *intentionality:no*, en apparence contradictoires, trouvent leur justification dans la théorie plotinienne de l'*autoctisis* de l'Un. En effet, dans les premiers chapitres du traité VII de la sixième *Ennéade*, Plotin argumente longuement le fait que l'Intelligence divine n'exerce pas de raisonnement (λογισμός), et donc à plus forte raison l'Un. Il n'y a aucun λογισμός en Dieu, c'est-à-dire dans l'Un et dans son activité créatrice de toute chose. La multitude procède de l'Un, par un acte de sa Volonté, comme conséquence de l'autovolition libre de l'Un (*will :yes*). L'Un est donc *causa sui*, principe qui renferme en lui-même sa cause et qui, dans sa liberté créatrice, réalise sa volonté, son essence : οὗτο δέ ἐστιν ἔχον ἐν ἑαυτῷ καὶ τὴν αἰτίαν et ἡ βούλησις αὐτοῦ καὶ ἡ οὐσία ταὐτὸν ἔσται (VI 8, 13.7 : « la volonté et l'essence de Dieu sont une seule et même chose »). Dans cette activité à la fois créatrice et autoproductrice, l'Un exerce sa volonté qui coïncide avec son essence. Cette opération d'*autoctisis* est bien libre mais elle est aussi non intentionnelle, c'est-à-dire nécessaire (*intentionality :no*).

Dans la colonne 'inferred' située à l'extrême droite du Tableau 1, les propriétés directement inférées par le système, grâce à un moteur de raisonnement[18], sont indiquées par la valeur *yes*.

À ce jour, 142 concepts et termes ont été définis.

18 Protégé inclut des moteurs d'inférence reposant sur la logique de description, tels que Pellet, Fact++, RacerPro, etc. qui détectent les incohérences formelles des modélisations effectuées et raisonnent sur les ontologies afin de déduire de nouvelles connaissances.

SOURCE TERM / CONCEPT	OBJECT/DATA PROPERTY	TARGET TERM / CONCEPT	INFERRED
τὸ_ἕν	hasSynonym	τὸ_ἀγαθόν	no
	hasSynonym	ὁ_Θεός	no
	isTranslatedAs	الواحد	no
	denotes	tha_One	no
THE_ONE	isNot	completion	no
	isNot	essence	no
	hasExternalActivity	Being	no
	hasExternalActivity	Intellect	no
	hasProperty	above_being	no
	hasProperty	impassibility	no
	hasProperty	indivisibility	no
	hasProperty	infinite	no
	hasProperty	omnipresence	no
	hasProperty	perfect	no
	hasProperty	power	no
	hasProperty	self_sufficiency	no
	hasProperty	simplicity	no
	hasProperty	unchangeable	no
	hasProperty	unmoved	no
	hasProperty	unpredicable	no
	isBeyond	Being	no
	isBeyond	Forms	no
	isBeyond	Intellect	no
	isBeyond	essence	no
	isBeyond	intelligible_world	yes
	isBeyond	soul	no
	isContemplatedBy	Being	no
	isContemplatedBy	Intellect	no
	isSourceOf	Being	no
	isSourceOf	Intellect	no
	isSourceOf	life	no
	isSourceOf	tha_Beauty	yes
	isThePrincipleOf	Intellect	no
	isThePrincipleOf	Being	yes
	possesses	life	no
	isRelatedTo	eternity	yes
	isDenotedBy	τὸ_ἕν	no
	isDenotedBy	τὸ_ἀγαθόν	yes
	isDenotedBy	ὁ_Θεός	yes
	intentionality	no	no
	will	yes	no

TAB. 1 – *Le term τὸ_ἕν et son concept THE_ONE.*

9. Les termes grecs et leur traduction arabe

Une telle structuration de la connaissance, c'est-à-dire la distinction entre plan terminologique et plan conceptuel — et à l'intérieur du plan terminologique la distinction entre terme grec et terme arabe — peut facilement faire ressortir tous les cas où le traducteur arabe ne se limite pas à traduire le texte grec, mais l'interprète et le commente.

Un cas très intéressant est représenté par la traduction arabe du terme grec θεός. Dans le corpus plotinien arabe, différents termes sont utilisés pour indiquer Dieu, chacun desquels met en évidence un aspect particulier de la divinité. Parmi eux, on trouve : الإله (*'ilāh* : le dieu en général) ; النُّور الأَوَّل (*'an-nūr al-'awwal* : littéralement « la lumière première ») ; الله : (*'allāh* : le nom propre du dieu musulman) ; البَارِي (*'al-bārī* : le créateur).

Voici un des nombreux passages illustrant les motivations théologiques qui sous-tendent le choix d'un certain terme par le traducteur, en particulier les raisons pour lesquelles il préfère dans un cas le terme الإله et dans l'autre le terme الله.

[…] Εἰ οὖν καὶ ὁ θεὸς αὐτοῖς κατὰ τὸν νοῦν ὕστερος καὶ γεννητὸς καὶ ἐπακτὸν τὸ νοεῖν ἔχων, ἐνδέχοιτο ἂν μηδὲ ψυχὴν μηδὲ νοῦν μηδὲ θεὸν εἶναι.

(Ennéades, IV 7, 8^3.11-13)

وإن لجّوا وقالوا : إنّ العقل بعد النفس والنفس بعد الطبيعة - لَزِم من قولهم أن يكون الإله - تبارك وتعالى - بعد العقل ، واقعا تحت الكون والفساد ، عالما بعَرَض ، وذلك محالٌ لأنّه إن أمكن أن يكون هذا الترتيب حقّا أمكن أن يكون لا نفس ولا عقل ولاً إله - وهذا محالٌ قبيحٌ جدًّا.

وأما نحن فنقول إنّ الله - عز وجلّ - علّة للعقل والعقل علّة للنفس والنفس علّة للطبيعة والطبيعة علّة للأكوان كلّها الجزئية. غير أنّه وإن كانت الأشياء بعضُها علّة لبعض فإنّ الله تعالى علّة لجميعها كلّها : غير أنّه علّة لبعضها بغير توسُّط وهو الذي جعل العلّة كما قلنا فيما سلف.

(*'aflutīn 'inda l-'arab*, maymar III, page 50, ligne 5)

Le passage de Plotin a été extrait du septième traité de la quatrième *Ennéade*, traité qu'il dédie tout entier au thème de l'immortalité de l'Âme. Il rejette en particulier la doctrine stoïcienne selon laquelle l'âme est le corps : s'il en était ainsi, l'on finirait par affirmer que la perfection naît de l'imperfec-

tion. Par absurde, Dieu serait généré par un principe qui lui serait supérieur et il n'y aurait donc ni âme, ni intelligence, ni Dieu[19].

Comme on peut le remarquer, la péricope arabe de la pseudo-*Théologie* est beaucoup plus longue que la péricope plotinienne. Nous avons ici un des nombreux passages typiques du corpus plotinien arabe dans lequel le traducteur ne se limite pas à traduire mais expose plutôt son propre credo.

La première partie du texte arabe représente une traduction assez fidèle de l'original grec. Dans ce passage où Plotin montre le caractère fallacieux du raisonnement stoïcien, en arrivant à la conclusion absurde de la non-existence de Dieu, le traducteur traduit le terme grec θεός par le terme arabe générique الإله, . Il marque son détachement par rapport au texte qu'il était en train de traduire en utilisant la troisième personne plurielle : « **وإن لجّوا وقالوا** » (« et s'ils s'opposent et disent »).

Cependant, un peu plus loin, dans un long passage qui n'a pas de correspondance dans le texte grec, il oppose la doctrine de l'existence et de l'unicité de Dieu au raisonnement par absurde de Plotin. C'est dans ce passage, introduit par **وأما نحن فنقول** (« par contre nous répondons ») que le traducteur choisit le terme الله, qui désigne par excellence Dieu, la cause première de toutes choses, dans la culture arabo-islamique.

10. Interrogation de la base de connaissance

Une base de connaissance offre une multitude de possibilités d'interrogation et d'exploitation automatisée des données. L'intégralité des informations qui y sont présentes, de quelque nature qu'elles soient, est exploitable pour la formulation de requêtes. Requêtes dont l'exhaustivité des résultats est garantie par la structuration et formalisation de l'information saisie et dont les

19 *Ennéades*, IV, 7, 10 : « Εἰ οὖν καὶ ὁ θεὸς αὐτοῖς κατὰ τὸν νοῦν ὕστερος καὶ γεννητὸς καὶ ἐπακτὸν τὸ νοεῖν ἔχων, ἐνδέχοιτο ἂν μηδὲ ψυχὴν μηδὲ νοῦν μηδὲ θεὸν εἶναι » : « Enfin, dans ce système, Dieu, par cela même qu'il possède l'intelligence, est postérieur, engendré, n'a qu'une intelligence adventice ; il en résulte qu'il n'y a ainsi ni âme, ni intelligence, ni Dieu ».

temps de réponse sont minimes. Ces résultats permettent en outre de vérifier la cohérence formelle de l'information codée.

10.1. Les outils Protégé

10.1.1. Représentation graphique des données

La plateforme Protégé offre deux systèmes de représentation graphique. Le module OWLViz permet de visualiser l'ontologie tout entière ou de l'examiner en partie, à la profondeur désirée et en incluant éventuellement le résultat des inférences faites par le raisonneur. Le cas échéant, les inférences sont distinguées de l'information définie par le concepteur de l'ontologie par une présentation de couleur différente.

Le module Ontograph permet, quant à lui, de rechercher un individu (terme ou concept) et de visualiser les relations ('propriétés d'objet') qu'il entretient avec d'autres individus présents dans la base de connaissance, mettant ainsi en évidence tout un réseau de connexions (cf. Fig. 2) :

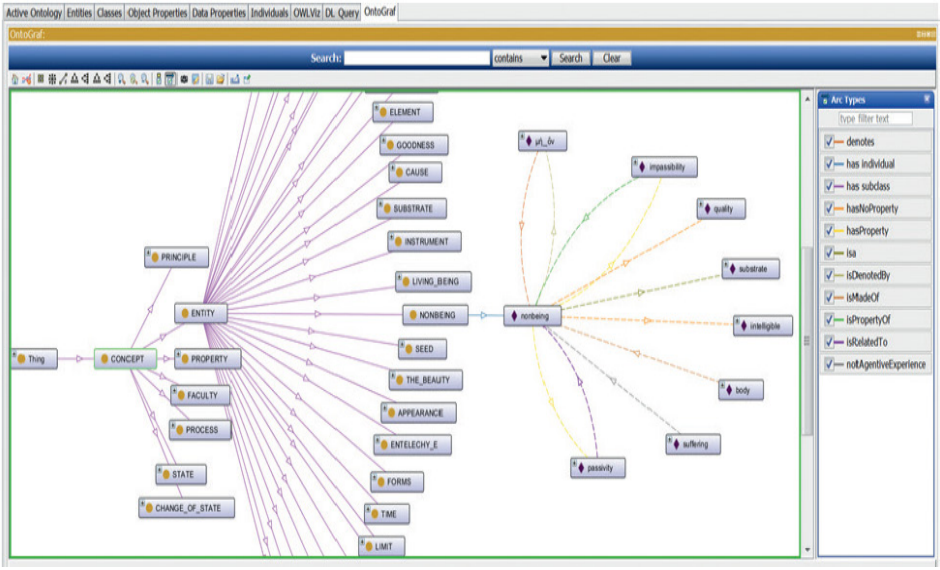


Figure 2 – *Ontograph : représenatation graphique du concept NONBEING.*

10.1.2. Système de requêtes

Protégé propose également deux puissants outils d'usage relativement simple dont le fonctionnement est basé sur des moteurs de raisonnement sémantique et qui permettent d'interroger une ontologie à travers un système de requêtes exprimées en langage formel.

Le premier, DL Query, utilise un langage de requête basé sur la syntaxe Manchester OWL[20]. À titre d'exemple, nous proposons ci-dessous trois requêtes exprimées en langage naturel suivies de leur traduction en syntaxe Manchester.

Quels sont les termes dénotant un procès dont l'agent est l'Intelligence ?

process and hasInvolvedAgent value Intellect

Quels sont les concepts auxquels ne peut être appliquée la notion de temps ?

time value "no"

Quelles sont les propriétés que partagent l'Intelligence et l'âme ?

isPropertyOf value Intellect and isPropertyOf value soul

Le second outil de requêtes, OWL2Query, utilise la syntaxe SPARQL. Ce langage est une recommandation du W3C (janvier 2008) et est considéré comme l'une des technologies clés du Web sémantique. SPARQL permet également de formuler une requête sous forme de graphe formé de triplets comportant des variables.

10.2. Le module Lexicon de la G2A Web Application

Le module *Lexicon* inséré dans la plateforme de philologie numérique *G2A Web Application* évoquée dans l'Introduction offre un système d'interrogation de la ressource termino-ontologique à travers une interface de type *Question Answering* conçue comme un ensemble de *templates* (ou patrons) en langage naturel semi-contrôlé. Chaque *template* permet de formuler plusieurs types de requêtes grâce aux menus déroulants qu'il propose. Ces requêtes sont ensuite converties et transmises au système en langage formel. Guidé par

20 Cf. URL : http://www.w3.org/TR/owl2-manchester-syntax/.

le *template* qu'il a sélectionné, l'utilisateur peut donc formuler sa requête en langage naturel.

Ainsi, s'il désire connaitre les relations lexicales d'un terme de son choix et les relations et propriétés du concept que ce terme dénote (*Query 1. Querying a term and its concept*), il sélectionnera dans le menu déroulant le terme choisi, par exemple ὕλη et obtiendra, sous forme de tableau et dans un format exportable, les résultats illustrés dans la Figure 3 :

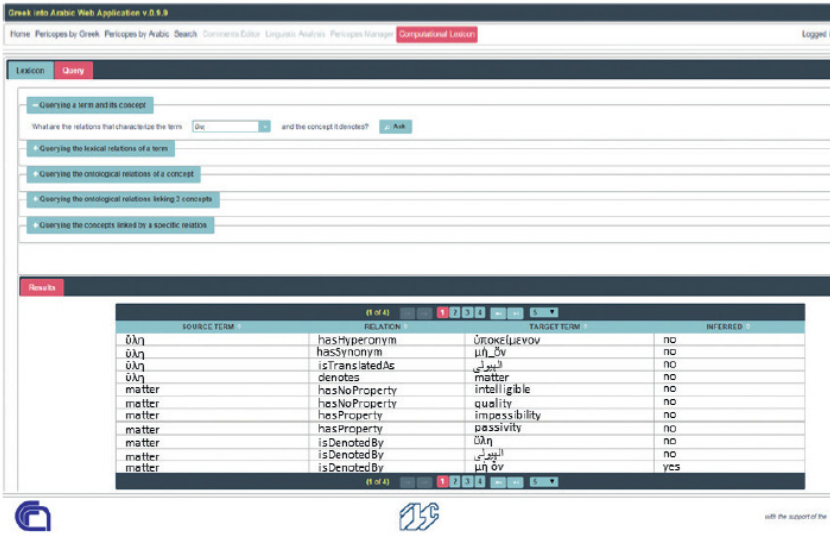


Figure 3 – *Requête sur un terme et son concept.*

L'utilisateur pourrait ne désirer connaître que les relations lexicales qu'un terme entretient avec d'autres termes dans le lexique plotinien (*Query 2. Querying the lexical relations of a term*) ; ou encore seulement les relations ontologiques qu'un concept de son choix entretient avec d'autres concepts (*Query 3. Querying the ontological relations of a concept*).

L'utilisateur peut également requérir les liens existants en trois concepts (*Query 4. Querying the ontological relations linking 3 concepts*). Par exemple,

la sélection des concepts *eternity*, *time*, *soul* fournit les résultats illustrés dans la Figure 4 :

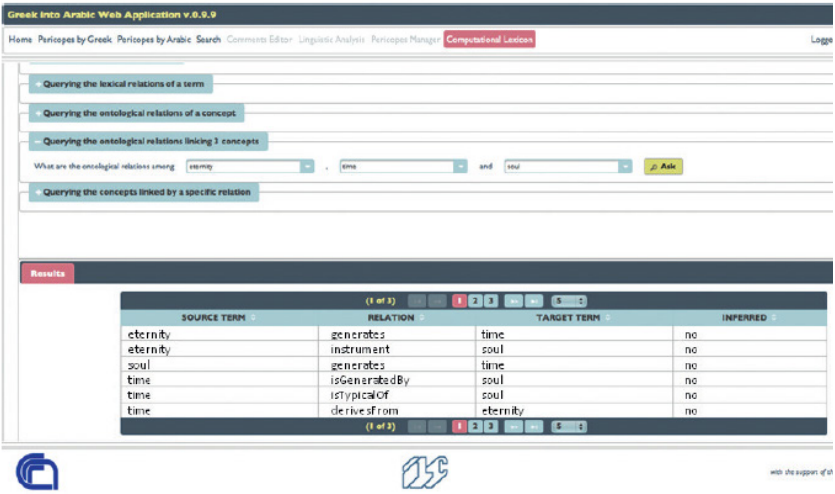


Figure 4 – *Requête sur les relations ontologiques reliant 3 concepts.*

Il lui est également possible de savoir quels sont les concepts associés à un concept sélectionné, par une relation déterminée (*Query 5. Querying the concepts linked by a specific relation*). Ainsi si la question que l'utilisateur se pose est « Qu'est-ce qui ne subit pas le πάθος, selon Plotin ? », il lui suffira de choisir le concept *suffering* dans la première fenêtre et la relation *notAgentiveExperience* dans la seconde, pour obtenir les résultats illustrés dans la Figure 5 :

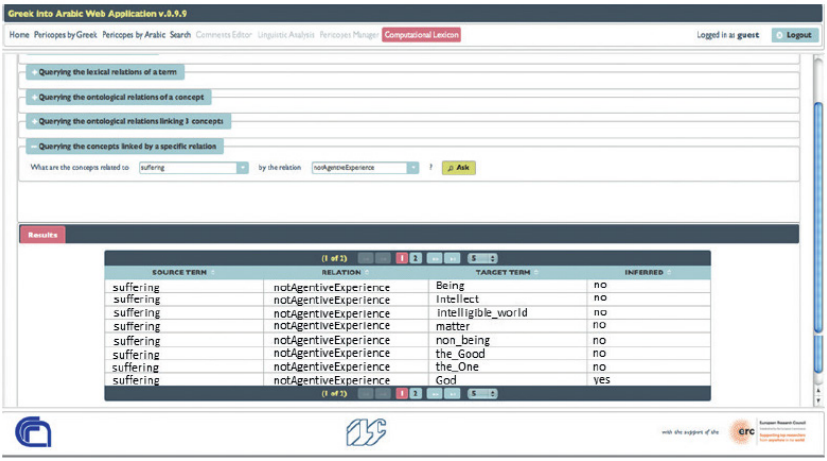


Figure 5 – *Requête sur les concepts reliés par une relation spécifique.*

Les recherches effectuées ci-dessus ne sont fournies qu'à titre d'exemples, dans le but d'illustrer tant la simplicité d'usage que les potentialités du système de requêtes. Nul doute qu'un expert du domaine saura formuler des requêtes bien plus ciblées et pertinentes pour l'exploration du domaine.

Dans la plateforme *Lexicon*, un outil de représentation graphique des données est en cours de mise au point. Il permettra de visualiser sous forme de réseau l'ensemble des liens que les termes ou les concepts du lexique entretiennent. Ceci pourrait permettre de faire émerger des connections qui n'apparaissent pas immédiatement de manière explicite et de suggérer ainsi des parcours alternatifs d'analyse et donc de compréhension de la pensée de l'auteur.

11. Conclusions

L'objectif des auteurs de la ressource termino-ontologique *Plotiterm* présentée dans cet article était de démontrer qu'une telle modélisation et formalisation des connaissances du domaine peut efficacement contribuer à appréhender de manière plus exhaustive la richesse de contenu des textes explorés afin d'en favoriser une herméneutique correcte. Cette structuration formelle permet au spécialiste d'effectuer aisément et rapidement tout type de recherche sur la terminologie utilisée ; elle favorise la perception de liens non explicites exis-

tant entre les différents concepts véhiculés par les termes dans le texte, contribuant ainsi à un approfondissement de certains des aspects les plus complexes du contenu conceptuel de l'œuvre. Une base de connaissance renfermant les données décrites tout au long de cet article a donc vocation à devenir un outil de recherche novateur pour des spécialistes accoutumés aux méthodes d'étude plus traditionnelles généralement utilisées dans leur domaine.

Références

A Greek & Arabic Lexicon (GALex). Materials for a Dictionary of the Mediaeval Translations from Greek into Arabic. (1992-). Endress G., Gutas, D. (eds.). Handbuch der Orientalistik, I.xi. Leiden : E.J. Brill.

Aussenac-Gilles, N., Biébow B., Szulman S. (2000). "Revisiting ontology design : a methodology based on corpus analysis". In Dieng, R., Corby, O. (eds). *Knowledge Engineering and Knowledge Management : Methods, Models, and Tools. Proceedings of 12th International Conference, EKAW 2000*, Juan-les-Pins, France, October 2000. Lecture Notes in Artificial Intelligence, Subseries of Lecture Notes in Computer Science. Berlin Heidelberg : Springer-Verlag, Vol. 1937, 172-188.

Badawi, A. (1966). *'aflutīn 'inda l-'arab*. Cairo : Maktabat al-Nahḍah al-'Arabiyyah.

Borst, W. N. (1997). *Construction of Engineering Ontologies for Knowledge Sharing and Reuse*. PhD thesis, University of Twente, Enschede, The Netherlands : Centre for Telematics and Information Technology.

Buitelaar, P., Cimiano, Ph., Haase, P., Sintek, M. (2009). "Towards Linguistically Grounded Ontologies". In Aroyo, L., Traverso, P., Ciravegna, F., Cimiano, P., Heath, T., Hyvönen, E., Mizoguchi, R., Oren, E., Sabou, M., Simperl, E. (eds). *The Semantic Web : Research and Applications. Proceedings of 6th European SemanticWeb Conference, ESWC 2009*. Lecture Notes in Computer Science. Berlin Heidelberg : Springer-Verlag, vol. 5554, 111-125.

Cabré, M.T. (1998). *Terminology. Theory, methods and applications*. Juan C. Sager (ed.). Amsterdam/Philadelphia : John Benjamins Publishing Company.

Cruse, A. (1986). *Lexical Semantics*. Cambridge : Cambridge University Press.

D'Ancona, C. (2005). "Tradizione greca e versione araba delle Enneadi : l'indipendenza reciproca e il caso del trattato Sull'immortalità dell'anima (IV 7[2])". In Finazzi, R.B. (a cura di). *Del tradurre. Da Occidente verso Oriente come incontro di lingue e culture. Atti della giornata di studio su*

"Traduzioni orientali e testi classici : lo stato della ricerca", Brescia, 8 ottobre 2004. Milano : I.S.U. Università Cattolica, 39-66.

Desprès, S., Szulman, S. (2008). "Réseau terminologique versus Ontologie". *Actes de la deuxième conférence TOTh*, Annecy, 5 et 6 juin 2008, 17-34.

Fellbaum, Ch. (1998). *WordNet. An Electronic Lexical Database*. Cambridge. MA : The MIT Press.

Gruber, T. R. (1993). "A translation approach to portable ontologies". *Knowledge Acquisition*, V (2), 199-220.

ISO 1087-1 (2000). *Travaux terminologiques – Vocabulaire – Partie 1 : Théorie et application.* Genève.

ISO 704 : 1987 (2000). *Travail terminologique – Principes et méthodes.* Genève.

Lenci, A., Bel, N., Busa, F., Calzolari, N., Gola, E., Monachini, M., Ogonowski, A., Peters, I., Peters, W., Ruimy, N., Villegas, M., Zampolli, A. (2000). "SIMPLE : A General Framework for the Development of Multilingual Lexicons". *International Journal of Lexicography*, XIII (4), 249-263.

Lexicon plotinianum. (1980). Sleeman, J.H., Pollet, G. (eds). Leuven : Leuven University Press. Leiden : E.J. Brill.

Moravcsik, J.M.E. (1975). "Aitia as Generative Factor in Aristotle's Philosophy". *Dialogue*, XIV, 622-636.

Murphy, M.L. (2003). *Semantic Relations and the Lexicon : antonymy, synonymy and other paradigms.* Cambridge : Cambridge University Press.

Nida, E. (1964). *Towards a Science of Translating.* Leiden : E.J. Brill.

Pigler, A. (2003). *Le vocabulaire de Plotin*, Paris : Ellipses.

Plotin (1924-1938). *Ennéades*. Texte établi et traduit par E. Bréhier, 7 vol. Paris : Les Belles Lettres.

Porphyre (1992). *Vie de Plotin*. Texte établi par Émile Bréhier. Introduction, notes et annexes par Sébastien Morlet. Paris : Les Belles Lettres.

Prandi, M. (2011). "Signes, signifiés, concepts : pour un tournant philosophique en linguistique". *Actes de la conférence TOTh*, Annecy – 26 et 27 mai 2011, 3-20.

Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge. MA : The MIT Press.

Pustejovsky, J., Havasi, C., Littman, J., Rumshisky, A., and Verhagen, M. (2006). "Towards a Generative Lexical Resource : The Brandeis Semantic Ontology". *Proceedings of Fifth International Conference on Language resources and Evaluation*, European Language Resources Association (ELRA), Genova, 1702-1705.

Radice, R., Bombacigno, R. (2004). *Plotinus*, Lexicon. Collana di lessici di filosofia antica. Milano : Biblia.

Reymonet, A., Thomas, J., Aussenac-Gilles, N. (2007). "Modelling ontological and terminological resources in OWL DL". In Buitelaar, P., Choi, K.-S., Gangemi, A., Huang, C.-R. (eds.). *OntoLex07 - From Text to Knowledge : The Lexicon/Ontology Interface*. Workshop at ISWC07 6th International Semantic Web Conference. Busan (South Korea), 415-425.

Roche, C. (2007). "Le terme et le concept : fondements d'une ontoterminologie". *Actes de la première conférence TOTh*, Annecy 1er juin, 2007, 1-22.

Roche, C. (2012). "Ontoterminology : How to unify terminology and ontology into a single paradigm". *Proceedings of Eight International Conference on Language Resources and Evaluation*, European Language Resources Association (ELRA). Istanbul. Turkey, 2626-2630.

Ruimy, N., Piccini, S., Giovannetti, E., Bellandi, A. (2013). "Lessicografia Computazionale e Terminologia Saussuriana". In Gambarara, D., Marchese, M. (a cura di). *Guida per un'edizione digitale dei manoscritti di Ferdinand de Saussure*. Alessandria : Edizioni dell'Orso, 161-179.

Sager, J.C. (1990). *A Practical Course in Terminology Processing*. Amsterdam : John Benjamins.

Studer, R., Benjamins, R.V., Fensel, D. (1998). "Knowledge engineering : Principles and methods". *Data & Knowledge Engineering*, XXV (1-2), 161-197.

Szulman, S., Biébow, B. (2004). "OWL et Terminae". *Actes de la 15e conférence d'Ingénierie des Connaissances IC*. Lyon : Presses Universitaires de Grenoble, 41-52.

Temmerman, R., Kerremans, K., Vandervoort, V. (2005). "La termontographie en contexte(s)". In Blampain, D., Thoiron, P., Van Campenhoudt, M. (eds). *Mots, Termes et Contextes. Actes des septièmes Journées scientifiques du réseau de chercheurs Lexicologie Terminologie Traduction*. Bruxelles, 429- 439.

Vossen, P. (1999). *Eurowordnet. A multilingual database with lexical semantic networks*. Dordrecht, Kluwer.

Summary

This article describes the termino-ontological resource *Plotiterm*, which is dedicated to modelling the terminology of Plotinus' *Enneads* and the translation of these works into Arabic, the pseudo-*Theology of Aristotle*.

This resource was developed in Protégé, an open-source ontology editor which supports the standard language OWL.

In *Plotiterm* the conceptual and terminological levels are separate but interconnected, which is in line with the new paradigms and methodologies which have been developed over recent years. In addition, the ontological structuration, a set of properties (DataType Properties) and relations (Object Properties) subdivided into lexical relations and conceptual relations, make it possible to express a wide range of information which characterizes the lexical and conceptual content of the terms. The data in *Plotiterm* can be queried or graphically explored using the tools provided by Protégé or using the module *Lexicon*, which is part of the Web based platform *G2A Web Application*.

Réingénierie de thésaurus pour l'indexation en Sciences du Langage. Retour d'expérience

Mario Marcon

ATILF UMR 7118 CNRS/Université de Lorraine – 44, avenue de la Libération
54000 Nancy
Mario.Marcon@atilf.fr
mario.marcon16@gmail.com

Résumé. Cet article présente une réflexion méthodologique issue de l'expérience de réingénierie du thésaurus *Thesaulangue* utilisé pour l'indexation des ouvrages en Sciences du Langage (désormais SdL) au Centre de documentation de l'ATILF. L'article décrit les étapes de la réingénierie du thésaurus en vue de la création d'une ontologie OWL. En conclusion, nous nous concentrons sur les pour et les contre de notre expérience.

1. Introduction

Le développement de l'ingénierie des connaissances et l'intérêt pour la gestion de terminologies ont favorisé la « représentation d'une modélisation d'un domaine partagée par une communauté d'acteurs » (Roche 2005 : 57), c'est-à-dire l'ontologie.

Des plateformes, comme *Ontology Design Patterns*[1], suggèrent des algorithmes d'alignement et des patrons de réingénierie de ressources non ontologiques pour la conception d'ontologies. Malgré le recours à des items différents (ex. *concept*, *descripteur*, *mot-clé*, *terme*, etc.), la réutilisation des contenus des ressources non ontologiques revêt une importance considérable pour la représentation – pour autant que possible – exhaustive des connaissances d'un domaine.

1 *Ontology Design Patterns* : http://ontologydesignpatterns.org/ (page consultée le 10/02/2016).

En ce qui concerne la structuration des connaissances en SdL, diverses ontologies se sont appuyées sur des ressources non ontologiques (Tab. 1 ci-dessous). Dans la plupart des cas, la modélisation a concerné un sous-domaine des SdL dont les connaissances sont issues de manuels ou de manuscrits (Ruimy *et al.* 2012, Tremblay 2009)[2]. La consultation des écrits scientifiques a parfois servi d'appui au recours à des ressources lexicographiques (Pareja-Lora 2013, Ruimy *et al.* 2012).

	Produit	**Langage**	**Finalité**	**Ressources de départ**
Tremblay (2009)	ontologie de la lexicologie	*frames* (*Protégé* v. 3.4)	didactique du lexique	LEC
Ruimy *et al.* (2012)	« lexique-thésaurus » de terminologie saussurienne	SIMPLE (OWL)	numérisation et pérennisation	lexiques et manuscrits de Ferdinand de Saussure
Pareja-Lora (2013)	module ontologique sur pragmatique	OWL 2	ontologie modulaire pour représentation données linguistiques	*Glossary of linguistic terms* (SIL) et manuel de pragmatique
Hyvönen *et al.* (2008), Lappalainen *et al.* (2014)	KTO (*Kielitieteen Ontologia*)	*General Finnish Upper Ontology* YSO (SKOS/OWL)	infrastructure nationale pour le Web sémantique	thésaurus d'indexation en SdL

Tab. 1 – *Ontologies en SdL.*

Un cas à part est représenté par KTO[3], à savoir une ontologie multilingue (anglais, allemand, estonien et finnois) issue de la réingénierie d'un thésaurus d'indexation en SdL. Il rejoint ainsi la tendance actuelle de réingénierie de thésaurus pour la création d'ontologies (Amarger *et al.* 2013)[4], d'autant plus

2 Approche synchronique et structuraliste de Saussure (Ruimy *et al.* 2012), formalisation sémantique de la Lexicologie Explicative et Combinatoire (Tremblay 2009).

3 KTO – *Kielitieteen Ontologia* : http://finto.fi/kto/en/index (page consultée le 10/02/2016).

4 La réingénierie des thésaurus a servi pour la création d'ontologies de domaine (Soergel *et al.* 2004), de bases de connaissances pour l'extraction d'autres thésaurus ainsi que d'on-

que la structuration du thésaurus s'est rapprochée de la structuration ontologique, comme le témoigne la norme ISO 25964-1 :2011 (Roche *et al*. 2014).

2. Contexte

Notre article présente une expérience de réingénierie d'un thésaurus en vue de la création d'une ontologie de domaine en SdL. Cette expérience s'inscrit dans le cadre du projet TermITH[5] (**Term**inologie et **I**ndexation de **T**extes en sciences **H**umaines) qui vise à indexer les textes scientifiques en sciences humaines, comme les SdL, à partir des termes contenus dans ces textes, la structuration de termes dans un référentiel pouvant faciliter la tâche d'indexation.

2.1. Courants terminologiques et représentations terminologiques

La structuration des termes en SdL doit trouver sa place à l'intérieur des deux macro-approches en terminologie : l'une évoquant la centralité du *concept* et des connaissances, l'autre du *sens* et du discours. L'*ontoterminologie* (Roche 2007), à la fois héritière et modernisatrice de la Théorie Générale de la Terminologie (TGT, Wüster 1981 [1974], Felber 1987 [1984]), a remis au centre le *concept* et, par conséquent, la représentation ontologique. En revanche, la *terminologie textuelle* (Bourigault & Slodzian 1999, L'Homme 2004) a privilégié le *sens* et la représentation de la variation terminologique en discours sous forme de dictionnaires spécialisés, banques terminologiques ou réseaux terminologiques.

Les deux macro-approches engendrent des représentations non spéculaires (Roche 2006). Il reste pourtant à établir dans quelle mesure cette inconciliabilité entre représentations doit persister.

tologies de domaine (Liang *et al*. 2006) ou pour l'intégration dans une ontologie de haut niveau, comme dans le cas de KTO (Hyvönen *et al*. 2008, Lappalainen *et al*. 2014). En lien avec un vocabulaire contrôlé ou avec l'extraction de connaissances d'un corpus, les contenus des thésaurus ont également fait l'objet respectivement de la création d'un thésaurus « noyau » (Cyrus et Preuss 2009) et d'une ontologie de domaine (Chrisment *et al*. 2008).

5 TermITH – ANR-12-CORD-0029 – ContInt (Contenus numériques et interactions) – http://www.atilf.fr/ressources/termith/ (page consultée le 10/02/2016).

2.2. Consensus des experts en SdL

La structuration des termes en SdL doit également essayer de proposer une conceptualisation valable pour la (quasi-)totalité des linguistes, car une ontologie se veut la « modélisation d'une intersubjectivité » (Roche 2007 : 17). Ainsi, une ontologie de domaine doit dépasser les particularismes théoriques de tel ou tel autre expert ou de telle ou telle autre école afin de proposer un noyau conceptuel partagé.

Le calcul de ce dénominateur conceptuel constitue l'enjeu principal puisque les sous-domaines en SdL, les approches et les théories présentent des positions divergentes en ce qui concerne des axiomes ou des descriptions linguistiques.

2.3. Relation lexique scientifique transdisciplinaire (LST) et terminologie

La tâche de structuration des termes en SdL doit prendre en compte l'interaction entre le LST (Drouin 2007, Hatier 2013, Tutin 2007) et les termes des SdL, aussi.

Le LST est l'ensemble d'unités lexicales – en sciences humaines et sociales (ex. économie, SdL, etc.), mais aussi en sciences appliquées et expérimentales (ex. biologie, médecine, etc.) – se référant aux outils abstraits et concrets ainsi qu'aux phases méthodologiques désignées par les experts (ex. *analyse*, *théorie*, *présenter*, *vérifier*, *méthodologique*, *opératoire*, etc.). Dans le cadre du projet TermITH, le LST a été utilisé en tant que *lexique d'exclusion* pour la désambiguïsation sémantique et en tant que *lexique introducteur*, à savoir marqueur de termes en contexte (Kister et Jacquey 2012, Jacquey *et al*. 2013). Depuis, le LST a fait l'objet d'une mise à jour et d'une classification sémantique (Hatier *et al*. 2014) (ex. classe « objet scientifique » : *notion*, *outil*, *question*).

Pour la conception de notre ontologie, le LST ne peut être négligé. Compte tenu de ses apports sur le plan linguistique, il reste à comprendre dans quelle mesure le LST peut participer à la classification et à la structuration des concepts en SdL.

2.4. Ressources non ontologiques

La structuration des termes en SdL doit également prendre appui sur un ensemble de ressources non ontologiques. Le Tab. 2 ci-dessous résume la nature des ressources linguistiques non ontologiques à notre disposition.

	Type	Unité fonctionnelle	Nb.	Structuration sémantique	Provenance
FRANCIS (SdL)	référentiel	mot-clé	6.088	NON	INIST
Candidats-termes *Scientext*	liste	candidat-terme	1.259	NON	ATILF/ LIDILEM
Thesaulangue	thésaurus	descripteur	872	OUI	ATILF

Tab. 2 – *Ressources linguistiques non ontologiques dans le projet* TermITH.

Les 6.088 descripteurs de FRANCIS[6] ont servi et servent de mots-clés pour l'indexation des documents au sein de l'INIST. La liste des 1.259 candidats-termes résulte de l'extraction automatique à l'aide de TTC-*TermSuite*[7] à partir du corpus *Scientext*[8], notamment des articles, des actes et des thèses en SdL. Ces deux ressources non ontologiques ne présentent aucune structuration linguistique ou conceptuelle.

En revanche, le thésaurus *Thesaulangue*[9] se caractérise par une structuration hiérarchique arborescente. Depuis sa conception dans les années 1970 comme outil d'indexation pour la documentation d'appui au *Trésor de la Langue Française* (TLF, Martin *et al*. 1971), le thésaurus a subi une refonte une dizaine d'années après la fin de l'entreprise TLF (Berthelier et Turcan 2006) et une spécialisation en SdL. Depuis, *Thesaulangue* comprend 20 micro-thésaurus dont les 872 descripteurs et 1.212 non-descripteurs sont organisés sur 6 niveaux hiérarchiques.

3. Méthodologie

Notre expérience a voulu explorer l'hypothèse de structuration d'une ontologie intégrant structuration conceptuelle et description linguistique. Pour ce faire, nous avons eu recours aussi bien à la TGT qu'à la terminologie textuelle.

Suivant l'exemple des ontologies mentionnées au § 1, nous nous sommes concentré sur un sous-domaine des SdL à partir du microthésaurus *Syntaxe*

6 Les variantes sont exclues du chiffre indiqué dans le Tab. 1.
7 TTC-*TermSuite* : http://termsuite.github.io (page consultée le 18/10/2019).
8 *Scientext* : http://scientext.msh-alpes.fr/ (page consultée le 10/02/2016).
9 *Thesaulangue* : http://apps.atilf.fr/revues/thesaulangue.pdf (page consultée le 10/02/2016).

de *Thesaulangue*. Ce dernier couvre environ 21% des descripteurs totaux. De plus, ses 5 rangs hiérarchiques ont permis d'approfondir la logique d'enchâssement des descripteurs. Le nombre relativement réduit de non-descripteurs (à peu près 6% du total) a plutôt donné un aperçu de la cohérence du traitement de la relation d'équivalence dans le thésaurus.

	Micro-thésaurus *Syntaxe*	***Thesaulangue***
Niveaux hiérarchiques	5	$1 \leq x \leq 6$
Descripteurs	186 (~ 21%)	872 (100%)
Non descripteurs	70 (~ 6%)	1.212 (100%)

Tab. 3 – *Comparaison entre* Thesaulangue *et son microthésaurus* Syntaxe.

Comme pour les autres microthésaurus, les descripteurs sont structurés d'après la relation hiérarchique « terme générique (TG) – terme spécifique (TS) ». La relation d'équivalence/synonymie « employé pour » (EP) établit les liens entre les descripteurs et les non descripteurs.

Nous avons validé le statut terminologique d'une désignation et de sa structuration à partir de nos connaissances en tant qu'expert en SdL. De même, nous avons eu recours aux marqueurs linguistiques définitionnels pour repérer les propriétés des concepts en SdL. Nous avons récupéré ces informations à partir de la consultation :

- des définitions encyclopédiques en SdL (Dubois *et al*. 2002 [1994], Ducrot & Schaeffer 1995 [1972], Neveu 2011) – les définitions encyclopédiques étant des contextes riches en connaissances, non pas des définitions terminologiques –,
- de deux ouvrages de référence en syntaxe (Arrivé *et al*. 1986, Riegel *et al*. 2009 [1994]),
- ainsi que du corpus TermITH d'articles annotés en SdL[10].

En nous inspirant des expériences présentées au § 1, nous avons exploité OWL comme langage de représentation. Contrairement aux expériences mentionnées au §1, nous avons spécialisé l'usage des primitives OWL (*classe*, *individu*, *propriété d'objets*, *propriété des données*) d'après l'acquis en TGT (*concept*, *relation logique*, *relation ontologique*) et en terminologie textuelle (désignation terminologique, informations lexico-sémantiques).

10 Plateforme *Smarties* : http://apps.atilf.fr/smarties/ (page consultée le 10/02/2016).

Primitives OWL		Acquis en terminologique et documentation
classe		facette
individu	>>	concept
propriété d'objet		relations logiques et ontologiques
propriété de données		désignation terminologique, définition, information

Tab. 4 – *Correspondance entre primitives OWL et notions en terminologie et en documentation.*

Comme le montre le Tab. 4 ci-dessus, nous avons également intégré la notion de *facette*[11]. La norme ISO 25964-1 :2011 présente la facette comme un « grouping of concepts of the same inherent category » (*Terms and definitions*, 2.20), la définition insistant sur le rapprochement entre thésaurus et ontologie. En même temps, elle souligne la valeur classificatoire autour d'un caractère inhérent partagé par les concepts du domaine à structurer. Ces caractères inhérents sont à repérer, à notre avis, dans les activités, phases et outils épistémologiques communs aux savoirs et désignés par le LST. C'est au niveau de la classification des concepts en SdL que nous avons exploité le LST.

4. Expérience de réingénierie

Grâce à l'éditeur *NeOn Toolkit*[12], nous avons réparti la réingénierie en 4 phases :

- phase 0 : application du patron de réingénierie ;
- phase 1 : implémentation des facettes ; classification des concepts et épuisement des connaissances par désambiguïsation des relations du microthésaurus *Syntaxe* à l'aide des ressources externes ;

11 Une définition moins récente de *facette* est donnée par le *Vocabulaire de la documentation* de l'ADBS : « catégorie non thématique permettant l'organisation d'un ensemble de notions […] en fonction de leur nature propre ou du point de vue […]. Ex. : *phénomène*, *processus*, *propriété*, *outil*, etc. » (http://www.adbs.fr/vocabulaire-de-la-documentation-41820.htm - page consultée le 10/02/2016).

12 *NeOn Toolkit* : http://neon-toolkit.org/ (page consultée le 10/02/2016).

- phase 2 : réagencement de la classification des concepts sous les facettes ; implémentation et établissement des relations entre concepts à l'aide des ressources externes ;
- phase 3 : implémentation de la description linguistique des termes.

4.1. Phase 0

Nous sommes parti du patron de réingénierie de Villazón-Terrazas *et al*. (2009) conçu dans le cadre de la méthodologie *NeOn*. Son application prévoit le statut terminologique de chaque descripteur. Avant son application, nous avons procédé à une validation terminologique manuelle des descripteurs. Sur un total de 186, 185 descripteurs sont des termes désignant des concepts en syntaxe. Seul le descripteur *type de phrase* a été rejeté.

Une fois vérifiée l'adéquation du patron, une arborescence de classes OWL a repris le contenu et la structuration du microthésaurus de façon spéculaire. Par exemple, la classe *nombre* est superordonnée aux deux sous-classes *pluriel* et *singulier*.

4.2. Phase 1

En vue d'appliquer notre modèle de représentation, nous avons créé 6 classes correspondant à 6 facettes dont les désignations appartiennent au LST : *élément*, *ensemble*, *fonction*, *processus*, *propriété*, *statut*. Chaque facette a facilité la première étape de classification des concepts, en s'éloignant au fur et à mesure de l'arborescence de départ.

Le processus de classification a abouti à une arborescence composée de 23 facettes (classes OWL) en Fig. 1.

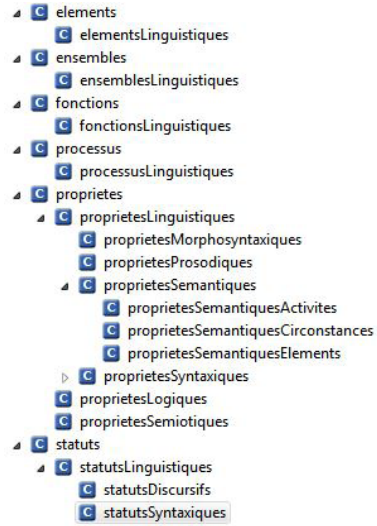


Fig. 1 – *Facettes (classes OWL) pour la réingénierie du micro-thesaurus* Syntaxe.

À ce stade, nous avons progressivement converti les sous-classes correspondant à des concepts en individus OWL.

Cette conversion informatique est allée de pair avec la désambiguïsation de la relation EP. D'après Kister *et al*. (2011), la relation EP dans *Thesaulangue* représente à la fois la relation sémantique *synonymie* et la relation logique *identique à*. Grâce à la consultation des ressources externes, notre analyse des non-descripteurs du microthésaurus *Syntaxe* a montré que cela n'est vrai qu'en partie (Tab. 6). Environ 83% (58 sur 70) des non-descripteurs révèlent que la relation EP traduit plusieurs types de relations logiques et ontologiques. Dit autrement, des concepts autonomes et distingués en syntaxe sont reliés par une relation d'équivalence conceptuelle. Par exemple, dans le cas de :

principale EP *indépendante*

la relation n'est ni d'équivalence conceptuelle ni de synonymie linguistique. Une (proposition) *principale* se distingue en effet d'une (proposition) *indépendante* grâce au caractère /subordination/ car une principale prévoit la présence d'au moins une subordonnée. En revanche, une indépendante n'entretient aucun lien grammatical avec d'autres propositions. Par conséquent,

les deux concepts /propositionPrincipale/ et /propositionIndépendante/ sont coordonnés entre eux et subordonnés au concept /proposition/. Seulement 16% (11 sur 70) des non-descripteurs montrent que la relation EP se situe sur le plan du discours, comme l'illustrent la nominalisation de l'adjectif *numéral* ou l'emploi quasi-synonymique entre *substantif* et *nom*[13].

Relations sémantiques	**Exemple**	**Total**
Quasi-synonymie	*substantif* EP *nom*	6
Synonymie	*adjectif numéral* EP *numéral*	5
		11

Relations conceptuelles	**Exemple**	**Total**
Concepts autonomes {Syntaxe}		
coordination logique	*principale* EP *indépendante*	55
subordination logique	*concession* EP *opposition*	
subordination partitive	*syntaxe* EP *parties du discours*	
processus > résultat	*clivage* EP *phrase clivée*	
Concepts appartenant à d'autres domaines	*discours direct* EP *style direct*	3
		58

Tab. 6 – *Désambiguïsation du traitement de la relation* EP *dans le microthésaurus* Syntaxe.

4.3. Phase 2

La désambiguïsation a engendré un enrichissement en individus de l'ontologie ainsi que la poursuite de leur réagencement dans les 23 facettes.

Une fois les connaissances du microthésaurus décryptées et épuisées, nous avons implémenté les relations conceptuelles recensées par la TGT (Felber 1987 [1984] : 101-108) et par la terminologie textuelle (L'Homme 2004 : 156) (*propriété d'objet* OWL). Plus précisément, nous avons retenu 9 relations con-

13 Une seule relation EP est un cas de variante orthographique : *coréférence* EP co-*référence*.

ceptuelles, à savoir 14 propriétés d'objet[14]. À ces relations, nous avons ajouté la relation ontologique *combinaison* pour exprimer la relation de dépendance syntaxique, et ce, en guise de test.

La vérification manuelle des relations encodées dans le microthésaurus a permis l'établissement des relations conceptuelles dans l'ontologie. Cette étape s'est appuyée sur les définitions encyclopédiques et sur les ouvrages de référence choisis (§ 3). Par exemple, le concept /conjonction/ est lié aux concepts /conjonctionDeCoordination/ et /conjonctionDeSubordination/par la relation de subordination logique, les derniers étant deux espèces du premier. La relation d'intersection partitive permet de rendre compte du lien entre le concept /morphosyntaxe/ et deux autres concepts – /morphologie/ et /syntaxe/ –, le premier partageant des caractères avec les derniers. En revanche, le concept /aspect/ est lié au concept /verbe/ par la relation d'instrumentation puisque le verbe est la partie du discours plus concernée par l'/aspect/, c'est-à-dire par la « catégorie grammaticale qui exprime la représentation que se fait le sujet parlant » (Dubois *et al*. 2002 [1994] : 53, s.v. *aspect*).

Relations	Propriété d'objet (OWL)	Exemple
Subordination logique	*hasSubordinate* *isSubordinatedTo*	/conjonction/ *hasSubordinate* /conjonctionDeCoordination/
Coordination logique	*isCoordinatedTo*	/genreFéminin/ *isCoordinatedTo* /genreMasculin/
Relation logique diagonale	*hasDiagonalRelationshipWith*	/propositionSubordonnéeFinale/ *hasDiagonalRelationshipWith* /propositionPrincipale/
Intersection logique	*hasLogicalIntersectionWith*	/adjectifQualificatif/ *hasLogicalIntersectionWith* /adjectifCaractérisant/
Détermination	*hasDetermined* *isDeterminedBy*	/adjectifSubstantivé/ *isDeterminedBy* /nom/

14 Nous avons ajouté 5 propriétés relevant de l'explication des relations converses de la détermination, de l'intersection, de la subordination logique, de la subordination partitive et de la succession.

Subordination partitive	*hasPartitiveSubordinate* *hasComprising*	/syntagmeAdjectival/ *hasPartitiveSubordinate* / déterminantDéfini/
Intersection partitive	*hasPartitiveIntersectionWith*	/morphosyntaxe/ *hasPartitiveIntersectionWith* /morphologie/
Intersection	*hasInstrument* *isInstrumentTo*	/aspect/ *hasInstrument* /verbe/
Succession	*hasSuccessor* *hasPredecessor*	/adjectifDéterminatif/ *hasSuccessor* /déterminant/
Combinaison	*combinesWith*	/adjectifpossessif/ *combinesWith* /nom/

Tab. 8 – *Propriétés d'objet pour le traitement des concepts.*

4.4. Phase 3

La dernière phase de réingénierie a porté sur les termes et sur d'autres informations linguistiques et documentaires (*propriété des données* OWL). Outre la consultation des études en terminologie textuelle, nous avons repris quelques catégories de données et d'attributs de la norme ISO 30042 : 2008 *TermBase eXchange* (TBX), quitte à les réadapter. Nous avons implémenté 16 propriétés de données qui correspondent à 9 catégories et attributs TBX.

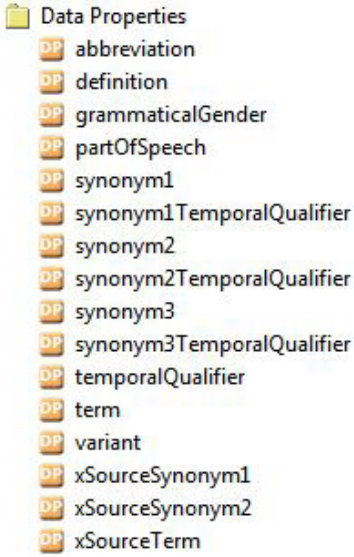


Fig. 2 – *Propriétés des données pour le traitement linguistique des concepts.*

La Fig. 2 montre l'ensemble des propriétés des données qui reflètent grosso modo la structure d'une fiche terminologique. Outre la désignation terminologique (*term*) et ses variantes graphiques (*abbreviation*, *variant*), on trouve des informations grammaticales (*grammaticalGender*, *partOfSpeech*), la définition (*definition*), la relation sémantique de synonymie (*synonym1*, *synonym2*, *synonym3*) ainsi que la marque d'usage – notamment de désuétude – du terme (*temporalQualifier*) et des synonymes (*synonym1TemporalQualifier*, *synonym2TemporalQualifier*, *synonym3TemporalQualifier*)[15]. La précision de la marque d'usage est surtout à envisager en fonction de la tâche d'indexation, l'ontologie devant gérer les diverses désignations conceptuelles.

En ce qui concerne les aspects documentaires, nous avons précisé la provenance des désignations terminologiques (*xSourceTerm*) et des synonymes éventuels (*xSourceSynonym1*, *xSourceSynonym2*). Plus précisément, nous avons repéré les désignations terminologiques et leurs synonymes dans l'ensemble des ressources non ontologiques TermITH : les corpus TermITH

15 Les informations grammaticales (*grammaticalGender*, *partOfSpeech*) et les marques d'usage (*temporalQualifier*) prennent des valeurs en anglais issues de la TBX.

(CTH) et *Scientext* (CTS) ainsi que le référentiel FRANCIS (INIST) et *Thesaulangue* (TLG).

5. Discussion

Au bout de la réingénierie, chaque concept en syntaxe devrait posséder une description qui ressemble à celle en Fig. 3. Nous utilisons la description du concept /déterminant/ pour montrer les pour et les contre de notre démarche de réingéniérie. Le concept /déterminant/ (individu OWL *determinants*) appartient à la facette *éléments linguistiques* (classe OWL *elementsLinguistiques*). La description conceptuelle et linguistique de /déterminant/ se trouve sous la rubrique *Facts*. La rubrique précise d'abord les diverses relations (propriétés d'objet OWL) que le concept /déterminant/ entretient avec d'autres concepts en syntaxe. Par exemple, /déterminant/ est superordonné à /article/ (*articles*) et /déterminantPossessif/ (*adjectifPossessif*), les articles (ex. *le*, *un*) et les déterminants possessifs (ex. *mon*, *notre*) étant des espèces de déterminants. Le concept /déterminant/ est aussi dans une relation de subordination partitive avec /syntagmeNominal/ (*syntagmeNominal*), un déterminant pouvant faire partie d'un syntagme nominal (ex. *mon chien*) et, par conséquent, dans une relation de combinaison avec /nom/ (*noms*). Par la suite, la rubrique *Facts* énumère les informations linguistiques (propriétés de données OWL) de /déterminant/ triées par ordre alphabétique. La désignation terminologique *déterminant* (propriété de données *term*) repérée dans les corpus TermITH et *Scientext* ainsi que dans le référentiel FRANCIS (propriété de données *xSourceTerm*), renvoie au concept /déterminant/. Plus précisément, *déterminant* est la désignation préférée à celle quasi-synonymique en discours d'*adjectif démonstratif* (propriété de données *synonym1*). Cette dernière est accompagnée d'une marque d'usage (propriété de données *synonym1TemporalQualifier*) soulignant qu'elle est désuète (valeur *outdatedTerm*).

```
Individual: determinants

  Types:
    elementsLinguistiques

  Facts:
    combinesWith noms,
    hasComprising syntagmeNominal,
    hasLogicalIntersectionWith adjectifsDeterminatifs,
    hasPredecessor adjectifsDeterminatifs,
    hasSubordinate adjectifDemonstratif,
    hasSubordinate adjectifExclamatif,
    hasSubordinate adjectifInterrogatif,
    hasSubordinate adjectifNumeralCardinal,
    hasSubordinate adjectifNumeralOrdinal,
    hasSubordinate adjectifPossessif,
    hasSubordinate adjectifsIndefinis,
    hasSubordinate adjectifsNumeraux,
    hasSubordinate articleDefini,
    hasSubordinate articleIndefini,
    hasSubordinate articlePartitif,
    hasSubordinate articles,
    isInstrumentTo referenceSemantique,
    isInstrumentTo referenceSyntaxique,
    definition "Elément grammatical qui précède et spécifie un nom dans un syntagme nominal
"@fr,
    grammaticalGender "masculine"@en,
    partOfSpeech "noun"@en,
    synonym1 "adjectif détérminatif"@fr,
    synonym1TemporalQualifier "outdatedTerm"@en,
    term "déterminant"@fr,
    xSourceSynonym1 "Dubois et al. (2002 : 140)"@fr,
    xSourceTerm "CTS, INIST, TLG"@fr
```

Fig. 3 – *Exemple de description du concept* /déterminant/ *dans l'ontologie.*

La quasi-synonymie sur le plan linguistique ne va pourtant pas de pair avec les concepts auxquels les désignations renvoient. Les deux désignations relèvent de deux conceptualisations de la fonction syntaxique du /déterminant/ : l'une (Arrivé *et al*. 1986, Dubois *et al*. 2002 : 140) a attribué à /déterminant/ des caractères le rapprochant du concept /adjectif/ ; l'autre a distingué clairement le /déterminant/ de l'/adjectif/ (Riegel *et al*. 2009 [1994] : 276), ce qui a engendré une nouvelle classification des espèces de /déterminant/. À cet égard, le microthésaurus *Syntaxe* révèle une conceptualisation hybride incohérente. D'une part, elle distingue les deux descripteurs *adjectif* et *déterminant*, suivant la deuxième conceptualisation. D'autre part, le descripteur *déterminant* entretient une relation TG – TS avec les descripteurs *adjectif démonstratif*, *adjectif indéfini*, *adjectif interrogatif*, *adjectif possessif* et *numéral* qui renvoient plutôt à la première conceptualisation. Comme le mon-

tre la Fig. 4, aucun autre non-descripteur n'est mentionné, excepté le synonyme *adjectif numéral*.

Fig. 4 – *Extrait du microthésaurus* Syntaxe *pour le descripteur* déterminant.

Il est évident que les deux conceptualisations nécessitent d'être mentionnées et d'interagir entre elles, vu la tâche d'indexation et la nature intersubjective d'une représentation ontologique. Dans des cas pareils, le nombre et la typologie des ressources sur lesquelles le concepteur s'appuie pour structurer son ontologie acquièrent toute leur importance. Le recours à une seule ressource – dictionnairique ou non – exclut des conceptualisations concurrentes et/ou divergentes. La dialectique théorique interne à un domaine se réduit ainsi à la vision du domaine portée par la ressource choisie. De plus, la vision synchronique de l'ontologie aplatit la dynamique de l'évolution conceptuelle à un moment donné qui correspond, en général, au moment le plus proche de la conception. Autrement dit, le concepteur tend à privilégier la conceptualisation la plus récente au moment où il structure les connaissances, sans tenir compte de l'évolution diachronique des concepts manipulés.

À ce stade, le concepteur peut soit représenter l'une ou l'autre conceptualisation soit harmoniser les divergences conceptuelles. Comme le concepteur n'est pas, en principe, un expert du domaine à représenter, il tend à faire con-

fiance aux ressources, sans remise en question de leurs contenus. Une telle démarche – mitigée, certes, par le recours aux experts du domaine – va, à notre avis, à l'encontre de la définition d'ontologie en tant que représentation intersubjective des connaissances (§ 1). Dans notre cas de figure, le concepteur et l'expert du domaine coïncident. Nous avons remis systématiquement en question les descripteurs de *Thesaulangue*. Cette démarche critique vis-à-vis de notre ressource non ontologique a fait l'objet d'une analyse ponctuelle des concepts et des désignations. Parallèlement à l'implémentation sous *NeOn Toolkit*, nous avons gardé une trace de chaque phase de réingénierie et de (re) structuration conceptuelle dans un fichier de texte indépendant. Cette trace documentaire nous a également permis de revenir en arrière sur nos choix, mais surtout de les motiver.

Pour revenir à la description du concept /déterminant/, donc, notre analyse nous a conduit au maintien des deux concepts /adjectifDéterminatif/ et /déterminant/. Comme le montre la Fig. 3, les deux concepts sont reliés par la relation d'intersection logique ainsi que par la relation de succession. La première s'explique par le fait que /adjectifDéterminatif/ et /déterminant/ partagent des caractères et se chevauchent, sans pour autant être équivalents[16]. La deuxième

16 D'après Ducrot & Schaeffer (1995 [1972]), la fonction de *déterminant* entre en SdL à partir de la grammaire de Port-Royal. Elle se stabilise et se spécialise en grammaire générative (Dubois *et al*. 2002 [1994] :140). Sous les déterminants, le dictionnaire de Dubois *et al*. (*ibid*.) regroupe les articles (définis), les possessifs, les démonstratifs, les adjectifs interrogatifs, les relatifs, les indéfinis et les numéraux. Excepté les articles, tous les autres sont classés sous les *adjectifs déterminatifs* par opposition aux *adjectifs qualificatifs*. En même temps, Dubois *et al*. (*ibid*.) précisent que la désignation *déterminant* peut également renvoyer au « constituant du syntagme nominal qui dépend du nom, tête ou constituant principal du syntagme nominal. En ce sens, les déterminants sont les articles, les adjectifs, les compléments du nom ». En guise de synthèse, on remarque que la définition encyclopédique de Dubois *et al*. (2002) renvoie à deux conceptualisations qui emploient la désignation *déterminant*, les deux partageant le concept /adjectif/. En revanche, la conceptualisation de Riegel *et al*. (2009 [1994]) sépare les adjectifs des déterminants, tout en rejetant la désignation *adjectif déterminatif*. Il s'ensuit que le déterminant est « le mot qui doit nécessairement précéder un nom commun pour constituer un groupe nominal bien formé » (Riegel *et al*. 2009 [1994] : 276). Tout en nous passant des détails sur la distinction entre *déterminants définis* et *déterminants indéfinis* qui réorganise les espèces de /déterminant/ grâce au concept /référence/, ainsi que sur les diverses désignations utilisées, nous remarquons que les conceptualisations de /déterminant/ de Dubois *et al*. (2002 [1994]) et Riegel *et al*. (2009 [1994]) se rapprochent dans la mesure où l'on se tient à la relation de subordination partitive avec /groupeNominal/ et à la relation de combinaison avec /nom/. D'où l'établissement de la relation d'intersection logique entre /adjectifDéterminatif/ et /déterminant/ dans notre représentation. De plus, une partie de la compréhension d'/adjectifDéterminatif/ chez Dubois *et al*. (2002 [1994] :140) a été reprise par /déterminant/

reflète la distance temporelle des conceptualisations (pas de leurs désignations), celle d'/adjectifDéterminatif/ précédant celle de /déterminant/[17].

Une conceptualisation harmonisatrice des connaissances demande une sélection représentative des diverses approches théoriques du domaine. Elle doit prévoir, d'une part, l'établissement de cartes conceptuelles analytiques pour chaque approche et, d'autre part, la mise au point d'une synthèse conceptuelle. Quoique la réingénierie porte sur un seul sous-domaine des SdL, un tel projet implique un travail chronophage. Compte tenu de nos soucis de précision et d'exhaustivité, la conception a été sensiblement ralentie. À l'heure où nous écrivons, la réingénierie du microthésaurus *Syntaxe* de *Thesaulangue* a abouti à une ontologie OWL composée de :

- 23 facettes (23 classes OWL) ;
- 331 concepts en syntaxe (331 individus OWL) issus des 185 descripteurs validés ainsi que de la désambiguïsation des 70 relations encodées dans le microthésaurus et de la consultation des ressources externes ;
- 10 relations entre concepts (correspondant à 14 propriétés d'objet OWL) ;
- 9 catégories et attributs tirés de la norme TBX pour la description linguistique (implémentés sous 16 propriétés de données OWL).

dont la finalisation a subi un arrêt pour répondre aux besoins contingents de TermITH. En effet, nous nous sommes tourné vers la conception d'un thésaurus ontoterminologique (Roche *et al*. 2014) en SdL suivant la norme ISO 25964-1 : 2011 à partir de l'intégralité des ressources non ontologiques (§ 2)[18].

6. Conclusions

Cet article a présenté une expérience de réingénierie de thésaurus en vue de la conception d'une ontologie OWL. Nous avons proposé un modèle – ambitieux, certes – de réingénierie pouvant concilier les divergences des cou-

chez Riegel *et al*. (2009 [1994] : 276), et représente un ensemble de caractères conceptuels partagés par /adjectifDéterminatif/ et /déterminant/.

17 Comme nous l'avons précisé (§ note 16), chez Dubois *et al*. (2002 [1994] :140), la compréhension d'/adjectifDéterminatif/ diffère de celle de /déterminant/ proposée par la suite par Riegel *et al*. (2009 [1994] : 276).

18 Le thésaurus contient 1.719 descripteurs et 713 non-descripteurs. Il est actuellement en phase de test au Centre de documentation de l'ATILF.

rants terminologiques et les contraintes pragmatiques liées à la sémantique du langage OWL et aux attentes du projet TermITH.

Notre expérience a montré que les ressources non ontologiques peuvent présenter des connaissances partielles, parfois dépassées, d'un domaine. D'où l'importance d'une sélection représentative de ressources externes à consulter pour évaluer et valider les contenus des ressources non ontologiques. Dans notre cas, nous soulignons, par exemple, l'absence – volontaire, vu le cadre expérimental – de concepts venant de la macrosyntaxe et du traitement informatique de la syntaxe. Nous avons également illustré que le décryptage de la sémantique régissant une ressource non ontologique constitue une étape indispensable pour envisager une structuration ontologique et que l'utilisation des patrons de réingénierie doit aller de pair avec une évaluation critique constante de la part du concepteur. Cette démarche d'interrogation des ressources non ontologiques doit avoir lieu tout au long de la structuration de l'ontologie pour que cette dernière reflète la richesse conceptuelle et désignationnelle d'un domaine, et ce, afin d'aboutir à une représentation réellement intersubjective des connaissances d'un domaine.

Il est évident qu'une telle structuration demande un effort d'analyse et de synthèse harmonisatrice. La recherche d'un équilibre entre gain et perte de connaissances acquiert un poids important, de même que la recherche d'un équilibre entre description conceptuelle et description linguistique. À ce propos, il est nécessaire de disposer d'une documentation retraçant l'acheminement vers la structuration ontologique, souvent lacunaire dans les structurations d'ontologies. La transparence des choix motivés par le concepteur est, à notre avis, l'une des garanties de l'interopérabilité entre ressources ontologiques et ressources non ontologiques, à côté de la réflexion, à juste titre, sur les standards informatiques.

Comme nous l'avons précisé, le processus de réingénierie de *Thesaulangue* a laissé la place à la conception d'un thésaurus en SdL pour répondre aux exigences du projet TermITH. Ce thésaurus constitue probablement la ressource non ontologique de départ la plus adéquate en vue de la conception d'une ontologie de domaine. Entre-temps, il sera opportun de reconsidérer l'architecture de l'ontologie à venir en fonction des caractéristiques du nouveau thésaurus ainsi que des besoins d'indexation en SdL.

Références

Amarger, F., Roussey, C., Chanet, J.-P., Haemmerlé, O. and N. Hernandez. (2013). « État de l'art : Extraction d'information à partir de thésaurus pour générer une ontologie ». In *Actes INFORSID 2013*, Paris, 29-31 mai 2013. Disponible en ligne : http://inforsid.fr/actes/2013/2013_1a_1%20Amarger.pdf (page consultée le 10/02/2016).

Arrivé, M., Gadet, F. et M. Galmiche. (1986). *La grammaire d'aujourd'hui. Guide alphabétique de linguistique française*. Paris : Flammarion.

Berthelier, V. et I. Turcan. (2006). « *Thesaulangue*. Le thesaurus de linguistique française du laboratoire ATILF ». *Revue française de linguistique appliquée* 11 (1), 139-143.

Boulogne, A. (2004). *Vocabulaire de la documentation*. Paris : ADBS éditions.

Bourigault, D. et M. Slodzian. (1999). « Pour une terminologie textuelle ». *Terminologies nouvelles* 19, 29-32.

Chrisment, C., Haemmerlé, O., Hernandez, N. et J. Mothe. (2008). « Méthodologie de transformation d'un thesaurus en une ontologie de domaine ». *Revue d'Intelligence Artificielle* 22 (1), 7-37.

Cyrot, C. et C. Preuss. (2009). « Réingénierie de thesaurus : une étude de cas ». *Documentaliste – Sciences de l'Information* 46 (3), 4-13. Disponible en ligne : http://www.cairn.info/revue-documentaliste-sciences-de-l-information-2009-3-page-4.htm (page consultée le 10/02/2016).

Drouin, P. (2007). « Identification automatique du lexique scientifique transdisciplinaire ». *Revue française de linguistique appliquée* 12 (2), 45-64.

Dubois, J., Giacomo, M., Guespin, L., Marcellesi, C., Marcellesi, J.-B. et J.-P. Mével. (2002 [1994]). *Dictionnaire de linguistique*. Paris : Larousse.

Ducrot, O. et J.-M. Schaeffer. (1995 [1972]). *Nouveau dictionnaire encyclopédique des sciences du langage*. Paris : Seuil.

Felber, H. (1987 [1984]). *Manuel de terminologie*. Paris : UNESCO/Infoterm. Disponible en ligne : http://unesdoc.unesco.org (page consultée le 10/02/2016).

Hatier, S. (2013). « Extraction des mots simples du lexique scientifique transdisciplinaire dans les écrits des sciences humaines : une première expérimentation ». In Boudin, F et Barrault, L. (éds). *Actes de la conférence TALN-RECITAL 2013*, 138-149. Disponible en ligne : http://aclweb.org/anthology/F13-5011 (page consultée le 10/02/2016).

Hatier, S., Tutin, A., Jacques, M.-P., Jacquey, E. et L. Kister. (2014). « Catégorisation sémantique des noms simple du lexique scientifique trans-

disciplinaire ». Communication présentée dans le cadre du *82e Congrès de l'ACFAS – Colloque 330 – Étude de lexiques à vocation particulière : approches théoriques, méthodologiques, pédagogiques et multidisciplinaires*, Montréal, 14-15 mai 2014.

Hyvönen, E., Viljanen, K., Tuominen, J. et K. Seppälä. (2008). « Building a National Semantic Web Ontology and Ontology Service Infrastructure — The FinnONTO Approach ». In Bechhofer, S., Hauswirth, M., Hoffman, J. et Koubarakis, M. (éds). *The Semantic Web : Research and Applications. Proceedings of 5th European Semantic Web Conference, ESWC 2008*. Berlin : Springer, 95-108.

ISO 25964-1 :2011. *Information et documentation – Thésaurus et interopérabilité avec d'autres vocabulaires – Partie 1 : Thésaurus pour la recherche documentaire*. Genève : Organisation internationale de normalisation.

ISO 30042 :2008. *Systems to manage terminology, knowledge, and content – TermBase eXchange* (TBX). Genève : Organisation internationale de normalisation.

Jacquey, E., Tutin, A., Kister, L., Jacques, M.-P., Hatier, S. et S. Ollinger. (2013). « Filtrage terminologique par le lexique transdisciplinaire scientifique : une expérimentation en sciences humaines ». In Aguado de Cea, G. et Aussenac-Gilles, N. (éds). *Actes TIA 2013*, Paris Villetaneuse : IRIT, 121-128.

Kister, L. et E. Jacquey. (2012). « Relations syntaxiques entre lexiques terminologique et transdisciplinaire : analyse en texte intégral ». In Neveu, F., Muni Toke, V., Blumenthal, P., Klingler, T., Ligas, P., Prévost, S., et Teston-Bonnard, S. (éds.). *Actes CMLF 2012*. Paris : EDPSciences, 909-919.

Kister, L., Jacquey, E. et B. Gaiffe. (2011). « Du thesaurus à l'onto-terminologie : relations sémantiques vs. relations ontologiques ». *Corela* 9 (1). Disponible en ligne : http://corela.revues.org/1962 (page consultée le 10/02/2016).

KTO – *Kielitieteen Ontologia* : http://finto.fi/kto/en/index (page consultée le 10/02/2016).

L'Homme, M.-C. (2004). *La terminologie : principes et techniques*. Montréal : Presses de l'Université de Montréal.

Lappalainen, M., Frosterus, M. et S. Nykyri. (2014). « Reuse of library thesaurus data as ontologies for the public sector ». *IFLA WLIC 2014 – Lyon – Libraries, Citizens, Societies : Confluence for Knowledge, Session 86 – Cataloguing with Bibliography, Classification Indexing and UNIMARC*

Strategic Programme. Disponible en ligne : http://library.ifla.org/819/ (page consultée le 10/02/2016).

Liang, A. C., Bauser, L., Sini, M., Keizer, J. et S. Katz. (2006). « From AGROVOC to the Agricultural Ontology Service / Concept Server. An OWL model for managing ontologies in the agricultural domain ». In Cuenca Grau, B., Hitzler, P., Shankey, C. et Wallace, E. (éds). *Proceedings of the OWLED*06 Workshop on OWL : Experiences and Directions*. CEUR-WS.org.

Martin, R., Vienney, R., et E. Martin. (1971). « La mécanisation de la documentation bibliographique au Centre de recherche pour un *Trésor de la langue française* ». *Le français moderne* 39 (2), 181-190.

NeOn ToolKit : http://neon-toolkit.org/ (page consultée le 10/02/2016).

Neveu, F. (2011). *Dictionnaire des sciences du langage*. Paris : Colin.

Ontology Design Patterns : http://ontologydesignpatterns.org/ (page consultée le 10/02/2016).

Pareja-Lora, A. (2013). « An Ontology-Driven Methodology to Reuse, Link and Merge Terminological and Language Resources ». In Aguado de Cea, G. et Aussenac-Gilles, N. (éds). *Actes TIA 2013*, Paris Villetaneuse : IRIT, 189-196.

Riegel, M., Pellat, J.-C. et R. Rioul. (2009 [1994]). *Grammaire méthodique du français*. Paris : PUF.

Roche, C., Damas, L., Roche, J. (2014). Multilingual Thesaurus : The Ontoterminology Approach. In *Proceedings of CIDOC 2014 – Access and Understanding – Networking in the Digital Era*.

Roche, C. (2007). « Le terme et le concept : fondement d'une ontoterminologie ». In *Actes TOTh 2007*. Annecy/Chambéry : Institut Porphyre, 1-22.

Roche, C. (2006). « Lexical and Conceptual Structures in Ontology ». In Ali, M. et Dapoigny, R. (éds). *Advances in Applied Artificial Intelligence 19th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Berlin : Springer, 1034-1041.

Roche, C. (2005). « Terminologie et ontologie ». *Langages* 164, 48-62.

Ruimy, N., Piccini, S. et E. Giovannetti. (2012). « Defining and Structuring Saussure's Terminology ». In Vatvedt Fjeld, R. et Torjusen, J. M. (éds). *Proceedings of the 15th EURALEX international Congress 2012*. Oslo : UiO, 828-833.

Scientext : https://scientext.hypotheses.org (page consultée le 18/10/2020).

Smarties : http://apps.atilf.fr/smarties/ (page consultée le 10/02/2016).

Soergel, D., Lauser, B., Liang, A., Fisseha, F., Keizer, J. et S. Katz. (2004). « Reengineering Thesauri for New Applications : the AGROVOC

Example ». Disponible en ligne : http://www.fao.org/3/af234e/af234e00.htm (page consultée le 18/10/2020).

TermITH – *Terminologie et Indexation de Textes en Sciences Humaines* : http://web.atilf.fr/ressources/termith/corpus.php (page consultée le 18/10/2020).

Thesaulangue : http://apps.atilf.fr/revues/thesaulangue.pdf (page consultée le 10/02/2016).

Tremblay, O. (2009). *Une ontologie des savoirs lexicologiques pour l'élaboration d'un module de cours en didactique du lexique*. Thèse de doctorat. Université de Montréal.

TTC-*TermSuite* : http://termsuite.github.io (page consultée le 18/10/2020).

Tutin, A. (2007). « Autour du lexique et de la phraséologie des écrits scientifiques ». *Revue française de linguistique appliquée* 12 (2), 5-14.

Villazón-Terrazas, B., Suárez-Figueroa, M. C. et A. Gómez-Pérez. (2009). « Pattern for Re-engineering a Term-based Thesaurus, Which Follows the Record-based Model, to a Lightweight Ontology ». In *Proceedings of the Workshop on Ontology Patterns (WOP 2009)*, Blomqvist, E., Sandkuhl, K., Scharffe, F., Svatek, V. (éds). CEUR-WS.org, 108-111.

Wüster, E. (1981). « L'étude scientifique générale de la terminologie. Zone frontalière entre la linguistique, la logique, l'ontologie, l'informatique et les sciences de choses ». In Rondeau, G., Felber, H. et Siforov, V. I. (éds). *Textes choisis de terminologie. I. Fondements théoriques de la terminologie.* Laval : Groupe interdisciplinaire de recherche scientifique et appliquée en terminologie (GIRSTERM), Université de Laval, 57-114.

Summary

The paper presents some methodological issues when re-engineering a thesaurus. More precisely, our experience is based on *Thesaulangue*, i.e. the thesaurus used for indexing documents in Linguistics at the ATILF's documentation centre. The paper describes the steps followed when re-engineering *Thesaulangue* in order to create an OWL ontology. In conclusion, the paper discusses the pros and cons.

Besoins de dénomination et classification dans les nomenclatures botaniques : termes et concepts en action

Claudio Grimaldi

claudio.grimaldi@uniparthenope.it

Résumé. Le long de l'histoire, notamment de la Renaissance à l'âge des Lumières, les objectifs de la botanique évoluent progressivement et s'orientent non plus seulement vers la description, mais aussi vers la classification et la création des nomenclatures voulant ranger les êtres du règne végétal dans des systèmes de classification artificiels ou naturels. Dans cette contribution nous proposons une réflexion concernant les systèmes de classification et nomenclature de la période classique afin d'analyser les nécessités des botanistes de l'époque, ainsi que les stratégies linguistiques retenues pour créer des dénominations adéquates pour ranger les êtres du règne végétal de manière univoque et sans ambiguïtés. En raison de leur valeur historique nous nous attarderons en particulier sur les nomenclatures proposées par Tournefort et Linné.

1. Introduction

La période qui va de la Renaissance à l'âge des Lumières correspond au moment de la disciplinarisation des sciences du vivant et en particulier au passage de la *Res Herbaria*, à savoir la discipline des plantes étudiée généralement par les médecins pour guérir les malades, à la Botanique, s'intéressant au règne végétal en entier, sans aucun rapport spécifique des plantes avec la médecine. Au fil de ces siècles les objectifs des botanistes évoluent progressivement et s'orientent non plus seulement vers la description, mais aussi vers la classification et la création des nomenclatures voulant ranger les êtres du règne végétal dans des systèmes de classification artificiels ou naturels.

D'un point de vue linguistique la naissance et l'évolution des classifications et nomenclatures de la botanique constituent des cas d'étude très intéressants

qui permettent de saisir les principes suivis dans ces systèmes d'arrangement des êtres vivants et de réfléchir sur les nécessités communicatives ressenties au sein des discours scientifiques de l'époque. En effet, pour le dire avec Brunot, « la langue des classifications botaniques fournissait pour la première fois un exemple d'un vocabulaire entièrement fabriqué, c'est-à-dire constitué logiquement suivant des principes arrêtés d'avance, formés de mots clairs, sonores, faciles à distinguer les uns des autres, aussi simples que pouvait le permettre l'expression de réalités qui ne l'étaient pas, et susceptibles enfin de se multiplier ou de se transformer par des applications nouvelles des mêmes principes, à mesure qu'il faudrait exprimer de nouvelles conceptions[1] ».

Dans cette contribution nous proposons une réflexion concernant les systèmes classificatoires et nomenclaturaux de la période classique afin d'analyser les nécessités des botanistes de l'époque, ainsi que les stratégies linguistiques retenues pour créer des dénominations adéquates pour ranger les êtres du règne végétal de manière univoque et sans ambiguïtés. En raison de leur valeur historique nous nous attarderons en particulier sur les nomenclatures proposées par Tournefort et Linné.

2. De la représentation à la classification : les nomenclatures botaniques et la naissance d'une terminologie structurée

L'utilisation de l'approche mathématique et expérimentale[2] au monde caractérise la naissance de l'idéal de science moderne qui marque la période

1 F. Brunot, *Histoire de la langue française. De l'origine à nos jours*, Paris, Armand Colin, tome VI, 1966, p. 616.

2 L'évolution de la notion d'expérience est bien évidente dans les dictionnaires et encyclopédies des XVII^e^ et XVIII^e^ siècles, où le mot « expérience » se charge de plusieurs signifiés nouveaux qui témoignent de l'évolution de la pratique scientifique de l'expérience même. Le mot « expérience », qui renvoie à une « épreuve qu'on fait de quelque chose » et à la « connaissance des choses acquises par un long usage » (*Dictionnaire de l'Académie française* de 1694), devient l'élément central de la pratique scientifique moderne puisque « pour raisonner juste sur les expériences et en tirer de l'utilité, il faut les suivre avec beaucoup d'application et répéter souvent les mêmes pour voir si elles ne varient point » (*Dictionnaire universel* de Furetière, édition de 1727, revue et augmentée par Basnage de Beauval et Brutel de La Rivière). D'après le *Trévoux*, l'expérience est « l'observation exacte des faits et des phénomènes que nous présente la nature, et de ceux que nous créons nous-mêmes, par de nouvelles combinaisons des corps, pour découvrir la cause de ces différents effets » et elles sont devenues si célèbres que « le mot d'*expérience* est devenu comme un terme d'art en Physique, et en plusieurs parties des Mathématiques ». Quant à l'*Encyclopédie*, Diderot indique que l'existence fournit l'épreuve d'une telle chose puisque « nous nous assurons par l'*épreuve*, si la chose a la qualité que nous lui croyons ; par l'*es-*

qui va du XVI^e siècle au XVIII^e siècle, où la science antique est progressivement assimilée de manière critique afin de donner naissance à une culture renouvelée qui conserve l'importance du passé. Le but est de produire de nouveaux principes et de nouvelles théories scientifiques qui peuvent satisfaire aussi bien l'évolution de la connaissance humaine que les besoins sociaux.

Au cours des XVI^e, XVII^e et XVIII^e siècles les sciences naturelles font l'objet de plusieurs recherches et travaux[3] qui conduisent à une nouvelle conception de la science de la botanique, qui pendant longtemps avait reflété un plaisir plutôt superficiel stimulé par la pratique courante et ancienne d'utiliser des herbes pour soigner les maladies humaines. Les plantes acquièrent de plus en plus un intérêt scientifique en quelque sorte désintéressé et spéculatif dépassant les compilations des découvertes pharmaceutiques ou les énumérations des plantes et de leurs propriétés sans souci de compréhension ou de description exacte du mécanisme végétal[4].

Toutefois, la connaissance du règne végétal et son usage ne sont pas allés de pair avec la création d'une nomenclature univoque, et trouver un cadre de référence où ranger les faits innombrables accumulés par les voyageurs et les naturalistes, ce sera l'œuvre notamment du XVIII^e siècle. De ce point de vue, la botanique prend progressivement la forme d'une description, d'une classification et d'une nomenclature, comme l'indique Magnin-Gonze[5], toutes trois strictement codifiées et reliées, dont les objectifs descriptifs, classifica-

sai, quelles sont ses qualités ; par l'*expérience*, si elle est » (entrée « épreuve, essai, expérience »).

3 Parmi les sciences naturelles, la minéralogie se développe en premier : les ouvrages sur les fossiles sont nombreux au XVI^e siècle, des premières tentatives de classification des minéraux s'esquissent et servent de base pour tous les ouvrages des siècles suivants. En France c'est notamment Palissy qui s'intéresse en premier aux connaissances minéralogiques, en soutenant que les coquilles fossiles trouvées sur les collines ne sont que des débris d'animaux marins. Cette idée fascinante ouvre la voie à une longue série de réflexions sur l'histoire du monde et de ses habitants, ainsi qu'à plusieurs recherches sur les êtres qui peuplent la planète. Cf., quant aux nomenclatures des sciences modernes, Ph. Selosse, « Entre Renaissance et Lumières : les nomenclatures des sciences nouvelles », F. Werner, B. Thörle (éds.), *Manuel des langues de spécialités*, Berlin, De Gruyter Mouton, 2016, p. 413-430. En ce qui concerne la naissance des terminologies de certaines sciences au cours des XVII^e et XVIII^e siècles, cf. C. Grimaldi, *Discours et terminologie dans la presse scientifique française (1699-1740). La construction des lexiques de la botanique et de la chimie*, Oxford, Peter Lang, 2017 ; M. T. Zanola, *Arts et métiers au XVIII^e siècle. Études de terminologie diachronique*. Paris, L'Harmattan, 2014.

4 Cf. A. Rey (éd.), *Dictionnaire culturel en langue française*, Paris, Le Robert, t. I, encadré « botanique », 2005, p. 1009.

5 Cf. J. Magnin-Gonze, *Histoire de la botanique*, Paris, Delachaux et Niestlé, 2004.

toires et nomenclaturaux visent à pouvoir classifier définitivement des millions d'espèces végétales et leur attribuer un nom. À ce propos, pour le dire avec Hoquet, « l'idéal serait de rebattre ces trois objectifs l'un sur l'autre : on a ainsi l'idée d'un nom qui serait également la description du spécimen et sa classification[6] ».

Les intentions de description des botanistes de la période de l'Antiquité jusqu'à la Renaissance sont bien évidentes dans les traités de cette époque où, bien qu'on fournisse plusieurs détails sur l'étymologie des noms des plantes, les informations recueilles restent insuffisantes pour reconnaître la plante elle-même parmi les nombreux êtres du règne végétal. De plus, la représentation graphique, qui devient presque symbolique, est le seul repère utile, mais insatisfaisant, pour l'identification des végétaux. Quelques modifications se produisent au cours des XV^e^ et XVI^e^ siècles, où les gravures qui accompagnent les ouvrages de botanique proposent sous un même nom les espèces végétales, ayant des affinités, rassemblées en groupes. La première tentative de classification scientifique des plantes est alors faite par les frères Bauhin qui établissent des homologies et interprètent les parties et les organes des espèces végétales. Bien qu'elles soient très intéressantes d'un point de vue épistémique parce qu'elles permettent de saisir les prémices de la discipline de la botanique et de discerner les plantes pour rendre compte de leur diversité[7], ces tentatives ne sont proposées que pour appréhender les plantes et les utiliser officinalement, dans une perspective de révérence du Divin[8].

Le manque d'une classification et d'une nomenclature partagées du règne végétal, ainsi que le désaccord sur les dénominations à utiliser entraînent des conséquences très importantes sous plusieurs points de vue : d'une part, le fait de désigner la même plante disponible sur le marché d'une manière différente provoque à la fois une confusion significative dans les échanges

6 Cf. Th. Hoquet, *Les fondements de la botanique. Linné et la classification des plantes*, Paris, Vuibert, 2005.

7 Cf. Ph. Selosse, « Entre Renaissance et Lumières : les nomenclatures des sciences nouvelles », cit., 2016.

8 Cf. Ph. Selosse, « Peut-on parler de classification à la Renaissance : les concepts d'« ordre » et de « classe » dans les ouvrages sur les plantes », *Seizième Siècle*, 8, 2012, p. 39-56 ; Ph. Selosse, « Les concepts de genre et d'espèce à travers l'évolution du modèle logique de la définition, de l'Antiquité à Linné », D. Prat, A. Raynal-Roques, A. Roguenant, J.-C. Lacharpagne (éds.), *Peut-on classer le vivant ? Linné et la systématique aujourd'hui*. Paris, Belin, 2008, p. 65-79.

commerciaux et plusieurs dangers au niveau médical[9] ; de l'autre, dans une perspective linguistique, étant donné que les herbiers et les travaux savants recourent à des dénominations différentes (noms communs en langue vernaculaire et noms latins), plusieurs noms concurrentiels désignant soit la même réalité végétale soit des êtres naturels divers continuent à se perpétuer dans l'usage avec facilité. Ces difficultés augmentent encore plus lors de la découverte d'espèces végétales nouvelles provenant des explorations des territoires de l'Amérique et de l'Asie aux XVI^{e} et XVII^{e} siècles.

Les nomenclatures botaniques proposées par Tournefort et Linné essaient donc de faire face à une confusion linguistique très forte présente dans la communication sur le règne végétal en entier. Cette confusion concerne aussi bien le plan de la dénomination et de la désignation des plantes que la catégorisation des réalités végétales dans des taxons censés regrouper les êtres vivants qui possèdent certains caractères taxinomiques communs. Il se peut, par exemple, que les noms vulgaires ou vernaculaires recouvrent ou ne recouvrent pas des espèces au sens des naturalistes, désignent des espèces différentes appartenant soit au même genre soit à des genres différents, ou encore correspondent à un seul nom scientifique[10].

D'après certains botanistes le recours au latin représente la possibilité de résoudre cette confusion linguistique et d'apporter de l'uniformité dans les dénominations, mais les choix linguistiques, qui ne sont pas forcément partagés par toute la communauté des savants, ne sont pas homogènes et ne favorisent pas la résolution des problématiques de dénomination des plantes. Ce n'est qu'avec le travail de classification présenté par Tournefort dans son ouvrage *Méthode pour connaître les plantes* (1694) que le règne végétal est doté d'une première tentative efficace de nomenclature rationnelle qui utilise le latin, mais qui a recours aussi à une autre stratégie dénominative très intéressante. En effet, dans sa nomenclature Tournefort, qui a une sensibilité très marquée pour les problèmes linguistiques liés à la dénomination des réalités du règne végétal[11], crée des polynômes, où à travers un qualificatif on nomme

9 Par exemple, un même nom peut finir pour désigner des plantes aux propriétés alimentaires et médicales, mais aussi d'autres aux propriétés uniquement toxiques et dangereuses pour la santé.

10 Dans ce dernier cas on peut parler de synonymie diatopique qui concerne l'existence de plusieurs dénominations concurrentielles produites dans un espace géographique restreint qui renvoient à la même réalité végétale qui a une seule dénomination scientifique.

11 À ce propos, Tournefort (*Éléments de Botanique ou Méthode pour Connaître les Plantes*, Paris, Imprimerie Royale, vol. 1, p. 37) affirme que « il faut mettre une grande différence entre nommer les plantes et les décrire ».

les espèces des plantes, qui sont attribuées chacune à un genre, qui est considéré comme étant l'unité synthétique formée par la réunion des espèces les plus voisines. Celles-ci sont donc régulièrement reparties sous leurs signes ou caractères communs, tout en suivant un souci de conception systématique et de clarification nomenclaturale. Le qualificatif employé par Tournefort sert généralement à indiquer l'aspect de la plante, ce qui lie la classification et la nomenclature proposées par le botaniste français aux travaux faits dans les siècles qui précèdent son ouvrage, où l'esprit de représentation et description occupait la plupart des tentatives de classification du règne végétal recueillies dans les herbiers.

L'apport de Tournefort à l'avancement d'une réflexion profonde sur les besoins de normaliser les dénominations du règne végétal est remarquable. Toutefois, sa nomenclature reste très confuse et la longueur des polynômes – étant donné que les polynômes peuvent compter sur plusieurs mots descriptifs placés après le nom du genre – ne résout pas certains problèmes linguistiques auxquels essaie de faire face Linné dans ses ouvrages.

Le botaniste suédois propose, en effet, une nomenclature qui, tout en prenant comme point de départ les travaux des savants qui l'ont précédé, nomme en latin plus de 7 700 espèces de plantes à fleurs qui sont divisées en classes, ordres, genres et espèces[12]. Dans son œuvre maîtresse *Species Plantarum* (1753) Linné établit vingt-quatre classes déterminées en fonction du nombre d'organes sexuels mâles contenant le pollen (les étamines) et de leur mode d'insertion par rapport aux organes sexuels femelles contenant les ovules (les pistils). Sur le plan de l'approche suivie, le système proposé par Linné a le mérite aussi bien de prendre en considération la biologie des plantes que d'affirmer une unité supérieure au genre où peuvent être regroupés tous les genres qui se ressemblent.

Sur le plan de la dénomination, les noms des plantes latins établis par Linné suivent un critère de création très rigide, mais en même temps très utile sur le plan de la classification. Le botaniste remplace les polynômes avec les binômes, formés par le nom du genre auquel l'espèce appartient et par une épithète qui sert à classifier l'espèce à l'intérieur du genre même et qui ne peut être utilisée qu'une seule fois dans un genre donné. Les objectifs de Linné

12 Ph. Selosse, « Peut-on parler de classification à la Renaissance : les concepts d'« ordre » et de « classe » dans les ouvrages sur les plantes », cit., 2012.

sont donc de nommer les espèces dont les dénominations acquièrent une force ontologique[13].

Par rapport aux nomenclatures et dénominations précédentes, celles proposées par Linné laissent donc de côté la valeur descriptive, associée, par exemple, aux qualificatifs utilisés par Tournefort, et acquièrent une valeur uniquement taxinomique. De fait, les épithètes linnéennes permettent de distinguer des espèces à l'intérieur d'un genre, mais ne fournissent aucun élément de reconnaissance de cette espèce puisque les épithètes perdent toute valeur représentative et descriptive.

Les innovations introduites par Linné marquent un tournant décisif dans l'histoire des classifications des règnes naturels[14] puisque pour la première fois une nomenclature réunit en elle les concepts de dénomination et classification. C'est sous cette perspective que grâce au travail de Linné s'impose la nouvelle acception de nomenclature, qui doit être conçue comme la méthode de structuration des dénominations[15]. Linné réussit donc à doter la botanique d'une langue « rationnelle, systématique et universelle[16] », où chaque mot possède une valeur bien précise au niveau conceptuel et où la nomenclature se configure comme une terminologie structurée taxinomiquement.

3. Conclusion

Les nomenclatures botaniques représentent des cas d'étude très intéressants du point de vue linguistique puisqu'elles permettent d'analyser en profondeur les rapports qui s'établissent entre termes et concepts pour l'élaboration de terminologies structurées taxinomiquement. En particulier, les nomenclatures botaniques de Tournefort et Linné démontrent de manière efficace le passage épistémique de la *Res Herbaria* à la Botanique, dont les objectifs sont de plus en plus liés à la classification des réalités du règne végétal sans aucun rapport direct avec l'exploitation des plantes au niveau médical. La valeur du modèle construit est soulignée par son adoption à la fois dans cette

13 Ph. Selosse, « Préface », D. Reynaud, Ph. Selosse (éds.), *Nomenclatures au dix-huitième siècle : la science, « langue bien faite » (tricentenaire Linné-Buffon)*, Lyon, Presses de l'Aristoloche, 2015, p. 6.

14 S. Müller-Wille, « La science baconienne en action : la place de Linné dans l'histoire de la taxonomie », Th. Hoquet, (dir.), *Les fondements de la botanique. Linné et la classification des plantes*, Paris, Vuibert, 2006, p. 57-102.

15 Ph. Selosse, « Préface », cit., 2015, p. 7.

16 P. Duris, « Traduire Linné en français à la fin du XVIII^e siècle », *Early Science and Medicine*, 12, 2006, p. 166-186.

discipline[17] et dans d'autres domaines scientifiques d'application (zoologie, minéralogie, médecine, chimie et géographie).

Sur le plan linguistique la nomenclature proposée par Linné joue un rôle très important[18] puisque pour la première fois elle attribue aux noms latins des plantes une valeur conceptuelle : ces noms servent pour situer l'espèce désignée à l'intérieur de la classification du règne végétal, mais les dénominations choisies, dépourvues de tout caractère descriptif, perdent la valeur représentative et évocatrice qu'elles avaient dans d'autres systèmes de classification précédents. Les épithètes créées par Linné deviennent des désignations symboliques permettant de distinguer les espèces, mais ne prétendent fournir aucun élément de reconnaissance extérieure de l'espèce elle-même. Tout en passant de la description à la classification, la nomenclature linnéenne, une terminologie ayant force ontologique privée de toute valeur évocatrice, repose donc sur un souci de plus en plus majeur de transmission exacte du concept, la notion d'unité taxinomique qui est au cœur du travail de Linné n'étant que de nature purement conceptuelle.

17 Cf. C. von Linné, *Systema naturae, sive Regna tria naturae systematice proposita per classes, ordines, genera et species*, Lugduni Batavorum, J. Haak, 1735 ; C. von Linné, *Critica Botanica*, Lugduni Batavorum, C. Wishoff, 1737 ; C. von Linné, *Philosophia Botanica*. Stockholmiae, C. Wishoff, 1751.

18 L'importance historique de la nomenclature de Linné proposée dans son ouvrage *Species Plantarum* est, d'ailleurs, soulignée par l'adoption de la date de publication de l'ouvrage linnéen (1753) comme point de départ de la nomenclature botanique utilisée aujourd'hui. Avant cette date, les noms, même sous forme de binômes, ne sont plus utilisables.

Références

Brunot, F. (1966). *Histoire de la langue française. De l'origine à nos jours*. Paris, Armand Colin, t. VI.

Dictionnaire de l'Académie française, <https://www.dictionnaire-academie.fr/>. Site consulté en septembre 2020.

Dictionnaire universel, français et latin, vulgairement appelé *Trévoux* (édition de 1771), <http://www.cnrtl.fr/dictionnaires/anciens/trevoux/>. Site consulté en septembre 2020.

Duris, P. (2007). « Traduire Linné en français à la fin du XVIII^e siècle ». *Early Science and Medicine*, 12, p. 166-186.

Encyclopédie ou dictionnaire raisonné des sciences, des arts et des métiers (1751-1772), <http://enccre.academie-sciences.fr/encyclopedie/>. Site consulté en septembre 2020.

Furetière, A. (1690), *Dictionnaire universel*, *contenant généralement tous les mots françois, tant vieux que modernes, et les termes de toutes les sciences et des arts*, <http://www.furetière.eu/>. Site consulté en septembre 2020.

Grimaldi, C. (2017). *Discours et terminologie dans la presse scientifique française (1699-1740). La construction des lexiques de la botanique et de la chimie*. Oxford, Peter Lang.

Hoquet, Ph. (2005). *Les fondements de la botanique. Linné et la classification des plantes*. Paris, Vuibert.

Linné, C. von (1735). *Systema naturae, sive Regna tria naturae systematice proposita per classes, ordines, genera et species*. Lugduni Batavorum, J. Haak.

Linné, C. von (1737). *Critica Botanica*. Lugduni Batavorum, C. Wishoff.

Linné, C. von (1751). *Philosophia Botanica*. Stockholmiae, C. Wishoff.

Linné, C. von (1753). *Species Plantarum*. Holmiae, Impensis L. Salvii.

Magnin-Gonze, J. (2004). *Histoire de la botanique*. Paris, Delachaux et Niestlé.

Müller-Wille, S. (2006). « La science baconienne en action : la place de Linné dans l'histoire de la taxonomie ». In Hoquet, Th. (dir.) (2006). *Les fondements de la botanique. Linné et la classification des plantes*. Paris, Vuibert, p. 57-102.

Rey, A. (éd.) (2005). *Dictionnaire culturel en langue française*. Paris, Le Robert, t. I.

Selosse, Ph. (2008). « Les concepts de genre et d'espèce à travers l'évolution du modèle logique de la définition, de l'Antiquité à Linné ». In Prat, D., Raynal-Roques, A., Roguenant, A., Lacharpagne J.-C. (éds.) (2008).

Peut-on classer le vivant ? Linné et la systématique aujourd'hui. Paris, Belin, p. 65-79.

Selosse, Ph. (2012). « Peut-on parler de classification à la Renaissance : les concepts d'« ordre » et de « classe » dans les ouvrages sur les plantes ». *Seizième Siècle*, 8, p. 39-56.

Selosse, Ph. (2015). « Préface ». In Reynaud, D., Selosse, Ph. (éds.) (2015). *Nomenclatures au dix-huitième siècle : la science, « langue bien faite » (tricentenaire Linné-Buffon)*. Lyon, Presses de l'Aristoloche, p. 5-13.

Selosse, Ph. (2016). « Entre Renaissance et Lumières : les nomenclatures des sciences nouvelles ». In Werner, F., Thörle, B. (éds.) (2016). *Manuel des langues de spécialité*. Berlin, De Gruyter, p. 413-430.

Tournefort, J. Ph. de (1694). *Elemens de Botanique ou Méthode pour Connaître les Plantes*. Paris, Imprimerie Royale, vol. 1.

Zanola, M.T. (2014). *Arts et métiers au XVIII^e siècle. Études de terminologie diachronique*. Paris, L'Harmattan.

L'implantation terminologique 20 ans après : le cas de la télédétection aérospatiale

Adam Renwick

Centre de recherche en terminologie et traduction (CRTT)
Université Lumière Lyon 2

La France dispose, depuis 1972, d'un dispositif d'enrichissement de la langue. Sous ses différentes formes (Commissions Ministérielles de terminologie entre 1972 et 1995, Commissions spécialisées de terminologie et de néologie entre 1996 et 2015, et groupes d'experts à partir de mars 2015) ce dispositif œuvre dans le but de « favoriser l'enrichissement de la langue française, de développer son utilisation, notamment dans la vie économique, les travaux scientifiques et les activités techniques et juridiques, [et] d'améliorer sa diffusion. » (Art. 1er, Décret nº 96-602 du 3 juillet 1996). Depuis la création de la toute première commission ministérielle en 1972, quelque 7 000 termes ont été recommandés dans le cadre de cette politique d'aménagement terminologique.

Cet aménagement terminologique ne représente qu'une partie d'une politique de la langue, pour citer Auger : « The terminology planning process comprises six different steps or functions, namely : (a) a research function ; (b) a standardization function ; (c) a dissemination function ; (d) an implantation function ; **(e) an evaluation and measurement function** ; and (f) a modernization fonction. » (Auger 1986, cité dans Quirion, 2003, 29-30, notre soulignement). L'aménagement terminologique entamé par la France constitue la fonction de normalisation *(b)*, mais c'est la fonction *(e)*, dénommée fonction d'évaluation de mesure, qui cherche à analyser l'implantation terminologique qui nous intéresse dans le cadre de cette étude.

1. Les études d'implantation

L'implantation terminologique est un champ de recherche relativement peu exploré, dont les études suivantes constituent la plus grande partie : sur l'hébreu (Allony-Fainberg (1974, 1983)), le catalan (Montané 2012, Montané et Cabré (2007, 2013), Cabré 2010), le turc (Karabacak 2009) et sur le fran-

çais du Québec (Quirion (2000, 2003, 2004), Auger (1999) et Martin (1998). Le français hexagonal a lui aussi fait l'objet de plusieurs études ponctuelles (Gaudin 1991, Daoust 1995), ainsi qu'une série d'études répondant à un appel de la délégation générale à la langue française. Cet appel a suscité des études réalisées en 1992-1993 par Chansou, Gouadec, Thoiron, Iwaz et Zaouche, Guespin et Gaudin, ainsi que par Fossat et Rouges-Martinez. Les résultats de ces études font l'objet du douzième numéro de *Terminologies nouvelles* de 1994, qui reprend les actes du séminaire qui a eu lieu à Maromme en décembre 1993. Les études de Chansou, Guespin et Gaudin, Gouadec, Fossat et Rouges-Martinez, et Thoiron, Iwaz et Zaouche, sont remises en forme et publiées dans un volume en 1997 de Depecker et Mamavi.

Reprises ensemble, toutes ces études ne s'intéressent qu'à une dizaine de domaines : le génie génétique, les transports, la santé et la médecine, la métallurgie, l'informatique, l'environnement, l'économie, l'halieutique, l'audiovisuel et la publicité, les sports, la télédétection aérospatiale et le droit. Qui plus est, on constate ces dernières années un manque d'études d'implantation, notamment en France, ainsi qu'un suivi de ces études. Nous nous proposons de nous concentrer sur l'étude d'implantation de Fossat et Rouges-Martinez (1997) (rédigée en 1993, publiée dans Depecker et Mamavi (1997)), qui s'est intéressée à l'utilisation des termes recommandés par la Commission Ministérielle de Terminologie de la Télédétection Aérospatiale (COMITAS) et d'analyser l'implantation des mêmes termes dans un corpus écrit dans le but de mieux comprendre l'implantation de ces termes.

2. L'étude de Fossat et Rouges-Martinez sur l'implantation des termes de la télédétection aérospatiale

La COMITAS a recommandé quelques centaines de termes dans sept arrêtés de terminologie publiés en 1980, 1982, 1984, 1986, 1987, 1988 et 1990, chez ses différents ministères de tutelle (en 1980, ministère des Universités, 1982-1987, ministère de l'Éducation nationale, et 1988-1990, ministère de l'Éducation nationale, de la jeunesse et des sports). Fossat et Rouges-Martinez (1997) tentent d'étudier l'implantation de cette terminologie recommandée, et leur étude commence en la délimitant à un échantillon de 59 termes représentatifs des nombreux sous-domaines de cette discipline. Ce travail de délimitation a été effectué par un expert, qui n'a pas participé aux étapes suivantes de l'étude (Fossat et Rouges-Martinez 1997 : 101).

La consultation d'experts du domiane s'est réalisée par le biais de questionnaires et d'entretiens avec des experts dans le domaine. L'expert a été invité à exprimer son attitude vis-à-vis du terme dans de nombreuses situations discursives, telles que, à l'écrit, dans un article de revue scientifique, un cours polycopié, une correspondance à titre professionnel ou, à l'oral, dans une soutenance, ou dans un colloque, pour ne nommer que quelques exemples des 22 situations de communication proposées. Il s'agit donc d'une étude qui analyse des attitudes et de l'utilisation reportées par les experts, au lieu d'une étude de l'utilisation des termes eux-mêmes, comme le note Quirion (2003 : 31). Cette première étape riche en détail sur l'attitude des experts s'est suivie par un entretien pour mieux saisir les raisons pour lesquelles un tel expert utiliserait un tel terme dans un tel énoncé. À partir de toutes les données recueillies, Fossat et Rouges-Martinez ont pu décrire quatre comportements possibles envers des termes chez les experts, dont la première est l'utilisation du terme recommandé, la deuxième est l'hésitation à utiliser le terme recommandé, la troisième est la substitution du terme recommandé par un autre terme, et la quatrième est de rejeter le terme recommandé (Fossat et Rouges-Martinez 1997 : 107).

Selon le taux de chaque type de réponse, Fossat et Rouges-Martinez (1997) ont pu diviser les 59 termes retenus pour leur étude en trois classes. La première, la classe A, comprend les termes dont le taux de réponses positives, appelé « indice d'implantation », est qualifié de « fort », c'est-à-dire que dans plus de 70 % des cas, les experts disent utiliser ce terme ; 30 termes (voir **Tableau 2**) appartiennent à cette classe A (Fossat et Rouges-Martinez 1997 : 104). Lorsque le taux de réponses positives est supérieur à 30 %, mais inférieur à 69,9 %, l'indice d'implantation est qualifié de « moyen » (17 termes appartiennent à cette classe B, voir **Tableau 3**) et enfin, les 12 termes dont l'indice d'implantation est inférieur à 30 % appartiennent à la classe C (voir **Tableau 4**), dont l'indice d'implantation est qualifié de « faible » (Fossat et Rouges-Martinez 1997 : 105).

Nous nous proposons d'utiliser cette étude de 1997 de Fossat et Rouges-Martinez comme point de référence pour examiner l'implantation de ces mêmes 59 termes dans un corpus de discours écrit des années 2010-2014 en comparant nos données avec celles de Fossat et Rouges-Martinez (1997) pour mettre en évidence d'éventuelles divergences et évolutions. Le corpus que nous consulterons est un corpus de 95 millions de mots, créé pour examiner l'implantation terminologique dans les sciences et techniques spatiales. Le corpus que nous avons construit se divise en sept sous-corpus, selon le type

de document et son degré de spécialisation. Le **Tableau 1** récapitule les points clés de ces corpus.

	Types de document compris	**Exemples de titres inclus**	**Mots (approx.)**
Corpus 1	Presse générale de grand tirage	*Le Monde* *Le Figaro*	35 000 000
Corpus 2	Documents didactiques	*CNESMag*, écoles d'été d'IN2P3, du GRGS)	1 100 000
Corpus 3	Discours officiel	*Journal officiel de la République, C.R.Acad.*	26 100 000
Corpus 4	Discours institutionnel	Rapports d'activités du CNRS, GRGS, Arianespace	2 100 000
Corpus 5	Discours spécialisé	Thèses de doctorat	20 200 000
Corpus 6	Discours demi-spécialisé	Magazines des sociétés d'astronomie	560 000
Corpus 7	Discours de vulgarisation	*La Recherche, Pour la science*	9 900 000
Corpus Total			94 960 000

Tableau 1 : Description des corpus

Pour comprendre l'implantation des termes, nous avons eu recours à trois méthodes principales. Premièrement, nous avons calculé la fréquence relative de chaque terme, car, si dans un corpus hypothétique, le terme *software* est relevé une fois par million de mots et le terme *logiciel* mille fois par million de mots, il est évident que le terme *logiciel* s'utilise plus fréquemment pour dénommer le concept de [logiciel]. La fréquence relative peut bien s'avérer utile, mais elle ne reflète pas l'ampleur de l'implantation du terme face à ses concurrents (Quirion 2003 : 32). C'est pour cette raison que nous avons eu recours à une deuxième méthode, le coefficient d'implantation. Le coefficient d'implantation, proposé par Quirion (2003), est le quotient du nombre d'occurrences du terme ciblé et du nombre d'occurrences du concept que ce terme dénomme ; il est aussi possible d'exprimer ce calcul sous la forme de l'**Équation 1.**

$$\textbf{Coefficient} = \frac{\textbf{occurrences du terme ciblé}}{\textbf{occurrences du concept}}$$

Équation 1 – (basée sur Quirion (2003 : 33)

Considérons un cas hypothétique où le terme *logiciel* s'utilise 96 fois pour dénommer le concept de [logiciel], et *software* est répertorié 4 fois pour dénommer ce même concept. Le nombre d'occurrences du concept est égal à 96 (occurrences de *logiciel*) + 4 (occurrences de *software* = 100. Si le terme ciblé est *logiciel*, son coefficient est 0,96, car 96 (nombre d'occurrences du terme ciblé) divisé par 100 (nombre d'occurrences du concept) fait 0,96. D'autre part, si le terme ciblé est *software*, le nombre d'occurrences de ce terme ciblé (4) divisé par le nombre d'occurrences du concept (100) donne un coefficient d'implantation de 0,04. Le coefficient est donc une indication de la proportion des cas dans lesquels un terme est utilisé pour dénommer un concept. Le recours à ce calcul permet, selon Quirion, une meilleure compréhension de l'implantation des termes parce que « the implantation coefficients indicates the degree of implantation of the given term ; the total of the implantation coefficients of each of the various désignations for a single notion amounts to 1 » (Quirion 2003 : 32). L'utilité de ce calcul peut néanmoins être remise en question lorsque le nombre d'occurrences du concept est bas, et notamment dans des cas où un concept est évoqué une fois au sein d'un texte, ce qui donne au terme un coefficient possiblement trompeur de 1.

Nous nous proposons de complémenter ces calculs en analysant l'implantation des termes dans le corpus de GoogleBooks, à l'aide du NgramViewer. GoogleBooks est un projet de numérisation de livres, et à partir du texte des livres numérisés, des n-grams ont été extraits. Un n-gram est une suite de *n* mots, donc *le chapeau* et le terme *équidensité colorée* sont des 2-grams, car contenant deux mots graphiques ; *monter sur le podium* et le terme *verrouillage sur le fouillis* sont des 4-grams, car contenant quatre mots graphiques. L'interface du NgramViewer de Google permet de rechercher des n-grams (jusqu'à n=5) dans le corpus de GoogleBooks, dont le corpus français comprend 45 milliards de mots dans des livres publiés entre 1500 et 2010 (Michel *et al.* 2010 : 176). Les résultats d'une recherche dans le corpus s'affichent sous forme d'une courbe, dont chaque point représente le nombre d'occurrences du n-gram recherché dans des textes rédigés au cours d'une année. Il est ainsi

possible de visualiser l'évolution de l'utilisation d'un n-gram dans un corpus de taille très importante.

Pour revenir au corpus que nous avons construit, il est aussi possible d'examiner dans quel(s) corpus apparaissent les termes. Nous notons que, au total, 44 des 59 termes retenus pour l'étude (c'est-à-dire les mêmes 59 termes que Fossat et Rouges-Martinez (1997) avaient retenus), sont présents dans notre corpus entier, dont 40 dans notre corpus de discours spécialisé. Cela indique que nos corpus recouvrent assez bien le discours du domaine de la télédétection. Il est aussi intéressant que six termes, *pixel, numérisation, capteur, fausse couleur, radiomètre* et *rayonnement électromagnétique*, soient relevés dans notre corpus de discours général, ce qui indiquerait que ces vocables se banalisent, malgré leur qualité de termes techniques. Pour *pixel*, notamment, il serait même question de voir si celui-ci est devenu un mot du discours général à titre complet, tout en respectant son acception technique, car continuant de décrire le « plus petit élément d'une surface de détection, d'une image échantillonnée ou d'une surface de visualisation, auquel on puisse affecter individuellement des caractéristiques visuelles » (*Journal officiel* du 15/11/2001).

Nous tournons notre attention aux classes que proposent Fossat et Rouges-Martinez (1997) ainsi qu'à l'analyse en détail de l'implantation de certains vocables de chaque classe qu'ils analysent.

3. La Classe A

Pour Fossat et Rouges-Martinez, lorsqu'un terme suscite un taux de réponses positives supérieur à 70 %, il appartient à la classe A, dont l'indice d'implantation est dit fort ; il y a 30 termes dans cette classe (Fossat et Rouges-Martinez 1997 : 113). Il nous a été obligé d'exclure quatre termes appartenant à cette classe de notre étude, *scène, détecteur, capteur* et *imagerie,* faute de moyen fiable qui permettrait d'examiner le contexte de chaque occurrence de chacun de ces vocables extrêmement récurrents dans nos corpus pour découvrir s'il s'agit d'une utilisation du terme ou non. Pour les 26 termes restants, nous avons calculé le coefficient d'implantation ainsi que la fréquence relative, et ces données sont présentées dans le **Tableau 2**. Parmi ces 26 termes, 13 ont un coefficient d'implantation égale à 1. Le terme avec le coefficient le plus bas est *infrarouge couleur* (0,6), et la médiane est égale à 1, la moyenne de 0,88. Ces deux dernières mesures indiquent que les termes de cette classe sont toujours « en bonne voie d'implantation », pour utiliser la description de Rouges-Martinez (1997 : 103). En revanche, les fréquences relatives de ces

termes sont très variées : la moyenne est de 7,0 occurrences par million de mots, la médiane est égale à 5,84. Le maximum est celle de *bande spectrale* (45 occurrences par million de mots dans le corpus de discours spécialisé), le minimum est celle de *infrarouge couleur* (0,2 occurrence par million de mots dans le corpus de discours spécialisé). On remarque également qu'une fréquence relative élevée ne va pas toujours de pair avec un coefficient d'implantation élevé, comme le montre *infrarouge couleur,* un concept peu fréquent dans nos corpus (évoqué 10 fois au sein de l'ensemble de nos corpus : 0,06 occurrence par million de mots), alors que son coefficient d'implantation de 0,6 démontre que, lorsque ce concept s'exprime, dans la majorité des cas, *infrarouge couleur* est utilisé pour le dénommer.

3.1. Spatiocarte

Fossat et Rouges-Martinez calculent un taux de rejet de 13,6 % pour le terme *spatiocarte,* et notent que ce « taux de rejet peut traduire le fait qu'il s'agit d'un terme récent » (Fossat et Rouges-Martinez 1997 : 107). Cela parce que *spatiocarte* est un terme recommandé en 1987 et dont l'implantation a été étudiée en 1992. Il s'agit du terme avec le taux de rejet le plus élevé parmi les termes de la Classe A. Nous calculons un coefficient d'implantation de 0,64 pour ce terme, ce qui montre que dans presque deux cas sur trois, *spatiocarte* s'emploie pour dénommer le concept. Cependant, avec seulement 11 occurrences du concept relevées au sein de nos corpus, il n'est pas possible de statuer de façon définitive pour ce qui est de son implantation.

A contrario, lorsqu'on consulte le corpus de GoogleBooks, la donne est entièrement différente : c'est le terme *spatiocarte* qui est utilisé, et les dénominations *carte satellite, carte satellitaire, carte d'image satellitaire* ainsi que les termes anglais *space map* et *satellite image map* qui sont totalement absents. Néanmoins, il faut souligner qu'il n'est pas possible d'examiner chaque occurrence de chaque dénomination dans son contexte à partir de l'interface du NgramViewer. Toutefois, le NgramViewer met en exergue une piste pour d'éventuelles recherches complémentaires pour vérifier si cette tendance relève du discours spécialisé de la télédétection aérospatiale.

3.2. Pixel

Pour Fossat et Rouges-Martinez, *pixel* est un terme avec un fort indice d'utilisation à l'oral, mais lorsqu'il s'agit de l'écrit, « Les interlocuteurs semblent s'être demandé si à l'écrit, plus formel, plus normatif, il était 'per-

mis' d'utiliser ce terme » (1997 :114). Alors que l'utilisation de ce terme à l'oral dépasse le cadre de notre étude, le taux d'occurrence de *pixel* dans notre corpus de discours spécialisé de 223 occurrences par million de mots, ainsi qu'un manque absolu de concurrents, permettent de constater une évolution dans ce domaine, car même dans le cadre formel de la thèse de doctorat, il est permis d'utiliser ce terme à l'écrit.

Classe A : Termes avec un fort indice d'implantation	**Indice d'implantation**	**Coefficient d'implantation**	**Fréquence (par million de mots)**
bande spectrale	87,7%	1,00	9,79
capteur	89,1%	**EXCLU**	**EXCLU**
composition colorée	85,7%	1,00	0,69
détecteur	87,3%	**EXCLU**	**EXCLU**
donnée brute	85,0%	0,97	2,55
données de terrain	81,6%	1,00	0,76
fausse couleur	81,1%	1,00	0,95
hyperfréquence	70,0%	0,85	0,48
image filtrée	81,1%	1,00	0,05
image prétraitée	86,1%	1,00	0,02
image radar	89,8%	1,00	2,94
image-satellite	83,6%	0,59	3,21
imagerie	79,3%	**EXCLU**	**EXCLU**
indice de végétation	79,8%	1,00	0,73
infrarouge couleur	85,7%	0,60	0,06
mosaïque	85,9%	0,30	1,93
numérisation	85,0%	0,99	5,60
pas d'échantillonnage	77,0%	0,99	1,36
pixel	91,1%	**EXCLU**	**EXCLU**
pouvoir de résolution	80,0%	0,59	0,66
radiomètre	89,3%	1,00	3,28
rayonnement électromagnétique	86,1%	1,00	2,71

Classe A : Termes avec un fort indice d'implantation	**Indice d'implantation**	**Coefficient d'implantation**	**Fréquence (par million de mots)**
réflectance	89,3%	0,80	1,94
rétrodiffusion	75,2%	0,92	3,19
scène	87,9%	**EXCLU**	**EXCLU**
signature spectrale	89,8%	0,67	4,79
spatiocarte	76,2%	0,64	0,07
thermographie	77,0%	1,00	0,23
trace du satellite	81,6%	1,00	0,33
traitement d'image numérique	75,9%	1,00	0,02
Median	0,85	1,00	0,95
Moyenne	**83,3%**	**0,88**	**1,90**

Tableau 2. Indices d'implantation, coefficients d'implantation et fréquence par million de mots de termes de la Classe B de Fossat et Rouges-Martinez. Basé sur Fossat et Rouges-Martinez (1997 : 105)

La consultation du corpus de GoogleBooks pour examiner l'utilisation du terme *pixel* ainsi que de ses concurrents démontre un problème pour le terminologue. Tandis qu'elles sont absentes de nos corpus, les dénominations *élément d'image* et *picture element* sont relevées dans ce corpus-ci, et cela relève du même phénomène que de dégager d'éventuelles pistes pour le terminologue. Deux autres faiblesses du NgramViewer se montrent lorsqu'on consulte les occurrences de *pixel, PEL, élément d'image et picture element.* Premièrement, il n'est pas possible de délimiter les occurrences de *pixel* au domaine qui nous intéresse ; cela a pour conséquence d'introduire le bruit des autres domaines qui utilisent aussi ce vocable. Deuxièmement, il n'est pas possible de distinguer entre les différents sens possibles de *PEL* : relevé dans le corpus de GoogleBooks, il pourrait s'agir d'une occurrence du terme qui fait l'objet de notre étude, d'un acronyme de finances (toutes les occurrences de *PEL* dans nos corpus correspondaient à *Plan Épargne Logement*), ou bien d'un autre acronyme ou même d'un nom propre en majuscules.

3.3. Signature spectrale

Le terme *signature spectrale* est relevé dans cinq de nos sept sous-corpus et Fossat et Rouges-Martinez lui attribuent un indice d'implantation de 89,8 % (1997 : 104). Nous calculons un coefficient d'implantation de 0,67, plus bas que ce que l'on pourrait attendre vu son indice élevé. Il s'avère que ce coefficient est bas en raison de ses nombreux concurrents relevés dans nos corpus (*réponse spectrale, signature spécifique, caractéristique d'une cible* et *spectral signature*). Sa fréquence est, en revanche, élevée dans notre corpus de discours spécialisé (21 occurrences par million de mots), mais ce chiffre baisse à 4,79 occurrences par million de mots dans notre corpus entier. Alors que dans nos corpus *signature spectrale* a un coefficient d'implantation de 0,67, le NgramViewer montre une rivalité entre les appellations *signature spectrale* et *réponse spectrale.* Dans ce cas, il semble qu'il y ait une influence de l'anglais par voie du calque du terme anglais *spectral signature*, car les deux appellations qui comprennent l'adjectif postposé *spectrale* s'utilisent beaucoup plus fréquemment que les termes qui ne le comprennent pas. Faute de contexte précis, il n'est pas possible de constater de divergence entre nos données et celles de GoogleBooks. Nous pouvons simplement dire que les données de GoogleBooks montrent qu'il y aurait une rivalité entre dénominations que nous n'avons pas observée dans nos corpus.

4. Classe B

Les termes dont l'indice d'implantation est supérieur à 30 %, mais inférieur à 70 %, appartiennent à la Classe B dans l'étude de Fossat et Rouges-Martinez (1997 : 105). Il y a 17 termes dans cette classe, présentés dans le **Tableau 3**, dont quatre sont absents de nos corpus (*donnée géocodée*, *équidensité colorée*, *image d'écran* et *image multidate*). Un terme, *cible*, a été exclu, selon le même raisonnement évoqué pour les quatre termes de la Classe A que nous avons dûs exclure de notre étude. Cinq termes de la Classe B ont un coefficient d'implantation égale à 1, la médiane est égal à 0,94 et la moyenne est de 0,63. En comparaison, les médianes des coefficients d'implantation sont très similaires entre les Classes A et B : (1,0 pour la Classe A, 0,94 pour la Classe B). Par contre, la différence entre les moyennes est plus importante (0,88 pour la Classe A, 0,63 pour la Classe B).

Classe B : Termes avec un indice d'implantation moyen	Indice d'implantation	Coefficient d'implantation	Fréquence (par million de mots)
cible	**60%**	**EXCLU**	**EXCLU**
donnée d'apprentissage	49,1%	1,00	0,50
donnée-image	51,8%	1,00	0,31
donnée géocodée	68,9%	0,00	0,00
équidensité colorée	55,4%	0,00	0,00
filtrage numérique	64,3%	1,00	0,02
image d'écran	41,6%	0,00	0,00
image diachronique	30,2%	0,00	0,00
imagette	66,8%	0,49	0,26
ligne de balayage	61,8%	0,90	0,09
lobe d'antenne	46,4%	1,00	0,20
orthophotoplan	37,9%	0,99	1,17
réalité de terrain	37,9%	0,02	0,10
scanneur	49,3%	0,02	0,01
segment de données	42,3%	1,00	0,05
vérité-terrain	59,5%	0,98	1,35
zone d'apprentissage	55,9%	1,00	0,04
Médiane	**51,80%**	**0,94**	**0,07**
Moyenne	**51,71%**	**0,63**	**0,26**

Tableau 3. Indices d'implantation, coefficients d'implantation et fréquence par million de mots de termes de la Classe B de Fossat et Rouges-Martinez. Basé sur Fossat et Rouges-Martinez (1997 : 105)

Bien que mitigée selon ces données, la différence entre ces classes est beaucoup plus marquée au niveau des fréquences : la fréquence moyenne des termes de la Classe A était, dans notre corpus entier 1,90 occurrence par million de mots, pour la Classe B, elle est égale à 0,26 occurrence par million de mots. Les médianes montrent une distinction similaire entre ces deux classes : dans la Classe A, la fréquence médiane est 0,95 occurrence par million de mots, dans la Classe B, elle est égale à 0,07 occurrence par million de mots. L'implantation des termes de la Classe B est donc beaucoup plus limitée que celle des termes de la Classe A.

4.1. Image d'écran

Appartenant à la Classe B, *image d'écran* a un indice d'implantation moyen (41,6%) pour Fossat et Rouges-Martinez (1997 : 105), mais il s'agit d'un terme entièrement absent de nos corpus, donc son coefficient d'implantation est 0. Cependant, nous attestons la présence de certains de ses concurrents (*image sur écran* et *copie d'écran*) dans nos corpus. Bien que le terme recommandé ne soit pas attesté dans nos corpus, il semble encourageant pour la politique d'aménagement terminologique dans laquelle la France s'est engagée que ce concept se dénomme avec des vocables français et que l'équivalent anglais *soft copy* ne soit pas relevé. Le NgramViewer montre encore une situation différente ; c'est *copie d'écran* qui se dit le plus souvent, mais puisqu'il ne nous est pas possible d'examiner les vocables dans leur contexte pour savoir si on a affaire à des termes ou non, on ne peut statuer de façon plus définitive au niveau de l'implantation de ce terme.

4.2. Scanneur

Pour Fossat et Rouges-Martinez (1997 : 119), l'indice d'utilisation de 49,3 % du terme *scanneur* révèle une certaine pénétration de ce vocable dans le discours des spécialistes. Cependant, dans nos corpus, nous attestons une seule occurrence du terme *scanneur*. Par contre, lorsque le suffixe nominal anglais *-er* remplace par son analogue français *-eur*, nous attestons 44 occurrences du terme sous la forme *scanner* dans le domaine qui nous intéresse. Malheureusement, il ne nous est pas possible de prononcer sur la prononciation ([skane] ou [skanœr] ou autre) de ce terme. Le terme polylexical *radiomètre à balayage* n'est relevé que cinq fois, et uniquement dans notre corpus de thèses, et le terme anglais *scanning radiometer* est entièrement absent de nos corpus.

Grâce à l'étiquetage morpho-syntaxique du corpus de GoogleBooks, le NgramViewer permet de distinguer les différentes catégories grammaticales d'un vocable. Ainsi, nous avons pu exclure les cas où *scanner* est employé comme verbe, mais, si nous nous fions à ce que nous a montré nos corpus, il nous restera un grand nombre de cas où *scanner* correspond non pas l'appareil utilisé dans la télédétection aérospatiale, mais aux appareils d'imagerie médicale et d'informatique. Par conséquent, nous ne pouvons que conclure que l'unité lexicale *scanner*, revêtue de toutes ses acceptions possibles, s'utilise

beaucoup plus fréquemment que l'unité lexicale *scanneur,* encore revêtue de toutes ses acceptions possibles

Alors il nous semble que la forme *scanneur* n'est pas aussi usitée que *scanner,* et donc que l'implantation mitigée de *scanneur* qu'ont décrit Fossat et Rouges-Martinez a continué de baisser face à *scanner* (1997 : 119). Nous soulignons que l'implantation compromise de ce terme ne relève pas du domaine de la télédétection, car dans sa recommandation du 10 octobre 1998, la Commission générale de terminologie et de néologie (CGTN) a recommandé que l'appareil informatique dénommé *scanner* en anglais devrait s'appeler *numériseur* ou *scanneur* (CGTN 1998) au lieu de la forme anglaise. Nous pouvons donc constater que, dans ce cas, l'aménagement terminologique ne porte pas ses fruits.

4.3. Imagette

Imagette est un terme avec un indice d'implantation de 66,8 %, donc il fait partie termes les plus implantés de la Classe B dans l'étude de Fossat et Rouges-Martinez (1997 : 116). Ses occurrences sont concentrées dans notre corpus de discours spécialisé et son coefficient d'implantation de 0,49 laisse la place à plusieurs concurrents : *quick look* (coefficient d'implantation de 0,1), *sous-image* (0,12) et *fenêtre de l'image* (0,29). En raison des facteurs expliqués plus haut, la consultation de GoogleBooks nous permet uniquement de dire que *imagette* est un vocable qui s'utilise très peu, et qu'un concurrent non-relevé dans nos corpus, *image de contrôle* s'utilise, mais encore, très rarement.

5. Classe C

Douze termes appartiennent à la Classe C que proposent Fossat et Rouges-Martinez en raison de leur faible indice d'implantation, c'est-à-dire les termes dont le taux de réponses positives est inférieur à 30 % ; ces termes sont présentés dans le **Tableau 4** (1997 : 105). Parmi ces 12 termes, ce sont uniquement *canal spectral, chatoiement* et *zone d'entraînement* (dont les coefficients d'implantation sont 1, 0,05 et 0,2 respectivement) qui sont attestés dans nos corpus.

Classe C : Termes avec un faible indice d'implantation	**Indice d'implantation**	**Coefficient d'implantation**	**Fréquence (par million de mots)**
canal spectral	23,6%	1,00	1,16
chatoiement	5,9%	0,05	0,13
classification dirigée	20,9%	0,00	0,00
classification non dirigée	15,9%	0,00	0,00
déversement-radar	8,2%	0,00	0,00
fouillis d'échos	5,4%	0,00	0,00
image chronoséquentielle	5,4%	0,00	0,00
rabattement-radar	8,7%	0,00	0,00
radargraphie	14,8%	0,00	0,00
tachèle	3,6%	0,00	0,00
zone d'entraînement	27,5%	0,20	0,01
zone d'essai	23,6%	0,00	0,00
Median	**0,12**	**0,00**	**0,00**
Moyenne	**13,625%**	**0,21**	**0,11**

Tableau 4. Indices d'implantation, coefficients d'implantation et fréquence par million de mots de termes de la Classe C de Fossat et Rouges-Martinez. Basé sur Fossat et Rouges-Martinez (1997 : 105)

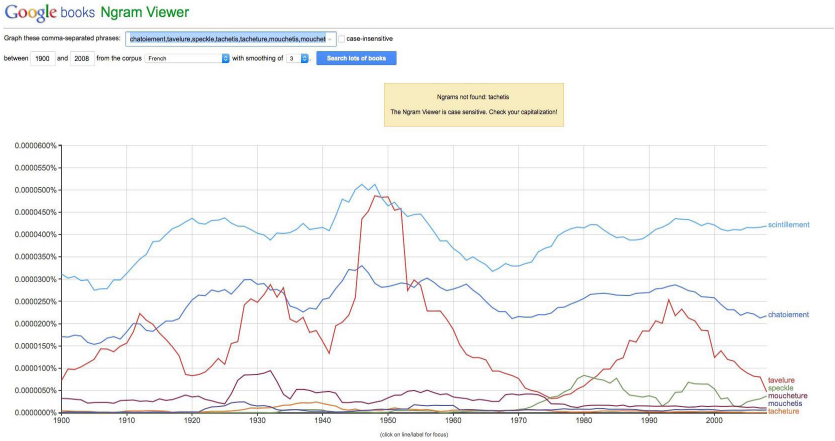
Le terme *zone d'entraînement* est un hapax et son concept est évoqué seulement cinq fois, ce qui rend hasardeuse toute conclusion à propos de son implantation à partir de son coefficient d'implantation. Il est évident que *canal spectral* ressort ici, faisant partie de cette classe de termes anciennement peu implantés chez les spécialistes, mais avec un coefficient d'implantation de 1,00. Contrairement à *zone d'entraînement*, il n'y a pas de hasard en jeu ici, car le concept de *canal spectral* est évoqué 111 fois (uniquement dans notre corpus de discours spécialisé) et son coefficient d'implantation de 1 montre qu'aucune autre appellation n'est relevée pour dénommer ce concept. Contrairement à *canal spectral*, le coefficient d'implantation de *chatoiement* montre que l'hésitation à utiliser ce terme persiste (Fossat et Rouges-Martniez 1997 : 122).

5.1. Classification dirigée

L'utilisation du terme *classification dirigée* dans nos corpus permet de témoigner d'une différence des habitudes des spécialistes dans ce domaine. En effet, ce terme recommandé est absent de nos corpus, alors que Fossat et Rouges-Martinez constatent que « ce vocable est plus activé en communication écrite pour [nos] interlocuteurs » (1997 : 121). Par contre, *classification supervisée,* le calque de l'anglais *supervised classification* s'emploie pour dénommer ce concept dans nos corpus ; le calque en question a un coefficient d'implantation de 1. Au lieu de parler d'un changement d'habitudes, nous préférons de parler ici d'une différence d'habitudes, en raison des différences entre les spécialistes confirmés que Fossat et Rouges-Martinez ont consultés, et les jeunes chercheurs rédacteurs des thèses comprises dans nos corpus. Le corpus de GoogleBooks, contrairement à nos données, montre une rivalité entre les appellations *classification dirigée* et *classification supervisée*. Que nos corpus montrent une préférence absolue pour un terme et que le corpus de GoogleBooks montre une rivalité entre appellations constituent encore des pistes de réflexion et d'analyse plus en profondeur, car il est évidemment possible que le discours spécialisé de la télédétection aérospatiale diffère selon le type de texte écrit et que ces textes ne fassent pas partie de nos corpus.

5.2. Chatoiement

Selon Fossat et Rouges-Martinez (1997 : 122) le terme *chatoiement* « suscite une grande hésitation […] et un fort rejet » chez les experts, constat que nous validons, car le coefficient d'implantation de ce terme, 0,05 montre une nette préférence à utiliser d'autres termes pour dénommer son concept. De plus, *chatoiement* ne s'utilise pas dans des termes polylexicaux dans nos corpus. En revanche, c'est *tavelure* qui s'emploie à la place de *chatoiement* en tant que terme monolexical (le coefficient d'implantation de *tavelure* est 0,65 et sa fréquence s'élève à 6,43 occurrences par million de mots au sein de notre corpus de discours spécialisé) ainsi qu'au sein des termes polylexicaux, tels que *champ de tavelure* ou *interférométrie des tavelures*, où aucun terme polylexicaux comprenant *chatoiement* n'est relevé dans nos corpus.



Graphique 1 : chatoiement, tavelure, speckle, tachetis, tacheture, moucheture et mouchetis dans le corpus de GoogleBooks interrogé par le Ngram Viewer.

L'utilité limitée du NgramViewer en terminographie est de nouveau évidente lorsqu'on l'interroge à propos du terme *chatoiement* et de ses concurrents comme le montre le **Graphique 1** ; *scintillement* peut relever de la langue commune, et contrairement à ce que nous avons trouvé dans nos corpus, *chatoiement* s'utilise plus fréquemment que *tavelure,* ce qui indiquerait que ce vocable peut adosser de multiples charges sémantiques. L'utilité du NgramViewer se limite à attester que certains vocables, notamment *speckle, moucheture, mouchetis* et *tacheture* ne s'utilisent que très rarement, mais que l'autre concurrent pour dénommer le concept, *tachetis* est absent.

En dépit de cette limitation, il paraît que le NgramViewer peut se montrer utile lorsque l'on recherche des termes polylexicaux, car entre les termes possibles *interférométrie de tavelure*, *interférométrie de chatoiement* et *interférométrie de speckle*, c'est uniquement *interférométrie des tavelures* qui est attesté, ce qui s'accorde avec les données de nos corpus. Bien qu'introduisant du bruit (notamment *attaques de tavelure* et *taches de tavelure*, termes de la botanique) il nous semble que la recherche des unités qui n'ont qu'une acception terminologique, soit capable de dégager des pistes dignes d'intérêt qui pourraient se confirmer dans d'autres corpus plus traditionnels.

6. Conclusions sur GoogleBooks

En analysant la présence ou l'absence de chaque terme, ainsi que de ses concurrents, dans le corpus de GoogleBooks, il est possible de constater que ce corpus n'est pas incompatible avec la répartition en trois classes que Fossat et Rouges-Martinez proposent. Cela parce que les termes de chaque classe se différencient assez facilement : 29 termes sur les 30 termes de la Classe A (96,7 %) sont présents dans le corpus de GoogleBooks, alors que ce chiffre baisse à 13 termes sur 17 termes (76 %) dans la Classe B et 8 termes sur 12 termes dans la Classe C. De plus, entre les termes rélevés dans chacune des Classes, les différences sont nettes : dans la Classe A, 27 termes sur 30 termes (90 %) s'utilisent plus fréquemment que toute autre dénomination, alors que ce chiffre baisse à 11 termes sur 17 termes (65 %) dans la Classe B et seulement 4 termes sur 12 (25 %) termes dans la Classe C. Cela étant dit, les faiblesses de ce corpus entraînent la conclusion évidente que la présence d'un vocable dans ce corpus ne signifie pas l'utilisation du terme. Néanmoins, nous constatons, dans un corpus très large, que certains vocables s'utilisent, que ce soit fréquemment (*pixel*), rarement (*chatoiement*) ou pas du tout (*image d'écran*).

La consultation de ce corpus a aussi permis de témoigner de certaines rivalités non attestées dans nos corpus, notamment entre *classification dirigée* et *classification supervisée* ainsi qu'entre *signature spectrale* et *réponse spectrale*. Bien qu'il soit impossible de dire si ces rivalités relèvent du discours spécialisé du domaine en question, elles représentent cependant des pistes qui seraient dignes de recherches complémentaires. De plus, lorsqu'on l'interroge avec des unités polylexicales qui n'ont qu'une acception terminologique, tels que *interférométrie de tavelure, interférométrie de speckle* ou *interférométrie de chatoiement*, le NgramViewer permet de dégager des pistes, et même de refléter ce qui a été trouvé dans des corpus, bien que risquant d'introduire du bruit dans certains cas.

7. Conclusion sur nos corpus

Concernant nos corpus, les coefficients d'implantation des Classes A et B sont relativement similaires, les moyennes étant respectivement de 0,85 et 0,64 tandis que la Classe C se différencie avec une moyenne beaucoup plus basse de 0,1. Toutefois, les fréquences moyennes sont de 1,90 occurrence par million de mots dans la Classe A, de 0,26 occurrence par million de mots dans la Classe B et de 0,11 occurrence par million de mots dans la Classe C.

Par conséquent, la Classe A se démarque de par sa fréquence largement supérieure aux autres. Bien que constituant trois classes généralement distinctes, il est plus difficile de généraliser pour chaque classe comme un tout, car chaque terme est assez particulier.

Quoi qu'il en soit, l'examen de l'implantation des termes en corpus a permis de constater certaines évolutions, notamment que l'hésitation à utiliser certains termes a disparu, et ce entre les termes qui étaient en bonne voie d'implantation selon Fossat et Rouges-Martinez, ainsi que parmi des termes très peu implantés. Pourtant, dans d'autres cas, nous soulignons que des hésitations persistent, notamment pour *spatiocarte*, qui peine à s'imposer face à *carte satellite*.

Nous avons démontré que dans certains cas, comme dans celui de *signature spectrale,* la politique d'aménagement terminologique porte ses fruits, que ce soit par l'implantation du terme recommandé, comme *signature spectrale* ou par l'implantation d'autres vocables français, tels que *copie d'écran* et *image sur écran* face à un équivalent anglais, mais que l'implantation d'un terme français n'est pas toujours garantie, comme nous montre le cas de *scanneur*.

Bien que les méthodes adoptées soient différentes de celles de Fossat et Rouges-Martinez, le fait que nos conclusions s'alignent montrent que, malgré les différences importantes (évoquées par Quirion : 30-31) entre analyser les attitudes et dires d'une terminologie (la méthodologie de Rouges-Martinez) et analyser le comportement linguistique lui-même (comme nous l'avons fait), nous avons montré qu'il est possible que ces deux méthodes d'analyse puissent les deux être utiles pour permettre une meilleure analyse et compréhension de l'implantation des termes.

Remerciements

L'auteur souhaite remercier Jean Soubrier et Pascaline Dury pour leurs commentaires et conseils sur un brouillon de cet article. Il souhaite également remercier Charlène Vandra, Nicolas Bouliteau et Anthony Chan-Ou-Teung pour leurs commentaires et relectures.

Références

Allony Fainberg, Y. 1983. « Linguistic and Sociodemographic Factors Influencing the Acceptance of Hebrew Neologisms. » Dans *International Journal of the Sociology of Language* 41. 9-40.

Allony Fainberg, Y. 1974 « Officiel Hebrew Terms for Parts of the Car : A Study of Knowledge, Usage and Attitudes ». Dans *International Journal of the Sociology of Language*. 1. 67-94.

Auger, Pierre. 1999. *L'implantation des officialismes halieutiques au Québec : essai de terminométrie*. Coll. « Langues et Sociétés » nº 37. Montréal : Office de la langue française.

Cabré Castellví, Maria Teresa. 2010. « La implantación de la terminología : conceptos, estragegias, recursos y algunos datos ». Dans *Jornadas de lenguajes de especilidad y terminología 25-26/01/2010*. [En ligne]. URL<http://www.ehu.eus/ehg/aurkezpenak/jar10/01_TCabre.swf> Consulté le 23/01/2015.

Chansou, Michel. 1997. « Étude d'implantation des arrêtés de terminologie - domaine : audiovisuel et publicité ». Dans Depecker, Loïc et Gina Mamavi (eds.). 1997. *La Mesure des mots - cinq études d'implantation terminologique. Rouen* : Publications de l'Université de Rouen. 133-233.

Chansou, Michel. 1994. « Évaluation d'une action de politique linguistique. Les travaux de la Commission ministérielle de terminologie de l'audiovisuel et de la publicité ». Dans *Terminologies nouvelles*. 12 : Implantation des termes officiels : Actes du séminaire de Rouen, déc. 1993. 107-111.

Commission générale de terminologie et de néologie. 2013. *Rapport annuel de la Commission générale de terminologie et de néologie*. Paris : Délégation générale à la langue française et aux langues de France.

Commission générale de terminologie et de néologie. 1998. « Vocabulaire de l'informatique ». Dans *Journal officiel de la République française* du 10 octobre 1998 15372-15375. NOR : CTNX9803169K.

Daoust, Denise. 1995. « Quelques facteurs sociodémographiques qui sous-tendent le choix des langues pour la terminologie en milieu de travail et leur interrelation avec des attitudes ». Dans *Meta : Translators' Journal*. 260-283. [En ligne]. URL<http://id.erudit.org/iderudit/002458ar> Consulté le 11/02/2015.

Depecker, Loïc et Gina Mamavi (eds.). 1997. *La Mesure des mots - cinq études d'implantation terminologique*. Rouen : Publications de l'Université de Rouen.

Delavigne, Valérie et François Gaudin. 1996. « À propos d'implantation terminologique. Questionner l'usage ou le sentiment linguistique ? ». Dans *Cahiers de linguistique sociale*. X. Le questionnement social. Pages 131-140. [En ligne] URL<https://hal.archives-ouvertes.fr/hal-00924220> Consulté le 23/01/2015.

Fossat, Jean-Louis et Josiane Rouges-Martinez. 1997. « Étude d'implantation des arrêtés de terminologie - domaine : télédétection aérospatiale ». Dans Depecker, Loïc et Gina Mamavi (eds.). *La Mesure des mots - cinq études d'implantation terminologique*. Rouen : Publications de l'Université de Rouen. 97-133.

Fossat, Jean-Louis et Josiane Rouges-Martinez. 1994. « Implantation terminologique en télédétection aérospatiale : méthodologie et mesure ». Dans *Terminologies nouvelles*. 12 : Implantation des termes officiels : Actes du séminaire de Rouen, déc. 1993. 123-135.

Gasquet, Évelyne et Isabelle Villebrun. 1994. « Observation de faits d'implantation terminologique en métallurgie ». Dans *Terminologies nouvelles*. 12 : Implantation des termes officiels : Actes du séminaire de Rouen, déc. 1993. 114-122.

Gaudin, François. 1992. « Enquête sur l'impact des arrêtés terminologiques auprès des magistrats ». Dans *La Banque des Mots*. 42. 59-66.

Gaudin, François et Valérie Delavigne. 1997. « L'enquête en terminologie : point de la question et propositions ». Dans *Terminologies nouvelles*. 12 : Implantation des termes officiels : Actes du séminaire de Rouen, déc. 1993. 37-42.

GoogleBooks Ngram Viewer. URL <http://books.google.com/ngrams> Consulté en février 2015.

Gouadec, Daniel. 1997. « Étude d'implantation des arrêtés de terminologie - domaine : informatique ». Dans Depecker, Loïc et Gina Mamavi (eds.). 1997. *La Mesure des mots - cinq études d'implantation terminologique*. Rouen : Publications de l'Université de Rouen. 235-493.

Gouadec, Daniel. 1994. « Implantation des termes officiels de l'informatique. Attestations, francisation, normalisation, assimilation ». Dans *Terminologies nouvelles*. 12 : Implantation des termes officiels : Actes du séminaire de Rouen, déc. 1993. 141-147.

Guespin, Louis. 1997. « Étude d'implantation des arrêtés de terminologie - domaine : Génie génétique ». Depecker, Loïc et Gina Mamavi (eds.). 1997. *La Mesure des mots - cinq études d'implantation terminologique*. Rouen : Publications de l'Université de Rouen. 1-45.

Karabacak, E. 2009. « Acceptance of Terminology Sanctioned by the Turkish Language Society : A Study of the Use of Economic Terms in Turkish Newspapers » dans *Terminology* 15 (2) : 145-178.

Martin, A. 1998. *Les mots et leurs doubles : étude d'implantation de la terminologie officialisée dans le domaine de l'éducation au Québec*. Coll « Langues et sociétés ». 36. Québec : Office québécois de la langue française.

Maurais, J. 1994. « Quelques aspects sociolinguistiques de l'implantation des décisions de normalisation terminologiques ». Dans *Actes du colloque sur la problématique de l'aménagement linguistique ; enjeux théoriques et pratiques*. 441-451. Chicoutimi : Office de la langue française.

Michel, J.-B. *et al.* 2010. « Quantitative Analysis of Culture Using Millions of Digitized Books ». Dans *Science*. 331 : 176-182. [En ligne] URL <https://www.sciencemag.org/content/ 331/6014/176.full> Consulté le 29 janvier 2014.

Ministère de la Culture. 1996. « Décret n° 96-602 du 3 juillet 1996 relatif à l'enrichissement de la langue française ». *Journal officiel de la République française*. 5 juillet 1996. 10169-10170. NOR : MCCB9600333D.

Ministère de la Culture et de la Communication. 2015. « Décret nº 2015-341 du 25 mars 2015 modifiant le décret nº 96-602 du 3 juillet 1996 relatif à l'enrichissement de la langue française. » *Journal officiel de la République française* du 27 mars 2015. Texte 71. NOR : MCCB1430182D.

Ministère de l'Éducation Nationale. 1982. « Terminologie de la télédétection aérospatiale ». *Dans Journal officiel de la République française* du 9 juin 1982 5408-5210 NC.

Ministère de l'Éducation Nationale. « Arrêté du 25 septembre 1984 relatif à l'enrichissement du vocabulaire de la télédétection aérospatiale ». Dans *Journal officiel de la République française* du 20 octobre 1984 9627-9635.

Ministère de l'Éducation Nationale. « Arrêté du 10 janvier 1986 relatif à l'enrichissement du vocabulaire de télédétection aérospatiale ». Dans *Journal officiel de la République française* du 17 janvier 1986. 829-834.

Ministère de l'Éducation Nationale. « Arrêté du 31 mars 1987 relatif à l'enrichissement du vocabulaire de la télédétection aérospatiale ». Dans *Journal officiel de la République française* du 17 avril 1987. 4373-4379. NOR : MENU8700201A.

Ministère de l'Éducation nationale, de la jeunesse et des sports. « Arrêté du 26 juillet 1988 relatif à l'enrichissement du vocabulaire de la télédétection aérospatiale ». Dans *Journal officiel de la République française* du 9 septembre 1988. 11520-11526. NOR : MENU8801216A.

Ministère de l'Éducation nationale, de la jeunesse et des sports. « Arrêté du 14 septembre 1990 relatif à la terminologie de la télédétection aérospatiale ». Dans *Journal officiel de la République française* du 26 septembre 1990. 11634-11640. MENZ9002179A.

Montané March, M. Amor. 2012. Terminologia i implantació : anàlisi d'alguns factors que influencien l'ús dels termes normalitzats de la informàtica i les TIC en llengua catalana. Thèse de doctorat de linguistique appliquée. Université Pompeu Fabra, Barcelona. [En ligne] URL <http://www.tdx.cat/bitstream/handle/10803/97295/tamm.pdf?sequence=1> Consulté le 18 juillet 2014.

Montané March, M. Amor et M. Teresa Cabré. 2013. « La implantación de la terminología oficial según el ámbito de uso de los textos ». *Actes de la IXe journée scientifique Realiter*. [En ligne] URL <http://www.realiter.net/wp-content/uploads/2013/09/MontanéCabre.pdf> Consulté le 14/04/2015.

Quirion, Jean. 2013. « Facteurs sociaux de la variation terminologique dans les enquêtes d'implantation terminologique- le cas du Québec dans les années suivant l'adoption de la Charte de la langue française ». Dans *Debate Terminológico*. 47-61. URL <http://seer.ufrgs.br/riterm/article/download/37171/24033> Consulté le 21/04/2014.

Quirion, Jean et Jacynthe Lanthier. 2006. « Intrinsic qualities favouring term implantation : verifying the axioms ». Dans Bowker, Lynne (dir.). *Lexicography, Terminology, and Translation. Text-based Studies in Honour of Ingrid Meyer.* Ottawa : University of Ottawa Press. 107-118.

Quirion, Jean. 2004. « État de la question sur la nature des facteurs d'implantation terminologique ». Dans *Actes du Colloque international sur la traduction : Traduction et Francophonie. Traduire en Francophonie*. 193-200. URL <http://www.colloque.net/archives/2003/volume_1/quirion.pdf> Consulté le 21/04/2014.

Quirion, Jean. 2003. « Methodology for the design of a standard research protocol for measuring terminology usage ». Dans *Terminology*. 9 :1. 29-49.

Quirion, Jean. 2000. Aspects évaluatifs de l'implantation terminologique. Thèse de doctorat de linguistique. Université de Montréal. [En ligne] URL <http://www.collectionscanada.gc.ca/obj/s4/f2/dsk2/ftp03/NQ52119.pdf> Consulté le 7 février 2015

Thoiron, Philippe, Jean Iwaz et Nadine Zaouche. 1997. « Étude d'implantation des arrêtés de terminologie - Domaines : Santé et médecine ». Depecker, Loïc et Gina Mamavi (eds.). 1997. *La Mesure des mots - cinq études d'implantation terminologique*. Rouen : Publications de l'Université de Rouen. 47-95.

Thoiron, Philippe, Jean Iwaz et Nadine Zaouche. 1993. « Résultats de l'enquête d'implantation des termes de santé et de médecine ». Dans *Terminologies nouvelles*. 12 : Implantation des termes officiels : Actes du séminaire de Rouen, déc. 1993. 77-103.

Linguistics Terminology in Greek
Translational Equivalents from English to Greek

Foteini Kalamida*

*Linguistics Department, School of Philology
University of Patras
University Campus, Rio Patras 26504, Greece
kalamida@upatras.gr

Summary. The present study is an attempt to examine the adaptation of the English linguistics terminology into the Greek language. Certain English terms that are in common among a number of dictionaries/glossaries are selected; then, the translations of these terms in Greek are compared and contrasted in order to find cases of terms that differ in their translation in the Greek language. So as to arrive to the optimal translation in Greek, the terms are evaluated using mainly the macro-rules proposed by Xydopoulos (2002). The main conclusion is that there is a need for a unified linguistics terminology in Greece since there are cases of English terms that can possibly cause problems, ambiguities and confusion to the users as, sometimes, their translations in Greek differ. Therefore, it is necessary to apply and follow the above macro-rules with systematicity so as to translate the linguistics terminology correctly and accurately.

1. Introduction

In the present paper[1], I attempt to examine the adaptation of the English linguistics terminology[2] into the Greek language. It appears that there are sev-

1 I would like to thank the organizing committee/reviewers and the audience of TOTh 2015-"Terminologie et ontologie: Théories et applications" for their constructive remarks, feedback and comments (Chambéry, France, 4-6 June, 2015).

2 The translational equivalents of the English terms are chosen not only for scholarly use, namely for linguists or people who have linguistic expertise but also for non-professionals/non-experts who wish to learn more about the linguistics terminology; the latter can do so by consulting the definitions of the terms in order to understand them better.

eral English linguistic terms, which are translated in a totally different way in the Greek language, therefore, it often happens that a single English term may have two or more diverse translational equivalents in the Greek language. In order to achieve such a task, I will apply a set of evaluation criteria with ultimate aim to find and propose the best and most accurate translation of each English term in Greek.

As for my own growing interest concerning the translation of the linguistics terminology, it began approximately in 2013, when I was first given the opportunity to see in progress the adaptation of the foreign linguistics terminology of the 6th edition of *A Dictionary of Linguistics and Phonetics* (Crystal, 2008) into Greek. By working upon this dictionary, I have noticed that there were some terms which have not yet been standardized in Greece by the Greek linguists and consequently, each linguist may use diverse Greek translations so as to refer to the same English term; besides, I have also found instances of terms, which have been standardized, but still, there are people who use other translations apart from the standardized one and this can easily create ambiguities. Finally, except for these terms, there were also instances of terms included as new entries in the above dictionary and they do not have any translation in Greek, up to now at least, so, the question that arises is how these entries will be translated in Greek. Taking all the above into account, it appears that is of great concern in the field of linguistics to work upon the unification of the Greek linguistics terminology.

Consequently, the basic part of this paper has to do with the checking of the linguistics terminology adaptation in Greek; the main aim is, after comparing and contrasting the translations of a number of terms by consulting certain dictionaries, to, finally, arrive to the best translation by applying certain evaluation criteria.

2. A few words about Terminology

To begin with, Dubuc (1985:19) explains that terminology is the art of finding, analyzing and then, creating a dictionary that represents a certain field so that it can cater for the speakers' needs and facilitate their communication. This way, terminology seems to be a practical activity that collects, classifies and translates terms.

Motsiou (1985:295-296), a Greek linguist, describes terminology as a code; she holds the view that terminology is an artificial and conventional system of terms, a metalanguage, whose purpose is to facilitate the exchange

of thoughts and finally, to ensure the objectivity, the accuracy and the unification of the modes of expression in a particular scientific domain. Each term should have one meaning, that is, it should denote only one certain concept, and it should be stylistically neutral and independent form the context.

Later on, Sager (1994 :8-10) notes that terminology combines elements both from linguistics and epistemology for its theoretical basis. Although terminology, in the widely used dictionaries, is regarded as a total of terms under a specific thematic framework, it nowadays includes also the semantic processing of the terms and the determination of the relationship among them. Besides, terminology, for Sager, (1990) is not a separate discipline as it does not have its own epistemology ; nevertheless, he acknowledges the significance of the principles and methods used in order to achieve the purposes of the terminology processing, that is the facilitation of communication among specialists.

According to another well-known Greek linguist, Mpampiniotis (1993), terminology is considered to be, par excellence, a kind of metalanguage, a collection of linguistic signs, words, and/or phrases which help to achieve specialized forms of communication between scientists, e.g. doctors, linguists, etc. For example, the metalanguage of linguistics includes the linguistic terminology and everything relevant to this (e.g. linguistic expressions); it is about a rather closed system of communication among scientists of a particular sector as opposed to the metalanguage used by other professionals, which appears to be simpler and closer to the everyday usage of each language.

Some years later, Cabré (1999 :1) states that terminology is the discipline that has to do with the study and compilation of specialized terms. More specifically, Cabré (1999 :32) notes that the term terminology refers to three different concepts ; the first concept refers to the whole "field of terminology", namely, the principles and conceptual bases that govern the study of terms. The second concept refers to the methodology, the guidelines used in the terminographic work. Finally, the third concept refers to the set of terms of a particular and special subject. For Cabré (1999 :32), terminology is a field with its own theoretical principles (terminological theory) and its own applied purposes (the writing of vocabularies, glossaries, dictionaries, and the standardization of designations), although its autonomy as an independent discipline has been questioned. Therefore, terminology is an interdisciplinary field of enquiry and deals with the specialized vocabulary that belongs to specific domains of usage (Cabré, 1999 :32).

3. Linguistics Terminology in Greece

3.1. Origins of Linguistics Terminology in Greece

In general, although Greek has been the leading source language for creating many terms in diverse languages, the rapid changes in the scientific and academic fields and the dominance of some communities in the creation of scientific products or products of thought have caused some rather "minor" languages, such as Greek, to follow these terminological developments (Floros & Grammenidis, 2012 :86-87).

As for the linguistics terminology across languages, it does have indeed a cross-linguistic character since it is transferred from language to language through the diffusion of science. This terminology originates either from the same language or from external or internal borrowing (Anastassiadi-Simeonidi (1986 :63-65), Cabré (1999 :88-89), Giannoulopoulou (2001 :23-24), Xydopoulos & Pavlakou (2009 :1065)).

More specifically, the linguistics terminology in Greek derives from the traditional terminology of the literary and historical studies concerning language (Xydopoulos 2002 :496). This terminology is mostly of Greek origin and it has been transported as such into other languages, especially in English (e.g. *larynx*, *morpheme*). However, nowadays, a great part of the linguistics terminology is mostly a product of research activity carried out and diffused through the English language ; as a consequence, the modern linguistics terminology in Greek is to a large extent imported through loan translations, especially from the English language (Xydopoulos, 2002 :496).

3.2. Lexicographic work on the Linguistics Terminology in Greece

The first Greek dictionary of linguistics terminology is the *Dictionary of Linguistic Terms* (Dimitriou, 1983). From then on, there have been several attempts to adapt the English terminology into Greek. To mention just a few, notable are the following dictionaries : *A Dictionary of Linguistics and Phonetics* by Crystal (4th & 6th edition, translated in Greek by Xydopoulos (2008 ; forthcoming)), the *Online Glossary of Linguistics Terminology* (Xydopoulos, 2007), the *Contemporary Dictionary of Linguistic Terms and Topics* (Sakellariadis, 2004), etc.

Nevertheless, through a lot of articles/essays/textbooks about linguistics terminology in Greek, one thing is for sure, that the writers have always refer-

red to the difficulties they encountered when trying to translate and standardize the linguistic terms in Greek. These difficulties arise because of the fact that there seems to be a lack of a consistent and unified usage of the Greek linguistics terminology, since diverse translations of a great number of English terms exist in Greek. As a consequence, users often feel confused as to which Greek translation is the most accurate and the most appropriate one.

4. Principles and Criteria in the Evaluation of the Linguistics Terminology

Generally, in order to evaluate the linguistics terminology there have been a lot of suggestions (see among others, Mpampiniotis (1993; 1994), Sager (1998a; 1998b), Kakridi-Ferrari (2001), Xydopoulos (2002; 2003; 2011), Tsakona (2007), Hellenic Society for Terminology (2010)). In the two sections that follow, I am going to present in brief, firstly, the principles and the criteria in the evaluation of the linguistics terminology suggested by Mpampiniotis (1993; 1994); then, I am going to present the four macro-rules set by Xydopoulos (2002; 2003; 2011) which are based on Mpampiniotis' criteria and which I took into consideration in order to evaluate the translation of the linguistic terms from English, the source language, to Greek, the target language.

4.1. Principles and Criteria suggested by Mpampiniotis (1993; 1994)

Mpampiniotis (1993; 1994:36-46) emphasizes the importance of setting some criteria in order to evaluate the scientific terms. These evaluation criteria are also the prerequisites for the terms to function linguistically and correspond to several communicative purposes. These criteria are the following:

a) the criterion of acceptability: it shows the extent to which the term is acceptable to the reader / listener. Namely, to what degree each term is used in an acceptable way in the texts and how this term can be satisfactory for the user who has already formed his/her linguistic sense;

b) the criterion of informativity: it means the extent to which a term is acceptable to the user;

c) the criterion of retrievability: it signifies how easily users can recall a term, and also, it means whether a term needs a lot of effort to be learnt or not;

d) the criterion of translatability : it indicates the extent to which a term can be translated to other languages and how easy it is to recognize from which foreign term the Greek term arises.

4.2. Macro-rules set by Xydopoulos (2002 ; 2003 ; 2011)

The methodological scheme proposed by Xydopoulos (2002 ; 2003 ; 2011) consists of four macro-rules, which are strongly associated with the four criteria proposed by Mpampiniotis (1993 ; 1994). Regarding these macro-rules, Xydopoulos (2003 :750) mentions that they help in the translation of the terms as they take into account morphological and semantic criteria as well as the scientific and pragmatic circumstances of usage of each term. So, the four macro-rules include the following steps :

a) checking the scientific term and its definition

b) checking the standardization degree of the term

c) checking the linguistic mechanisms of productivity

d) avoidance or elimination, as much as possible, of any ambiguities.

5. Dictionaries/glossaries used so as to compare and contrast the translation of the linguistic terms & Methodology

So as to choose which terms I would discuss for the needs of the present study, I consulted a number of dictionaries/glossaries. The main reference points of my work were :

1) the *Dictionary of Linguistics and Phonetics* by Crystal, 4th edition (2003/2008) (translated in Greek by Xydopoulos),

2) the *Dictionary of Linguistics and Phonetics* by Crystal (in English) (2008),

3) the *Dictionary of Linguistics and Phonetics* by Crystal, 6th edition (2008/forthcoming) (translated in Greek by Xydopoulos),

4) The *Contemporary Dictionary of Linguistic Terms and Topics* by Sakellariadis (2004),

5) The most recent *Glossary by the Hellenic Society for Terminology* ((ELETO, 2014),

6) The *Online Glossary of Linguistics Terminology* by Xydopoulos

(OGLT, 2007).

Then, I collected as many terms[3] as possible that were in common among the above dictionaries. Then, I tried to compare and contrast the translations of the terms; namely, I mean that I tried to find cases of terms that differ a lot in their translation in the Greek language. Furthermore, except for the common terms, I found some new entries in the 6th edition of Crystal's dictionary (2008) which were not included in the previous edition, therefore, they were not translated in Greek before. So, I collected a certain number of these and I tried to propose the best translation I could, taking into account and applying the macro-rules discussed earlier in section 4.2.

6. Review of the Terminology Adaptation from English to Greek

6.1. Examples[4] of English terms that differ in their Greek translation

6.1.1. Term: Back-formation

• Definition:

Back-formation: "[…] an abnormal type of word-formation where a shorter word is derived by deleting an imagined affix from a longer form already present in the language" (Crystal, 2008:48-49).

• Translations in Greek:

Term	Crystal (2003/2008)	Crystal (2008/ forthcoming)	OGLT (Xydopoulos, 2007)	Sakellariadis (2004)	ELETO (2014)
Back-formation	αναδρομικός σχηματισμός [anaðromi'kos sximati'smos]	αναδρομικός σχηματισμός [anaðromi'kos sximati'smos]	αναδρομικός σχηματισμός [anaðromi'kos sximati'smos]	υποχωρητικός σχηματισμός [ipoxoriti'kos sximati'smos]	---

3 The common terms among these dictionaries were approximately 191.

4 Due to the limited space of the present article, I shall provide only a certain number of examples of terms; for more examples, see Kalamida (2015).

- No Standardized translational equivalent in Greek.
- Productivity of the Greek linguistic terms:

1) αναδρομικός-ή-ό, αναδρομή, αναδρομικότητα, ανάδρομος, αναδρομικά

2) υποχωρητικός-ή-ό, υποχωρώ, υποχώρηση, υποχωρητικότητα

- Definitions of the Greek linguistic terms:

1) αναδρομικός: "used for a word that is formed from another word in analogy to a respective scheme, giving the impression that it is the basis of the formation" (DMG[5], 1998/2005).
2) υποχωρητικός: "used for somebody who easily retreats/ recedes" (DMG, 1998/2005).

- Taking all the above into account, I suppose that the Greek term *υποχωρητικός σχηματισμός* does not render exactly the English term *back-formation,* therefore, in my opinion, the optimal Greek translational equivalent for the term *back-formation* is *αναδρομικός σχηματισμός*.

6.1.2. Term: Blending

- Definition:

Blending: "A process […] in which two elements which do not normally co-occur, according to the rules of the language, come together within a single linguistic unit (a blend)" (Crystal, 2008:57).

- Translations in Greek:

Term	Crystal (2003/2008)	Crystal (2008/ forthcoming)	OGLT (Xydopoulos, 2007)	Sakellariadis (2004)	ELETO (2014)
Blending	αμαλγάμωση [amal'γa-mosi]	1) αμαλγάμωση [amal'γamosi] 2) μίξη ['miksi] 3) σύμφυρση ['simfirsi]	1) αμαλγάμωση [amal'γamosi] 2) σύμφυρση ['simfirsi]	σύμφυρση ['simfirsi]	σύμμειξη ['simiksi]

5 Abbreviation used for: Mpampiniotis, G. 1998/2005. *Λεξικό της Νέας Ελληνικής Γλώσσας* [ΛΝΕΓ]. *Dictionary of Modern Greek [DMG]* (2nd edn., 2nd printing 2005). Athens: Lexicology Centre. [in Greek].

- No Standardized translational equivalent in Greek.
- Productivity of the Greek linguistic terms:

1) σύμφυρση, συμφυρμός, σύμφυρμα, συμφυρματικός, συμφύρω

2) αμαλγάμωση, αμάλγαμα

3) σύμμειξη - -

- Definitions of the Greek linguistic terms:

1) αμάλγαμα: "denoting two or more meanings from the same morpheme" (DMG, 1998/2005:131).
2) σύμφυρση: […] "two cognate linguistic elements blend, namely, they interconnect in the speaker's mind and they result in one" (DMG, 1998/2005:1692).
3) σύμμειξη: - (not found).

- Taking all the above into consideration, it seems to me that the optimal Greek translational equivalent for the term *blending* is *σύμφυρση* since *σύμφυρση* means mixing, and *blending* is exactly about mixing two elements.

6.1.3. Term: Cleft Sentence

- Definition:

Cleft sentence: "A term used in grammatical description to refer to a construction where a single clause has been divided into two separate sections, each with its own verb, one of which appears in a dependent *wh*-clause (relative clause)" (Crystal, 2008:79).

- Translations in Greek:

Term	Crystal (2003/2008)	Crystal (2008/forthcoming)	OGLT (Xydopoulos, 2007)	Sakellariadis (2004)	ELETO (2014)
Cleft sentence	- δίπτυχη - πρόταση ['ðiptixi 'protasi]	- δίπτυχη - πρόταση ['ðiptixi 'protasi]	- δίπτυχη - πρόταση ['ðiptixi 'protasi]	ρηγματική πρόταση [riγmati'ki 'protasi]	--

- No Standardized translational equivalent in Greek.

• Productivity of the Greek linguistic terms:

1) δίπτυχος-η-ο --

2) ρηγματικός-ή-ό, ρήξη, ρήγμα, ρηγματώδης-ης-ες, ρυγνύω, ρηγμάτωση

• Definitions of the Greek linguistic terms:

1) δίπτυχος: "[…] whatever is comprised of two parts" (DMG, 1998/2005:515).

2) ρηγματικός: "split the unity of a set" (DMG, 1998/2005: 1539).

• Considering both the two definitions of the Greek terms above and the fact that a *cleft sentence* includes a main clause and a dependent clause and it has a meaning that could be expressed by a simple sentence, I would say that the first translation, namely *δίπτυχη πρόταση* is more correct so as to translate the English term *cleft sentence* since it describes something that consists of two parts; instead, the second Greek term, namely *ρηγματική πρόταση*, seems to imply separation and I suppose that separation may not be the case in a *cleft sentence*, as the main aim is to put emphasis on a particular constituent.

6.1.4. Term: Output

• Definition

Output: "1. A use of the general sense of the term in generative grammar to refer to a sentence which is produced after the application of a rule or set of rules […], 2. In optimality theory, an output is the optimal representation of an input form, following the application of the mechanisms of the theory" (Crystal, 2008:345).

• Translations in Greek:

Term	Crystal (2003/2008)	Crystal (2008/ forthcoming)	OGLT (Xydopoulos, 2007)	Sakellariadis (2004)	ELETO (2014)
Output	Εξαγόμενο [eksa΄γomeno]	εξαγόμενο [eksa΄γomeno]	εξαγόμενο [eksa΄γomeno]	1) εξαγόμενο [eksa΄γomeno] 2) έξοδος [΄eksoðos]	έξοδος [΄eksoðos]

• No Standardized translational equivalent in Greek.

• Productivity of the Greek linguistic terms:

1) εξαγόμενο, εξαγόμενος-η-ο, εξάγω, εξαγωγή, εξαγωγέας, εξαγωγικός-ή-ό,

εξαγώγιμος-η-ο

2) έξοδος, εξοδούχος

- Definitions of the Greek linguistic terms:

1) εξαγόμενο: "[…] something that arises as a consequence […]" (DMG, 1998/2005: 620).

2) έξοδος: "[…] the opening or the way that leads outside, the movement from inside to outside […]" (DMG, 1998/2005: 630-631).

- Although both translations, namely *εξαγόμενο* and *έξοδος*, seem to be appropriate and they translate correctly the English term *output*, I suppose that since *έξοδος* is already used so as to translate the English term *coda* (Crystal, 2003/2008:493), then, in order to eliminate such an ambiguity, it would be better to avoid this translation. Besides, another fact in favor of the translation *εξαγόμενο* is that analogically to the translation of the term *input* as *εισαγόμενο* (Crystal, 2008/forthcoming), perhaps, *εξαγόμενο* is more appropriate to translate the term *output.*

6.2. Examples of English terms that do not have a translational equivalent in Greek (up to now):

6.2.1. Term: Attract

- Definition:

Attract: "In the minimalist programme, a term which identifies the driving force behind move" (Crystal, 2008:43).

- Possible Greek translational equivalents for attract:

Term	Collins (1997:46)	Stavropoulos & Hornby (1977/1997:29)
Attract	1) προσελκύω [prosel'kio] 2) αποκτώ [apo'kto] 3) τραβάω [tra'vao] 4) γοητεύω [γοi'tevo]	1) έλκω ['elko] / ελκύω [el'kio] 2) προσελκύω [prosel'kio]

- Since the definition of *attract* is about a driving force behind move, then, I suppose that the optimal translational equivalents are the terms *ελκύω / έλκω / προσελκύω* which have exactly this meaning, namely, *ελκύω / έλκω* mean "drag, pull something towards me"(DMG, 1998/2005:588) and *προσελκύω* means "pull / bring something towards me […]" (DMG, 1998/2005:1490).

6.2.2. Term: Frication

- Definition:

Frication: "In acoustic phonetics, the turbulent noise produced by the vocal organs engaged in the production of fricatives and other consonants involving a "constricted airflow" (Crystal, 2008:199).

- Possible Greek translational equivalents for frication:

Term	Crystal (2003/2008)	Stavropoulos & Hornby (1977/1997:226)
Frication	--	--
Friction	τριβή [tri'vi]	τριβή [tri'vi] τρίψιμο ['tripsimo]
Fricative	τριβόμενος-η-ο ['trivome-nos-i-o]	--

- In order to avoid any ambiguity, the term *frication* cannot be translated as the term *friction*, namely as *τριβή*. Hence, I suppose the term *frication* will surely contain the stem *τριβ-* in Greek; as for the second constituent, that is *–ation*, according to the Online Etymology Dictionary (2001-2014), it is "a word-forming element for making nouns of action". Therefore, it could be translated, in my opinion, as *τριβοποίηση* (analogically to *lexicalisation*: *λεξικοποίηση* (Xydopoulos, 2008:369)).

6.2.3. Term: Spike

- Definition:

Spike: "In acoustic phonetics, a burst of acoustic energy with an abrupt onset and very short duration" (Crystal, 2008:448).

- Possible Greek translational equivalents for spike:

Term	WordReference (2014)	Stavropoulos & Hornby (1977/1997 :659)
Spike	1) ακίδα [a'kiða] 2) σουβλί [su'vli] 3) απότομη άνοδος [a'potomi 'anoðος] 4) καρφί [kar'fi]	1) μύτη ['miti] 2) καρφί [kar'fi] 3) ακίδα [a'kiða] 4) στάχυ ['staxi]

- As it is mentioned in the definition of the term *spike*, it has an abrupt onset ; therefore, taking this into account, I suppose that the translational equivalent *απότομη άνοδος* is acceptable and suitable for the term *spike* since it has the meaning of *sudden ascent* and it has to do with a burst of acoustic energy and, usually, a burst brings a rise.

6.2.4. Term : Stage

- Definition :

Stage : "In semantic theory, a temporal part of an individual. The individual is seen as existing over a limited span of time, not typically including its entire lifetime" (Crystal, 2008 :450)

- Possible Greek translational equivalents for stage :

Term	Collins (1997 :776)	Stavropoulos & Hornby (1977/1997 :668)
Spike	1) σκηνή [ski'ni] 2) εξέδρα [e'kseðra] 3) στάδιο ['staðio]	1) σκηνή θεάτρου [ski'ni θe'atru] - 2) θέατρο ['θeatro] - 3) πεδίο δράσεως [pe'ðio ðraseos] - 4) φάση['fasi] - 5) στάδιο ['staðio] - 6) βαθμίδα [va'θmiða] - 7) τμήμα διαδρομής ['tmima ðia-dro'mis] - 8) στάση ['stasi]

- Taking the definition of *stage* into consideration and the Greek translations as well, I suppose that since *stage* refers to a temporal part of an individual, then, the best translations are *στάδιο* and *φάση*. The translation *φάση*, however, already exists in the linguistics terminology to describe the term *phase* (Crystal, 2003/2008 :507 ; OGLT). Hence, in order to avoid

any ambiguities caused since the term *phase* is translated as *φάση*, I suggest translating the term *stage* as *στάδιο*.

7. Conclusions

In this paper, I dealt with the linguistics terminology adaptation and I attempted to examine the Greek translations of a number of English linguistic terms. Taking into consideration all the information provided earlier about the Greek linguistics terminology, the most basic points that arise from this study are the ones that follow.

It appears, that there are instances of English terms, which are translated in a similar way in Greek, and, therefore, unification seems to exist. However, there are still cases of English terms that can possibly cause problems, ambiguities and confusion to the users since their translations in Greek differ. There are also the new entries from Crystal (2008) that do not have yet any translational equivalent in Greek. Thus, there is a need for a unified linguistics terminology in Greece. I suppose that this can be achieved by applying and following the macro-rules, mentioned earlier, with systematicity so as to translate the linguistics terminology correctly and accurately.

References

A. Dictionaries

Collins, H. (1997). *English Greek Dictionary*. Glasgow : HarperCollins Publishers.

Crystal, D. (2003/2008). *Λεξικό Γλωσσολογίας και Φωνητικής. A Dictionary of Linguistics and Phonetics*. (Translated into Greek by G. Xydopoulos.) Athens : Patakis. [in Greek].

Crystal, D. (2008). A *Dictionary of Linguistics and Phonetics* (6th edn.). Oxford : Blackwell.

Crystal, D. (2008/forthcoming). *Λεξικό Γλωσσολογίας και Φωνητικής. A Dictionary of Linguistics and Phonetics*. (Translated into Greek by G. Xydopoulos.) Athens : Patakis. [in Greek].

Dimitriou, S. (1983). *Dictionary of Linguistic Terms* (Volumes I and II). Athens : Kastaniotis. [in Greek].

Harper, D. (2001-2015). *Online Etymology Dictionary*. In http://www.etymonline.com/index.php ?allowed_in_frame=0&search=&search mode=-none [1.9.2015].

Mpampiniotis, G. (1998/2005). *Λεξικό της Νέας Ελληνικής Γλώσσας* [ΛΝΕΓ]. *Dictionary of Modern Greek* (2nd edn., 2nd printing 2005). Athens: Lexicology Centre. [in Greek].

Sakellariadis, G. (2004). *Σύγχρονο Λεξικό Όρων και Θεμάτων Γλωσσολογικών* [ΣΛΟΘΓ]. *Contemporary Dictionary of Linguistic Terms and Topics*. Athens: Savvalas. [in Greek].

Stavropoulos, D.N. & A.S. Hornby. (1977/1997). *Oxford English-Greek Learner's Dictionary* (15th impression). Oxford: Oxford University Press.

Valeontis, K. (ed.). (2014). "General and Specific terms of Terminology (4th edition): English-Greek-French Glossary". TEE – ELOT / Technical Committee 21 *"Orologia – Orologikoi poroi*", a collective member of ELETO. Athens, 40- 75. In http://www.eleto.gr/download/Bodies/TE21-TerminologyTerms_GR.pdf [1.9.2015]. [in Greek].

WordReference.com. (2014). *Online Language Dictionaries: Αγγλοελληνικό λεξικό*. In

http://www.wordreference.com/engr/a [1.9.2015].

Xydopoulos, G. J. (ed.). (2007). *Online glossary of linguistics terminology: Greek- English & English-Greek* [OGLT]. Ioannina: Department of Linguistics, University of Ioannina. In http://users.uoi.gr/gjxydo/lexicon/glossary.html [1.9.2015].

B. Other Literature

Anastassiadi-Symeonidi, A. (1986). *Neology in Modern Greek*. Yearbook of the Faculty of Philosophy of the Aristotle University of Thessaloniki, Volume 65. Thessaloniki. [in Greek].

Cabré, M. T. (1999). *Terminology: Theory, methods, and applications* (transl. J. DeCesaris). Amsterdam: John Benjamins.

Dubuc, R. (1985). *Manuel pratique de terminologie* [2nd edn.]. Brossard (Quebec): Linguatech.

Floros, G. & S. Grammenidis. (2012). "Secondary Term Formation in Greek: Theoretical and Methodological Considerations". In Cabré, T. *et al.* (eds) *Terminology* 18(1) *Special issue: Neology in Specialized Communication*, 86- 104.

Giannoulopoulou, G. (2001). "The contribution of Ancient Greek and Latin in the formulation of the Scientific Terminology of the European Languages: A historical study". Announcement at the 3rd Conference of ELETO *Greek Language and Terminology*, 19-29. [in Greek].

Kakridi-Ferrari, M. (2001). "Translation of Foreign Terms". In A.-F. Christidis (ed. in collaboration with M. Theodoropoulou) *An Encyclopedic Guide to Language.* Thessaloniki: Kentro Ellinikis Glossas, 203-205. [in Greek].

Kalamida, F. (2015). *A Lexicological Study of the linguistics Terminology in Greek* (Unpublished Master Thesis). University of Patras, Greece.

Motsiou, V. (1985). "For the Unification of the Greek Linguistic Terminology". *Proceedings of the 3rd Annual Meeting of the Department of Linguistics, Faculty of Philosophy, Aristotle University of Thessaloniki*, 26-28 April 1982. Thessaloniki, 295-305. [in Greek].

Mpampiniotis, G. (1993). "The linguistic side of scientific terms". *To Vima* 20.6.1993. [in Greek].

Mpampiniotis, G. (1994). "Scientific Metalanguage. The Scientific Terms of Greek". In G. Mpampiniotis (ed.) *Language as Value. The Example of Greek*. Athens : Gutenberg, 31-46. [in Greek].

Sager, J.C. (1994). Terminology : Custodian of knowledge and means of knowledge transfer. *Terminology* 1/1, 7-16.

Sager, J.C. (1998a). "Terminology, applications". In M. Baker (ed.) *Routledge encyclopedia of translation studies*. London : Routledge, 251-255.

Sager, J.C. (1998b). "Terminology, standardization". In M. Baker (ed.) *Routledge encyclopedia of translation studies*. London : Routledge, 255-258.

Tsakona, V. (2007). "Bilingualisation in practice : Terminological issues in bilingualising a specialised glossary". *International Journal of Lexicography* 20 (2): 119-145.

Valeontis, K. (2010). "The internationally standardized interlingual 'principles of term formation' (according to ISO)". In *Orogramma*, 102. Maroussi, Greece, 05- 06/2010. [in Greek].

Xydopoulos, G. J. (2002). "Problems in Rendering English Linguistic Terms into Greek" in *Studies in Greek Linguistics*. Proceedings of the 24th Annual Meeting of the Department of Linguistics, Faculty of Philosophy, Aristotle University of Thessaloniki, 27-29 April 2001. Thessaloniki, 495-506. [in Greek].

Xydopoulos, G. J. (2003). "Translating a Scientific Textbook : difficulties and possible solutions". Proceedings of the Conference *Metafrazontas ston 20o Aiona : Taseis kai Prooptikes*. Aristotle University of Thessaloniki, 740-747. [in Greek].

Xydopoulos, G. J. (2008). *Lexicology : An Introduction to the Analysis of Words and Dictionaries*. Athens : Patakis. [in Greek].

Xydopoulos, G. J. (2011). "The Linguistics Terminology in Greek : The case of Semantics". In G.J. Xydopoulos, M. Mpaltatzani, A. Tsaggalidis & G. K. Giannakis (eds.) *12 meletes gia ti Glossologia*. Ioannina, 2004-2008. Athens : Kontili Publications, 125-140.

Xydopoulos, G. J. & M. Pavlakou. (2009). “The handling of the linguistics terminology in two monolingual contemporary dictionaries in Modern Greek : One first critical review”. In Baltazani, M., G. K. Giannakis, A. Tsangalidis & G. J. Xydopoulos *Proceedings of the 8th International Conference on Greek Linguistics* (Ioannina, Greece, 30 August-2 September 2007). [in Greek].

Résumé

La présente étude est une tentative d’examiner l’adaptation de la terminologie linguistique anglaise dans la langue grecque. Certains termes anglais qui sont en commun entre un certain nombre de dictionnaires sont sélectionnés ; puis, les traductions de ces termes en grec sont comparées et contrastées afin de trouver des cas de termes qui diffèrent dans leur traduction dans la langue grecque. Afin d’arriver à la traduction optimale en grec, les termes sont évalués en utilisant principalement les macro-règles proposées par Xydopoulos (2002). La conclusion principale est qu’il y a un besoin pour une terminologie linguistique unifié en Grèce car il y a des cas de termes anglais qui peuvent éventuellement causer des problèmes, des ambiguîtés et de la confusion aux utilisateurs puisque, parfois, leurs traductions en grec sont différentes. Par conséquent, il est nécessaire d’appliquer et de suivre les macro-règles avec systématicité de manière à traduire la terminologie linguistique correctement et avec précision.

Indices lexico-syntaxiques pour la reconnaissance des termes de la forme N_Adj dans les textes intégraux

Laurence Kister*, Évelyne Jacquey*, Mario Marcon*, Sabine Barreaux**

* Atilf UMR 7118 CNRS/Université de Lorraine - 44, avenue de la Libération – 54000 Nancy
Prénom.Nom@atilf.fr
** Inist-CNRS -2, allée du parc de Brabois – 54500 Vandœuvre-lès-Nancy
Prénom.Nom@inist.fr

Résumé : L'objectif de cette expérience est de confronter les résultats d'une extraction automatique de candidats-termes aux attentes annotateurs humains afin de proposer des indices de présélections des constructions *N_Adj* susceptibles d'acquérir le statut de terme.

L'extraction automatique de candidats-termes à partir de textes intégraux et leur évaluation font l'objet de recherche depuis les années 90 : (Daille 1996), (Bourigault *et al.* 2001 et 2004), (Drouin 2003 et 2004), (Bachimont *et al* 2005), (Kupsc 2007), entre autres. De nombreuses recherches relatives à la validation de candidats-termes extraits de textes intégraux présentent des listes de candidats, contextualisés ou non par des exemples d'usage, à des experts (Bourigault *et al.* 2001). Comme l'a montré (L'Homme 2004a), les occurrences de candidats-termes peuvent varier sur différents plans. Les variations morpho-syntaxiques et syntaxiques sont bien connues aujourd'hui et ont été intégrées dans différents outils d'extraction terminologique : Fastr (Jacquemin, 1997), Yatea (Aubin et Hamon, 2006), Acabit (Daille, 1994, 2003 ; Toussaint *et al.*, 1998), TermoStat (Drouin, 2003), TTC-TermSuite (Daille *et al.* 2011)[1]. Les variations sémantiques (Jacquemin, 1996) et les variations dues

1 Yatea - http://taln09.blogspot.fr/2009/03/description-lextracteur-de-termes-yatea.html (page consultée le 29/09/2015).
Acabit - http://taln09.blogspot.fr/2009/03/acabit-acquisition-de-termes-partir-de.html (page consultée le 29/09/2015).
Fastr - http://perso.limsi.fr/jacquemi/FASTR/ (page consultée le 29/09/2015).

aux différents emplois des formes peuvent correspondre à des occurrences de termes, mais la nature de leur emploi - terminologique ou de langue courante - constitue une question encore largement ouverte.

Pour ce qui nous concerne, nous cherchons à utiliser la reconnaissance automatique de termes dans des documents scientifiques pour contribuer à la mise au point d'une aide à l'indexation des textes qui les contiennent. Notre proposition se veut complémentaire de celles évoquées ci-dessus, dans la mesure où l'évaluation porte sur le caractère terminologique des différentes occurrences des candidats-termes contenues dans les résumés des articles à indexer. Pour cela, nous fondons nos travaux sur l'analyse d'un corpus de notices bibliographiques rédigées par l'Inist et rassemblées pour constituer un corpus de référence.

Nous examinerons les indices lexicaux-syntaxiques susceptibles de constituer des arguments en faveur de l'accession des *N_Adj* au statut de terme après avoir présenté le contexte de l'expérience.

1. Annotation sémantique et terminologique en sciences humaines et sociales : contexte

Ce travail s'inscrit dans le cadre du projet TermITH[2] (**Term**inologie et **I**ndexation de **T**extes en sciences **H**umaines) qui vise à l'annotation sémantique et terminologique de documents afin de faciliter l'indexation des documents par les termes qu'ils contiennent. Il associe six partenaires dont les activités gravitent autour du traitement de la langue, de l'informatisation de son traitement, de la terminologie et de l'indexation : Atilf UMR 7118 CNRS/ Université de Lorraine, Inist UPS 76 CNRS, Lidilem EA 609 Université Stendhal, Lina UMR 6241 CNRS/Université de Nantes/ École des mines de Nantes, Inria Nancy-Grand Est et Inria Saclay.

Les sciences du langage sont une discipline à laquelle les six partenaires sont familiarisés, aussi nous avons choisi cette discipline pour mettre au point la méthodologie. Pour évaluer la méthodologie, nous utilisons trois autres disciplines relevant des sciences humaines et sociales : les sciences de l'information et de la communication, la psychologie et l'archéologie. Une cinquième discipline qui ne relève pas des sciences humaines sociales a été

TTC-TermSuite - https://logiciels.lina.univ-nantes.fr/redmine/projects/termsuite (page consultée le 29/09/2015).

2 TermITH - ANR-12-CORD-0029 - Contint. (Contenus numériques et interactions) - http://www.atilf.fr/ressources/termith/ (page consultée le 29/09/2015).

retenue pour envisager la transposition de la méthodologie sur une discipline de science dite exacte, la chimie.

Nous cherchons à repérer les termes par recoupement d'informations telles celles apportées par la désambiguïsation sémantique des occurrences de termes, les **l**exiques **s**cientifiques **t**ransdisciplinaires (LST) et des ressources terminologiques (Vocabulaire de la linguistique de l'Inist et Thesaulangue[3]).

1.1. Extraction automatique des candidats-termes

Afin de disposer d'un ensemble de candidats-termes pour nos investigations, nous avons constitué un corpus de 150 résumés extraits des notices bibliographiques disponibles à l'Inist, par discipline. Le choix de sélectionner des documents déjà indexés humainement constitue une liste de contrôle : nous avons une première idée des contenus et des candidats susceptibles d'apparaître.

Les 750 notices ont été soumises à une extraction automatique des candidats-termes par TTC-TermSuite[4] qui présente l'avantage de traiter les documents simultanément avec deux extracteurs :

- TermoStat mis au point par Patrick Drouin (Observatoire de linguistique Sens-texte - Montréal), un outil d'acquisition automatique de termes par comparaison de corpus spécialisés et non-spécialisés. Il analyse les documents fournis en entrée au format *texte brut* pour restituer une liste de candidats-termes. Les documents sont étiquetés par TreeTagger. Les candidats-termes proposés peuvent être mono-lexicaux ou poly-lexicaux et sont accompagnés d'un score en fonction de leur fréquence et de leur coefficient de spécificité dans le corpus analysé et dans un corpus prétraité, le corpus de référence.
- Acabit mis au point par Béatrice Daille (Lina – Nantes) permet l'acquisition de termes sur un corpus pré-étiqueté et désambiguïsé en deux étapes : une analyse linguistique doublée d'un regroupement de variantes (analyse par des transducteurs et production de candidats-termes binaires) et un filtrage statistique (tri des candidats- termes à partir d'un corpus de référence et de termes valides associé à la comparaison de mesures statistiques en fonction de leur capacité à distin-

3 Thesaulangue est un thesaurus des sciences du langage développé et maintenu par le centre de documentation de l'atilf.

4 TTC-TermSuite : https://code.google.com/p/ttc-project/ (page consultée le 29/09/2015).

guer les unités linguistiques des séquences ne correspondant pas à des termes).

1.2. Annotation humaine

Les occurrences de candidats-termes obtenues automatiquement font l'objet d'une évaluation et d'une validation humaine. Cette procédure d'annotation humaine est destinée à distinguer les occurrences de candidats-termes qui font l'objet d'un emploi terminologique de celles qui font l'objet d'un emploi en langue courante :

le sujet$_{[terme]}$ *de la phrase…* vs. *le sujet*$_{[non\text{-}terme]}$ *de cet article concerne…*

La méthodologie que nous employons (Kister et Jacquey 2012), comme celle développée par (Delavigne 2001) sur des extraits de textes de vulgarisation scientifique dans le domaine de l'énergie nucléaire, applique successivement différents critères de différenciation aux candidats-termes grâce à une interface de pastillage, la plateforme Smarties[5] :

- validité syntagmatique du candidat-terme
- appartenance du candidat-terme au lexique scientifique disciplinaire
- emploi terminologique du candidat-terme dans sa discipline d'apparition

1.2.1. La plateforme Smarties

Smarties est une interface de visualisation et de validation humaine des résultats obtenus après soumission des documents à TTC-TermSuite. Les candidats-termes sont projetés dans les documents et apparaissent entre crochets précédés d'une pastille. Le passage de la souris sur une pastille entraîne un changement de couleur du candidat pour faciliter la prise de décision.

Après projection, toutes les pastilles sont vertes et l'annotateur les fait passer au rouge quand il ne valide pas le découpage syntagmatique proposé, l'appartenance au vocabulaire scientifique disciplinaire ou l'emploi terminologique. Après chaque étape de validation, le document subit un traitement qui élimine les pastilles rouges et les crochets qui encadrent les candidats non validés par l'annotateur.

5 Smarties - https://apps.atilf.fr/smarties/ (page consultée le 29/09/2015)

La première tâche de l'annotateur est de valider les différents découpages syntagmatiques afin de conserver les candidats-termes complexes mais aussi les composants simples de ces candidats-termes. Cette première étape d'annotation reste relativement longue car les candidats-termes proposés sont nombreux.

→●[●[structures] à ●[●[attribut]] de l'●[objet]]
●[●[**structures**] **à** ●[●[**attribut**]] de l'●[objet]]
→●[●[structures] à ●[●[attribut]] de l'●[objet]]
↘
●[●[structures] à ●[●[**attribut**]] **de l'**●[**objet**]]
↘ ↘
●[●[structures] à ●[●[**attribut**]] de l'●[**objet**]]

Après élimination des pastilles rouges et des crochets encadrant les candidats-termes non validés, nous obtenons la suite :

● [●[structures] à ●[●[attribut]] de l'●[objet]]

qui correspond à la série de candidats-termes syntagmatiquement valide qui est examinée à l'étape suivante.

Fig. 1 –Titre et premier paragraphe du texte 394 après intervention de l'annotateur pour la validation syntagmatique

La seconde tâche d'annotation consiste à valider l'appartenance des candidats-termes au vocabulaire scientifique disciplinaire.

→●[**structures**] à ●[●[attribut] de l'●[objet]]
↘
●[structures] à ●[●[**attribut**] **de l'**●[**objet**]]
↘ ↘
●[structures] à ●[●[**attribut**] de l'●[**objet**]]
●[structures] à ●[●[attribut] de l'●[objet]]

Pour l'exemple ●[structures] à ●[●[attribut] de l'●[objet]], il n'y a pas de modification au cours de la seconde étape. L'annotateur considère tous les candidats-termes issus de la première étape de validation comme des candidats-termes appartenant au vocabulaire scientifique disciplinaire qu'ils soient mono-lexicaux (structures, attribut, objet) ou poly-lexicaux (attribut de l'objet).

Fig. 2 –Titre et premier paragraphe du texte 394 après intervention de l'annotateur pour la validation appartenance au vocabulaire disciplinaire

La dernière étape de l'annotation consiste à examiner le type d'emploi : les candidats-termes peuvent faire l'objet d'un emploi terminologique, d'un

emploi en langue courante ou relever du LST (S. Hatier 2014). À cette étape l'annotateur sélectionne et valide les candidats-termes qui font l'objet d'un emploi terminologique.

Pour l'exemple que nous avons choisi, le candidat *structure* est jugé non spécifique au vocabulaire des sciences du langage car il appartient au LST, il n'est donc pas retenu par l'annotateur.

→● **[structures]** à ●[●[attribut] de l'●[objet]]
●[structures] à ●[●[attribut] de l'●[objet]]
structures à ●[●[attribut] de l'●[objet]]
↘
structures à ●[●**[attribut] de l'●[objet]**]
↘ ↘
structures à ●[●[attribut] de l'●[objet]]
structures à ●[●[attribut] de l'●[objet]]

Le résultat de l'annotation sur cet exemple nous donne trois termes possibles : deux mono-lexicaux (*attribut* et *objet*) et un poly-lexical (*attribut de l'objet*).

Fig. 3 –Titre et premier paragraphe du texte 394 après intervention de l'annotateur pour la validation présence d'un emploi terminologique

La fig. 3 montre que des nombreux candidats-termes proposés à l'issue de l'extraction automatique seuls deux candidats poly-lexicaux (*attribut de l'objet* et *grammaire constructionnelle*) et deux candidats mono-lexicaux (*construction* et *allostructions*) ont été retenus par l'annotateur humain.

2. Résultats de l'annotation/validation humaine

Lors du travail de validation, les annotateurs[6] ont repéré des constructions qui pourraient constituer des termes qui n'ont pas été proposés par TTC-TermSuite. Par exemple, pour la construction que nous avons évoquée ci-avant, ils considèrent qu'il existe une structure complexe qui pourrait fonctionner comme un terme en raison de sa fréquence d'apparition dans les textes du domaine : **[**structures à attribut de l'objet**]**

Les annotateurs ont relevé des constructions linguistiques qu'ils utilisent habituellement comme des termes dans leur domaine d'expertise et ont fait un relevé systématique de ces constructions que nous appelons les termes-attendus. À l'issue de la validation de l'emploi terminologique des candidats-termes, les annotateurs ont proposé 1583 termes-attendus pour les cinq disciplines :

- 1057 poly-lexicaux (66.6%) qui comportent deux noms ou un nom et un adjectif
- 259 (16,4%) poly-lexicaux qui comportent plus de deux composants nominaux et/ou adjectivaux que nous laissons de côté dans cette expérience - les non-examinés[7]
- 267 mono-lexicaux (17%).

Les différents comptages que nous avons effectués font apparaître que les termes-attendus les plus nombreux sont des poly-lexicaux de la forme *N_Adj* (735 occurrences de *N_Adj*, variantes en nombre et variations orthographiques comprises - soit 47% des termes-attendus). Nous avons choisi de nous intéresser aux 572 termes-attendus de la forme *N_Adj* (36%) qu'il reste après élimination des doublons (candidats au pluriel, variantes orthographiques telle

6 Nous remercions les ingénieurs documentalistes de l'Inist pour leur expertise : Nathalie Antonot (chimie), Frank Arnould (psychologie), Valérie Bonvallot (sciences de l'information et de la communication), Flora Badin et Sabine Barreaux (sciences du langage) et Bernard Sampité (archéologie).

7 Les candidats de ce groupe correspondent aux patrons : [N_prép_(dét)_N] – [N_prép_dét_(N_prép_N)] – [N_participe passé_dét_ N] – [N_Adj_Adj_Adj_coord_Adj].

la présence ou non d'un tiret entre le *N* et l'*Adj*, présence ou non de majuscule, etc.)

Notre objectif est d'examiner la forme des *N_Adj* attendus afin de déterminer si certaines propriétés du *N* et de l'*Adj* favorisent l'accès de l'ensemble *N_Adj* au statut de terme dans son domaine d'apparition. Nous posons l'hypothèse de la contamination adjectivale, c'est-à-dire de l'existence de relations particulières entre le *N* et de l'*Adj* qui entraînent ou favorisent la probabilité pour le poly-lexical d'accéder au statut de terme.

Différents travaux s'attachent au fonctionnement de l'adjectif de relation (*AdjRel*) dans la construction de termes poly-lexicaux (Daille 2001 et Maniez 2002, entre autres). Le rôle des adjectifs dans la construction des poly-lexicaux est une problématique courante en terminologie notamment lorsqu'on s'attache à la structuration de la terminologie (L'Homme 2004a, Mendes 2003). L'Homme (2004b) s'intéresse tout particulièrement aux déverbaux et dénominaux qu'elle regroupe dans la classe des adjectifs dérivés sémantiques (*ADS*) : la structuration terminologique est impactée par les liens que ces adjectifs entretiennent avec les bases nominales et verbales dont ils découlent. Pour ce qui est des relations sémantiques exprimées par les *Adj*, L'Homme (2004b) signale que les adjectifs sont généralement préfixés ou suffixés. Ce phénomène particulièrement marquant en sciences exactes se vérifie sur les candidats poly-lexicaux attendus en chimie (*poudre poly-cristalline, polycondensation interfaciale*).

Afin d'envisager une amélioration du repérage automatique, nous nous proposons d'observer la structure morpho-syntaxique des *N_Adj* ainsi que les propriétés des *N* et/ou des *Adj* qui les composent. Les différents points que nous retenons sont :

- *N* appartient au LST : *analyse formelle* - 214 occurrences de *N* sur 572 occurrences de *N_Adj* (37%)
- *Adj* dénominal : *abondance informationnelle* - 81 occurrences d'*Adj* sur 572 occurrences de *N_Adj* (14%)
- *Adj* entraîne la localisation dans le temps, dans l'espace ou les deux simultanément avec des *Adj* formés à partir de *NP : gisement mous-térien*, *abri magdalénien* - 77 occurrences d'*Adj* sur 572 occurrences de *N_Adj* (13%)
- *Adj* avec une forme identique au *N* correspondant *: analyse sémantique* - 33 occurrences d'*Adj* sur 572 occurrences de *N_Adj* (6%)

- *Adj* complexes - préfixation, suffixation, composition avec ou sans troncature : *test neuropsychologique*, *donnée thermodynamique* - 36 occurrences d'*Adj* sur 572 occurrences de *N_Adj* (6%)
- *Adj* déverbal : *donnée numérisée* - 25 occurrences d'*Adj* sur 572 occurrences de *N_Adj* (4%)

Lors de nos observations, nous avons remarqué 28 *N* qui entrent dans la composition de *N_Adj* d'au moins deux des cinq disciplines. Parmi ces 28 *N,* 14 appartiennent au *LST* (50%).

Le tableau 1 fait apparaître des *N_Adj* candidats-termes comportant un *N* relevant du LST complétés par un *Adj* mais aussi des *N* ne relevant pas du LST complété par un *Adj*. L'hypothèse de la contamination adjectivale se vérifie pour les *N* appartenant au LST mais doit aussi être envisagée pour des N sans lien avec le LST. L'hypothèse de la contamination demande à être vérifiée à plus grande échelle que le *N* appartienne ou non au LST.

N≠LST au moins deux disciplines	**N**=LST au moins deux disciplines
Comportement [linguistique, prosocial][8]	**Activité** [photocatalytique, picturale]
Eau [distillée, stagnante]	**Analyse** [plurifactorielle, physico-chimique]
Échelle [macro-régionale, micrométrique]	**Donnée** [dendrochronologique, numérisée]
Environnement [géologique, numérique]	**Espace** [méditerranéen, multimédias]
Gestion [différentielle, documentaire]	**Evolution** [morphologique, paléoenvironnementale]
Langage [corporel, graphique]	**Forme** [abstraite, graphique]
Langue [commune, étrangère]	**Identité** [sociale, socio-discursive]
Matière [première, osseuse]	**Milieu** [hétérogène, saharien]
Niveau [socio-culturel, aurignacien]	**Modèle** [atomistique, éditorial]
Outil [sémantique, archéologique]	**Organisation** [documentaire, interne]
Préparation [motrice, pharmaceutique]	**Période** [critique, protohistorique]
Représentation [mentale, figurative]	**Production** [langagière, lithique]
Ressource [cognitive, documentaire]	**Relation** [sociale, structurée]
Structure [membranaire, morphologique]	**Traitement** [numérique, phonologique]

Tab. 1 – N entrent dans la composition de N_Adj d'au moins deux disciplines

8 Pour chaque *N* nous proposons deux exemples de *N_Adj* issus de deux disciplines distinctes.

Nos observations et comptages montrent que 15 *Adj* apparaissent dans des *N_Adj* d'au moins deux disciplines. Un seul de ces adjectifs figure dans le LST : *statistique*. Par ailleurs, on peut remarquer que 7 de ces adjectifs ont une forme identique à celle du *N* correspondant.

Adj = forme identique au *N* correspondant	*Adj* ≠ forme différente du *N* correspondant
[archive, ressource] **électronique**[9]	[monde, compétence] **académique**
[comportement, développement] **linguistique**	[analyse, risque] **chimique**
[environnement, lecture] **numérique**	[forme, langage] **graphique**
[émotion, vide] **primaire**	[alcoxyde, pièce] **métallique**
[mémoire, ambiguïté] **sémantique**	[évolution, sensibilité] **morphologique**
[analyse][10] **statistique**	[bibliothèque, réussite] **scolaire**
[espace, monde] **virtuel**	[élite, relation] **sociale**
	[analyse, attention] **spatiale**

Tab. 2 – Adj entrent dans la composition de N_Adj d'au moins deux disciplines

3. En guise de conclusion

Les observations que nous avons effectuées sur les termes-attendus relevés par les annotateurs experts au cours de cette première expérience laissent entrevoir des patrons lexicaux syntaxiques permettant de prévoir l'accession de certains *N_Adj* au statut de termes poly-lexicaux. Les indices liés à l'*Adj* paraissent plus nombreux que ceux liés au *N*: l'appartenance du *N* au LST n'apparait pas comme un indice très performant. Les termes-attendus repérés par les annotateurs comportent un nombre de *N* appartenant au LST et un nombre de *N* n'appartenant pas au LST à peu près identiques. Pour les *Adj* les patrons suivants peuvent être proposés bien qu'ils demandent une vérification de leur validité sur un corpus étendu : les articles en texte intégral correspondant aux notices par exemple mais aussi d'autres documents des différentes disciplines.

***- N + Adj* dérivé d'un nom de lieu**
magdalénien – La Madeleine – *abri magdalénien*

9 Pour chaque *Adj* nous proposons deux exemples de *N* issus deux disciplines distinctes.

10 *Analyse* est le seul *N* utilisé avec *statistique* : on le retrouve dans plusieurs disciplines.

- ***N* + *Adj* dénominal**
  académie – *compétence académique*
- ***N* + *Adj* déverbal**
  numériser – *donnée numérisée*
- **N + Adj identique au N lui correspondant**
  sémantique – *ambiguïté sémantique*

D'autres adjectifs auxquels nous ne nous sommes pas encore intéressés entrent dans les compositions des *N_Adj* qui fonctionnent comme des termes. Par exemple nous avons remarqué *moyen,* un adjectif qualificatif, récurent dans les *N_Adj* attendus dans le domaine de l'archéologie : *paléolithique moyen*, *néolithique moyen*, *mésolithique moyen*, *pléistocène moyen*. Les particularités des *N_Adj* dans les différents domaines scientifiques demandent elles aussi à être prises en compte.

Référence

Aubin S. et Hamon T., (2006). Improving Term Extraction with Terminological Resources. *Advances in Natural Language Processing*, *(5th International Conference on NLP, FinTAL*.

Bourigault D., Aussenac-Gilles N. et Charlet J., (2004). Construction de ressources terminologiques ou ontologiques à partir de textes : un cadre unificateur pour trois études de cas. M. Slodzian (ed.)*, Revue d'Intelligence Artificielle, Numéro spécial sur les techniques informatiques de structuration de terminologies*, Hermès, 18(1), 87-110.

Bourigault D., Jacquemin C. et L'Homme M.C., (2001). *Recent Advances in Computational Terminology.* Amsterdam-Philadelphie : John Benjamins

Daille B., (2003). Conceptual structuring through term variations. In F. Bond, A. Korhnen, D. MacCarthy and A. Villacicencio (eds.), *Proceedings ACL 2003 Workshop on Multiword Expressions : Analysis, Acquisition and Treatment*, 9-16.

Daille B., (2001). *Qualitative terminology extraction : identifying relational adjectives*. In D. Bourigault, C. Jacquemin et M.C. L'Homme (eds), *Advances in computaional terminology,* John Banjamins, 149-166.

Daille B., (1996). ACABIT : une maquette d'aide à la construction automatique de banques terminologiques, in A. Clas, P. Thoiron et H. Béjoint, (eds.), *Lexicomatique et Dictionnairitique*, FMA, Beyrouth, 123-136.

Daille B., (1994). Approche mixte pour l'extraction de terminologie : statistique lexicale et filtres linguistiques, Thèse en informatique fondamentale, Université Paris 7.

Daille B., Jacquin Ch., Monceaux L., Morin E. et Rocheteau J., (2011). TTC-TermSuite : une chaîne de traitement pour la fouille terminologique multilingue, Démonstration, Taln'11, Montpelier, 27 juin-1er juillet.

Delavigue V., (2001). Repérage de termes dans un corpus de vulgarisation : aspects méthodologiques, *TIA*, Nancy, 33-43.

Drouin P., (2004). Acquisition des termes simples fondée sur les pivots lexicaux spécialisés. *TIA,* Strasbourg.

Drouin P., (2003). Term extraction using on-technical corpora as a point of leverage. *Terminology*, 9(1), 99-117

Jacquemin C., (1997). Variation terminologique : Reconnaissance et acquisition automatiques de termes et de leurs variantes en corpus. HDR en informatique fondamentale, Université de Nantes, Nantes.

Jacquemin C., (1996). What is the tree that we see through the window : A linguistic approach to windowing and term variation. *Information Processing & Management*, 32(4), 445-458.

Kister L. et Jacquey E., (2012). Relations syntaxiques entre lexiques terminologiques et lexiques transdisciplinaires : analyse en texte intégral, *CMLF 2012*, Lyon.

Kupsc A., (2007). Extraction automatique de termes à partir de textes polonais, *TALN*, Toulouse.

L'homme MC., (2004a). La terminologie : principes et techniques, Montréal : les presses de l'Université de Montréal.

L'homme MC., (2004b). Adjectifs dérivés sémantiques (ADS) dans la structuration des terminologies. Actes du colloque *Terminologie, ontologie et représentation des connaissances*, Université Jean Moulin Lyon 3, 22-24 juin.

Maniez F., (2002). Distinguer les termes des collocations : étude sur corpus du patron <Adjectif – Nom> en anglais médical, *TALN'02*, Nancy.

Mendes S., (2003) Sémantique de l'adjectif et ses restrictions de sélection. Actes *EDIT*, Université Paul Sabatier, Toulouse.

Toussaint Y., Namer F., Daille B., Jacquemin J. et Hathout N., (1998). Une approche linguistique et statistique pour l'analyse de l'information en corpus, *TALN'98*, Paris.

Mots-clés

poly-lexicaux – terminologie – annotation terminologique – annotation sémantique – extraction terminologique

Summary

The aim of this paper is to compare the results of an automatic terms extraction with the expectations of information managers. We propose to suggest clues to preselect French *N_Adj* which may be considered as terms.

False borrowings in fashion language : a corpus-driven analysis

Valeria Fiasco

Università degli studi Roma Tre
Dipartimento di Lingue, letterature e culture straniere - Roma
valeria.fiasco@gmail.com

Abstract. This study investigates the contemporary language of Italian fashion magazines and the most remarkable phenomena of linguistic interference which characterise its terminology. The analysis focuses on English and French false borrowings as used in one of the major fashion magazines and traces their creative integration into Italian. In particular, such terms are identified and discussed as regards their spelling, morphological features, frequency and etymology. The objective is to demonstrate the importance and the amount of False Anglicisms and False Gallicisms in the language of fashion as a language for specific purposes (LSP).

1. Introduction

Fashion is a complex phenomenon whose relationship with culture and society is one of reciprocal exchange (Barthes 1970, Catricalà 2004, Lopriore 2015). They influence each other in a continuous dialogue in which language plays a key role and also stands out for its characteristics. Therefore language and image highlight the multidimensional nature of this language for specific purposes (LSP).

Fashion magazines are specialised publications which promote fashion products, luxury goods and patterns of behaviour. In the last two decades the spreading of fashion magazines became massive with publications and editions from all over the world, especially in west countries.

Fashion and image are strictly interwoven since there is a never-ending connection to the iconic side. In fact fashion communicates through drawings, symbols, colours and words. Barthes (1970 :5) talks about *image-cloth-*

ing and *written-clothing*. The first one concerns a photograph or a drawn, while the second concerns the garment depicted in the photograph, but transformed into language. The *image-clothing* is made up of forms, lines and colours, while the *written-clothing* is made up of words and its nature is verbal. Obviously Barthes also talks about a third garment : the *real clothing* which is represented by the *image-clothing* and the *written-clothing*. As a consequence the description of fashion garments is a social fact, because it constitutes a significant element of the mass culture.

Moreover fashion and language are both characterised by change, since they are both linked to time. Fashion, like language, has two dimensions : a diachronic dimension because it is subject to progressive transformations through space and time on the basis of consumer tastes and fashion specialists' decisions and a synchronic dimension referring to a state, that is fashion which changes each year.

Fashion writing is also marked by creativity because it plays on words and uses a funny and fascinating language in order to catch reader's attention and arouse reader's curiosity. Therefore it is source of many neologisms formed by elements coming from different languages, in particular English, French, Italian.

2. The language of fashion as a language for specific purposes

The language of fashion constitutes an interesting research field as it represents a heterogeneous *ensemble* composed of technical, semi-technical and general language words (Bhatia 2005, Gotti 1991, 2003) which form different specialised languages and in turn correspond to different professional activities and profiles, i.e. tailor, designer, fashion stylist, fashion editor.

Therefore the language of fashion is composed of technical terms which refer to specific professions connected to the fashion system (embroiderer, tanner), terms borrowed from the industry of textiles and leathers, terms referring to chemical processes to realize clothes or leathers. A lot of them are not mentioned in general dictionaries, because of their specificity. These terms are characterised by morphological, syntactic, rhetorical and stylistic features which typify the specialised discourse like monoreferentiality, lack of emotion, precision, monosemy, transparency. In this case variation is marginally involved. If we refer to the three different situations in which a specialist uses the specialised language discussed in Gotti (2006 :17), in this situation the specialist addresses other specialists to talk about issues related

to their professional field, i.e. tailors and embroiderers who work in a *haute couture atelier*.

The language of fashion also includes semi-technical terms, as Cabré maintains these terms "can be attributed to a borderline area between general language and special language" i.e. open-toe, fashion show, outfit, trench. For this reason Gotti (2006) and other linguists state that the lexis of specialised languages is more productive than that of general language and one of the reasons is that specialised language has a higher lexical concentration, borrowings, prefixes and suffixes, even if they can also be found in general language.

On the opposite there is the language of fashion magazines, which is closer to the general language and it can be characterised as hybrid since it includes technical, semi-technical and general language words. It is connotative, emotional, polysemic, productive, i.e. jeans, pochette. It expresses the dynamism and it is the proof that the boundaries among these three groups of words are quite scarce (Bhatia 2004). The language used in fashion magazines is closer to the language of advertising. Its public is made up of specialists and laymen with the purpose of spreading trends and styles.

3. Linguistic interference and false borrowings

Linguistic interference is a topical issue at the present time as we live in a world where we can easily communicate, travel and relate to people coming from different cultures and languages. This situation of language contact give rise to diverse linguistic phenomena, like borrowings, calques and false borrowings.

The linguistic interference occurs provided that:

1. Two languages coexist in the same context
2. There is a bilingual speaker in that context

When the interference is just related to a single speaker, there are no relevant effects, but by the time the interference spreads among other speakers, the innovation becomes stable (Gusmani 1998).

When a language shapes some of its elements on another language and the relationship between the two languages involved can be proved, we can talk about loanwords or borrowings. If this relationship cannot be demonstrated, we cannot talk about borrowings on the basis of seeming similarities. Therefore where there is a clear divergence between the archetype and the supposed borrowing, the word can be defined pseudo-borrowing or false

borrowing. False borrowings are words whose form (pronunciation, spelling or morphology) reflects the donor language structures, even though they do not exist as such or are used with a totally different meaning. At the same time they are considered authentically foreign words by receptor language speakers and they are accepted in the vocabulary of the receptor language. Speakers use the linguistic material of the donor language and reinterpret it in order to create a new word, thus reshaping the vocabulary of the receptor language (Furiassi 2010).

The two donor languages being discussed here are English and French, in other words the two loan-giver languages mainly used in this kind of magazines. English has acquired the status of global language of international communication. French is the modern language that has provided, and still provides, the greatest number of borrowings in English. It has always been the language of fashion, representing the style, the elegance and the refinement of French *couturiers*.

4. Objective and research questions

The main objective of this research is to demonstrate the amount of false borrowings (false Anglicisms and false Gallicisms) as used in the language of fashion magazines, in particular this study refers to the Italian edition of the fashion magazine ELLE. The Italian language of fashion makes a considerable and extensive use of borrowings of English and French origin. More and more TV programs and magazines employ foreign words to refer to concepts connected to the fashion world and reach out to a wider audience, providing new words and expressions. For this reason it is interesting to examine how fashion language is shaped.

This investigation is also intended to investigate how the amount of false borrowings is composed, more false Anglicisms or false Gallicisms, thus examining their frequency, use and the semantic fields which they belong to.

Moreover it is important to consider the possible adaptations of false Anglicisms or false Gallicisms to the Italian language and investigate if the Italian corpus contains English and French words which correspond to false Anglicisms and false Gallicisms.

5. Research design and methodology

The corpus is drawn from the Italian edition of the fashion magazine ELLE as it is one of the most important and international fashion magazines and it is highly representative of fashion language. It plays on words, it displays creativity and productivity, thus representing a source of many neologisms.

The analysis focuses on a corpus of twelve issues covering a three year period (2007-2009, four issues per year) and scrutinises the texts' fashion-specific terms in three languages: Italian, English and French. In order to trace a complete and heterogeneous inventory of fashion language, the issues were selected on the basis of those containing more linguistic material, i.e. the issues which previewed and presented the new collections and fashion shows. Each issue has been wholly examined by extracting, dividing and classifying all Anglicisms and Gallicisms, genuine or false.

The methodology adopted to collect the corpus is the *corpus-driven approach*, according to which the corpus serves as an empirical basis from which lexicographers extract their data without former assumptions and expectations. Therefore genuine and false borrowings have been retrieved without making reference to borrowings already selected in dictionaries. Unlike the *corpus-based approach* where data is used to confirm, support and find proof for pre-existing linguistic theories or retrieve illustrative samples, the *corpus-driven approach* is an inductive approach where conclusions are drawn only on the basis of data investigation. After the collection of the corpus, the retrieved data have been compared with dictionaries and online corpora.

The corpus is synchronic since it focuses on a limited period of time, but it can also be integrated with new data from the following or previous years. It is also a specialized corpus as it belongs to the fashion domain. Finally it is realized by native speakers that use their mother tongue to express themselves.

The linguistic aspects investigated are:

- Spelling: spelling alteration analysis of false borrowings as regards the adaptation process to the receptor language, Italian.
- Morphology: the assimilation process of borrowings whereby the new word is a mix of the receptor language and donor language material.
- Frequency: quantitative analysis in order to measure the incidence of false Anglicisms and Gallicisms in the corpus.

6. False Anglicisms

The corpus includes an amount of thirty-two false Anglicisms. Henceforth the complete list of false Anglicisms grouped into four semantic fields:

Garments		
Baby doll	Cargo	Pull
Ball dress	Golf	Slip
Balloon	K-way	Smoking
Bermuda	Montgomery	Tight
Bomber	Plaid	Trench
Boxer	Polo	

Tab. 1 – *False Anglicisms pertaining the semantic field of Garments*

Accessories	**Materials and Fabrics**
Beauty case	Pile
Trolley	

Tab. 2 – *False Anglicisms pertaining the semantic fields of Accessories, Materials and Fabrics*

General language words	
Backstage	Miss
Dark	Nude Look
Dark lady	Restyling
Gadget	Revival
Jolly	Sportswear
Liberty	Total look

Tab. 3 – *False Anglicisms pertaining the semantic field of General language*

6.1. Spelling

Some false Anglicisms undergo a process of spelling adaptation into Italian which produces different orthographic variants. The corpus only contains two cases of this phenomenon which mainly concern compounds:

Beauty case - *Beautycase*

English translation: *cosmetic bag, cosmetic box, dressing case*

Baby doll - *Babydoll*

English translation: *babydoll nightdress, babydoll pajamas*

6.2. Morphology

Some other false Anglicisms are subjected to adaptation through derivational morphemes. The corpus includes just one false Anglicism which is adapted using a diminutive suffix. The false Anglicism *golf*, which refers to a pullover or a sweater, displays two diminutive suffixes for the singular form: *-ino, -etto*

Golf - *Golf**INO*** - *Golf**ETTO***

And a diminutive suffix for the plural form: *-ini*, which is the plural form of *golfino*:

Golf**ino** - Golf**ini**

6.3. Etymology

Since false borrowings don't exist in the hypothetical target-language, it is interesting to reconstruct their story and the events which caused these innovations.

The identification of false Anglicisms was preceded by a crosscheck on two important corpora: the British National Corpus (BNC)[1] and the Time Magazine Corpus of American English[2]. These corpora didn't contain the

1 http://corpus.byu.edu/bnc/

2 http://corpus.byu.edu/time/

false Anglicisms, but only the words which correspond to them in English. Henceforth some of the most interesting false Anglicisms will be discussed:

Baby doll

Italian meaning: a frilly short sleeveless nightdress with matching short pants

English translation: *baby doll nightdress*, *baby doll nightie*, *baby doll pyjamas*

Etymology: attested by 1963. This word derives from the name of the film "Baby doll" by Elian Kazan, released in 1956. The protagonist Carol Baker wore this garment in the film. It reminds a baby little dress.

Example: Una blusa che pare un **baby doll**, un paltò vestaglia […] (ELLE November 2007)

Translation: A blouse which looks like a baby doll, an overcoat which looks like a dressing gown […]

Ball dress

Italian meaning: a long dress usually worn to formal occasions or during a special event

English translation: *ball gown, ballgown, gown*

Etymology: this word is not mentioned in English dictionaries, but it appears seven times in the BNC, in a section concerning the 80s and the 90s, whereas it appears three times in the *Time Magazine Corpus of American English* in a section concerning the 60s. In English this word was probably used in the past and now has been replaced with *ball gown* (16 occurrences in the BNC and 19 occurrences in the *Time Magazine Corpus of American English*) and *gown* which is highly used in both corpora. In Italian this word is currently used in the language of fashion magazines.

Example: **Ball dress** in organza di seta con ruches (ELLE Aprile 2008)

Translation: Silk organza ball gown with ruches

Balloon

Italian meaning: a stitching which gives the fabric a rounded form like that of a balloon. It may refer to a sleeve, a skirt, a dress.

English translation: *ballooned, balloon slevees*

Etymology: In English the word *balloon* is not used in relation to the fashion world, since it relates to "an inflatable coloured rubber bag for party decorations, a circle containing the words of a character in comics, air bal-

loon, in surgery is an inflatable plastic tube used for dilating obstructed blood vessels".

This false Anglicism in Italian has two possible origins:

a) From the past participle of the verb "to balloon", "ballooned", that is puffed out like a balloon

b) From the meaning of "ball" or "hot-air balloon". Consequently in Italian it refers to a type of dress, skirt or sleeve through semantic shift.

Examples:

1. Giacchino di pelle con maniche **balloon** (ELLE Aprile 2007)

Translation: Balloon sleeves leather jacket

2. Gonna **balloon** in duchesse di seta (ELLE Novembre 2008)

Translation: Satin duchesse ballooned skirt

Bomber

Italian meaning: a short jacket with an elastic band on the waist and a zip on the front. It reminds the jackets worn by British Royal force pilots during the second world war. In Italian the false borrowing is the ellipsis of the English compound *bomber jacket*.

English translation: *bomber jacket*

Etymology: attested by 1983. According to the etymological Italian dictionary Zanichelli (1999), the word *bomber* comes from the French "bombé" in his turn deriving from the French noun "bombe" (bomb) referring to the round form.

Example: **Bomber** di cotone lamé (ELLE March 2007)

Translation: Lamé cotton bomber jacket

Cargo

Italian meaning: loose trousers with a large external pocket on each leg, for men and women. The model and colours remind army trousers. The Italian false Anglicism derives from the ellipsis of the English compound *cargo pants*.

English translation: *cargo pants, combat trousers, combats*

Example: Pantaloni **cargo** di cotone impermeabilizzato (ELLE November 2007)

Translation: Waterproof cotton cargo pants

Golf

Italian meaning : wool or cotton sweater. In Italian the false Anglicism *golf* has three diminutive forms[3]: *golfino* (sing.), *golfini* (plur.), *golfetto* (sing.).

English translation : *pullover, sweater, jersey, jumper* (BrE)

Etymology : attested by 1915. Its meaning is presumably explained by the fact that golf players formerly worn the pullover, thus becoming a popular garment.

Example : Per lei, il cappotto di cachemire e angora e i **golfini** girocollo (ELLE November 2007)

Translation : For her, cashmere and angora coat, crewneck pullovers

Montgomery

Italian meaning : a knee-length coat made of wool with a hood and fastened with toggles

English translation : *duffel coat, duffle coat*

Etymology : attested by 1953. This eponym derives from the name of the general Bernard Law Montgomery who habitually wore it. This garment spread during the II World War.

Example : Da festa il **montgomery** rosa (ELLE November 2009)

Translation : pink duffel coat for the party

Polo

Italian meaning : a short-sleeve cotton or wool t-shirt with a collar and two-three buttons at the neck

English translation : *polo shirt*

Etymology : attested by 1965. In India English officers wore it to play polo. Afterwards the French tennis champion René Lacoste put the polo shirt on the market with his name and the crocodile symbol and it became a success.

Example : […]**polo** di cotone piquet e accento wasp (ELLE September 2009)

Translation : piqué cotton polo shirt, wasp style

3 Ibid. 6.2

Smoking

Italian meaning: a man's evening jacket worn on formal occasions, usually black, but also blue or white. The collar, the lapels and sometimes the buttons are made of silk.

English translation: *dinner jacket (BrE), tuxedo (AmE)*

Etymology: attested by 1981. This false Anglicism is the ellipsis of the English compound *smoking jacket***:** so called because it was formerly worn by men at home for smoking. The jacket had silk facing over the collar and the lapels and it was fastened by a belt on the waist. Subsequently the *compound smoking* jacket was replaced by *dinner jacket.*

Example:

1. **Smoking** di raso con maxi maniche a sbuffo e pantaloni a sigaretta (ELLE September 2009)

Translation: Satin dinner jacket with maxi puff sleeves and drainpipe trousers

2. **P**igiama-**Tuxedo** di satin (ELLE April 2008)

Translation: Satin pjyama-tuxedo

Tight

Italian meaning: a man's morning suit worn on formal or special occasions and composed of a black or grey jacket with two tails, a grey or white waistcoat, pinstripe trousers and a grey top hat.

English translation: *morning suit, tailcoat*

Etymology: attested by 1970. The origin of the word is uncertain, but it probably derives from the English "tight" which means "close-fitting, snug"[4].

Example: Gilet tipo quello da **tight**, ma portato sullo chiffon (ELLE March 2007)

Translation: Morning suit waistcoat, worn on chiffon

Trench

Italian meaning: a gabardine belted double-breasted raincoat.

English translation: *trench coat, raincoat, mackintosh, macintosh*

Etymology: attested by 1933. It is the ellipsis of the English compound *trench coat. Trench* derives from the old French "trenche" which means

4 Battaglia S. (1961) *Grande Dizionario della Lingua Italiana*, Torino, UTET

"ditch, pit" and the word "coat", whose meaning is "overcoat, jacket", derives from the old French too. English officers worn it during the first world war. It became popular thanks to Greta Garbo who wore it in the film "Destino" in 1928.

Example : **Trench** di lana doppio petto a taglio militare (ELLE November 2007)

Translation : Wool double-breasted trench coat, military style

6.4. Frequency

To conclude this investigation on false Anglicisms, this paragraph examines their frequency. The three most frequent false Anglicisms are *pull* (88), *smoking* (50) and *trench* (67). The quantitative analysis focused also on English words corresponding to false Anglicisms in order to compare their recurrence in the corpus. On the whole the corpus comprises more false Anglicisms (417) than their English matching words (38) since Italian seems to prefer the false Anglicism. This is caused by the international status of English which has a great influence on the Italian language. As a consequence the majority of the words which represent the English translation of false borrowings is not used in Italian.

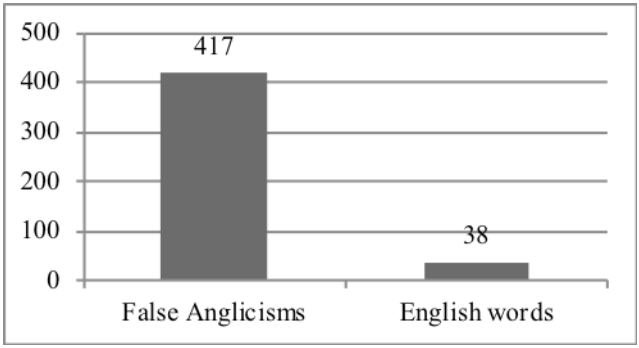


FIG. 1 – *Chart representing the quantitative difference between the number of False Anglicisms and the words which represent their English translation.*

In particular the frequency of some of the most relevant false Anglicisms will be treated in the following list:

False Anglicisms	Frequency	English words	Frequency
Pull	88	Pullover - Sweater - Jumper	9 - 2 - 0
Golf – Golfino	2 -3	Pullover - Sweater - Jumper	9 - 2 - 0
Smoking	50	Dinner jacket - Tuxedo	1 - 9
Trench	67	Trench coat - Mackintosh	0
K-way	1	Cagoule – Windbreaker	3 - 0

TAB. 4 – *Quantitative difference between the number of some of the most relevant False Anglicisms and the words which represent their English translation.*

The table shows a clear preference with regard to the false Anglicism, even if there is a word which goes against the prevailing trend: *k-way*, which is a registered trademark attested in Italian by 1981. As the table shows, it appears only one time in the corpus, whereas its English matching word, *cagoule*, appears 3 times.

7. False Gallicisms

The corpus here investigated includes only four false Gallicisms and none of them displays spelling or morphological adaptations, but they are all cases of semantic shift, or rather words that have undergone semantic changes. This phenomenon occurs because what is intended by speakers is not exactly the same each time. If a word is used with a different meaning and it is shared by the speech community, the new meaning becomes stable.

Materials & Fabrics	**Accessories**
Animalier	Sabot
Garments	**Specialized terms**
Duvet	Torchon

TAB. 5 – *False Gallicisms divided into semantic fields*

7.1. Etymology

In order to identify false Gallicisms, a check among dictionaries, a corpus of spoken French "Discours sur la ville, corpus de français parisien des années 2000" and some of the most important fashion brands websites has been undertaken. Henceforth false Gallicisms will be discussed :

Animalier

Italian meaning : textile, leather, fur made from, or as if from, the skin of an animal, for example leopard skin, zebra skin (with black and white stripes), panther skin.

French translation : *(tissu) imprimé panthère, léopard*

Etymology : from French *animalier* "a painter or sculptor of animal subjects, animal keeper (laboratory), zoo (parc animalier), referring to animals". In Italian this is the hyperonym of *leopardato (leopard skin), zebrato (zebra stripes), maculato (spotted).*

Example : Bomber di seta a stampe **animalier** (ELLE September 2009)

Translation : Leopard skin silk bomber

Duvet

Italian meaning : a down-filled jacket

French translation : *veste en duvet, doudoune*

Etymology : attested by 2003. Form French "duvet", "down".

Example : Via libera a **duvet**, cappotti e trench (ELLE November 2009)

Translation : Green light to duvet jackets, coats and trench coats

Sabot

Italian meaning : slip-on shoes, with or without heels

French translation : mule

Etymology : from French "sabot", from "savate" (slipper) and "bot" (foot) uncertain origin. In French the word *sabot* refers to wooden-soled shoes having no fastenings on the heel.

Example : **Sabot** di vinile e pelle (ELLE April 2007)

Translation : Vinyl and leather pump

Torchon

Italian meaning : Twisted necklace or bracelet. Particularly referring to the semantic field of jewellery. In general it may also refer to textiles or leathers unite or be united by twisting or twining together.

French translation : *Entrelacé*

Etymology : attested by 2003. From French "torchon" derived from "torche" intertwined straw rope.

Example : Collana a **torchon** di perle (ELLE November 2008)

Translation : Intertwined pearl necklace

7.2. Frequency

As the table shows, the corpus contains only false Gallicisms, whereas the words which represent their French translations are not included.

False Gallicisms	Frequency	French words	Frequency
Animalier	36	(Tissu) imprimé panthère	0
		(Tissu) imprimé léopard	0
Duvet	8	Doudoune	0
		Veste en duvet	0
Sabot	24	Mule	0
Torchon	4	Entrelacé	0

TAB. 6 – *Quantitative difference between the number of False Gallicisms and the words which represent their French translation.*

As far as the expression *(tissu) imprimé panthère*, *léopard* is concerned, the corpus only contains the word *imprimé* (six times) with different collocations : silk, leather, blouse, cotton, patent without any reference to the meaning of the word *animalier.*

8. Conclusions

This investigation showed the intense relationship among different languages and the complexity of the phenomenon of linguistic interference which may produce both borrowings and false borrowings. This research on false borrowings made use of different tools to confirm and support the hypothesis

about these words, i.e. dictionaries, corpora and specialized websites about English and French fashion most important brands.

On the whole the occurrence of false Anglicisms and false Gallicisms is quite different:

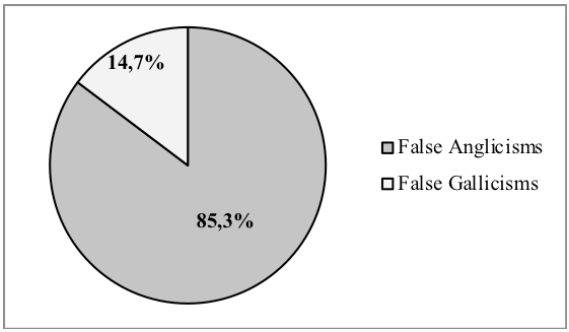


FIG. 2 - *The corpus includes 417 False Anglicisms and 72 False Gallicisms*

The quantitative difference between false Anglicisms and false Gallicisms is probably the reflection of the prestige attributed to the American culture which exercise a socio-historical and political hegemony over the contemporary world. The more a language is used and known by speakers coming from different cultures, the more the word flow in other languages increases so that bilingual speakers create neologisms composed of orthographic and morphological English material and Italian semantic material. Even though France is fashion world's leader, in this period of history, French does not have the same diffusion and the same influence as English.

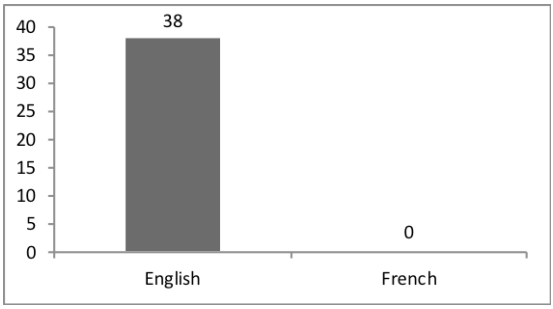


FIG. 3 – *Quantitative difference between the occurrence of English and French words corresponding to false borrowings.*

This table confirms what has been said before about the status of English. The influence of English on other languages implies the presence of both false borrowings and real donor language words. These words probably spread with contribution of bilingual speakers and advanced level speakers.

References

Beccaria G.L. (1983) I linguaggi settoriali in Italia, Milano, Bompiani

Beccaria G.L. (2006) Per difesa e per amore : la lingua italiana oggi, Milano, Garzanti

Bhatia V. K. (2005) Worlds of written discourse : a genre-based view, London, Continuum

Bowker L., Pearson J. (2002) Working with Specialized Language : a practical guide to using corpora, London and New York, Routledge

Cabré Castellví M.T. (1999) Terminology : Theory, methods and applications, Amsterdam/Philadelphia, John Benjamins Publishing Company

Catricalà M. (2004) Per filo e per segno : scritture della moda di ieri e di oggi : atti del Convegno, Roma, Museo Boncompagni-Ludovisi, 10 dicembre 2003 / a cura di Maria Catricalà, Soveria Mannelli, Rubbettino

Deroy L. (1967), L'emprunt linguistique, Paris, Les Belles Lettres

Furiassi C. (2010), False Anglicisms in Italian, Polimetrica International Scientific Publisher, Monza

Gotti M. (1991) I linguaggi specialistici, Scandicci, La nuova Italia

Gotti M. (2003) Specialized discourse : Linguistic Features and Changing Conventions, Bern, Peter Lang

Gusmani R. (1973) Aspetti del prestito linguistico, Napoli, Libreria Scientifica Editrice

Gusmani R., Interlinguistica, in Lazzeroni R. (1998) Linguistica Storica, Roma, Carocci

Hunston S. (2002) Corpora in Applied Linguistics, Cambridge, Cambridge University Press

Lopriore L (2007). Fashion in city magazines : the global and the local in Time Out (pp.375-390). In : Jottini L., Del Lungo G., Douthwaite J. eds.. Cityscapes : Islands of the Self. Language Studies. vol. 2, p. 1-587, Cagliari : Cooperativa Universitaria Editrice Cagliari

Lopriore L. Furiassi C. (2015). The influence of English and French on the Italian language of fashion : Focus on false Anglicisms and false Gallicisms. In Furiassi C. & Gottlieb, H. (eds), Pseudo-English : Studies

on False Anglicisms in Europe - Language Contact and Bilingualism, Berlin :De Gruyter Mouton
Migliorini B, (1957), Saggi Linguistici, Firenze, Le Monnier
Sager J.C. (1990) A Practical Course in Terminology Processing, Amsterdam/ Philadelphia, John Benjamins Publishing Company
Sinclair J. (1991) Corpus Concordance Collocation, Oxford, Oxford University Press
Tognini-Bonelli E. (2001) Corpus Linguistics at work, Amsterdam/ Philadelphia, John Benjamins Publishing Company

Résumé

Cet étude examine le langage des magazines italiens de mode et les phénomènes d' interférence linguistique les plus importants qui caractérisent sa terminologie. L'analyse focalise l'attention sur les faux emprunts d'origine anglais et français utilisés par l'un des principaux magazines de mode et décrit leur intégration créative en italien. En particulier, ces termes sont identifiés et traités en ce qui concerne leur orthographe, leur morphologie, leur étymologie et enfin leur fréquence. Le but de cette recherche est de démontrer l'importance et la quantité des faux anglicismes et gallicismes dans le langage de mode comme langue de spécialité.

Articles courts

Une démarche ontoterminologique pour la rétro-ingénierie de logiciels

Françoise Deloule*, Yves Dumond*, Christophe Roche*
Philippe Dugerdil**, Eric Harth**
Isabelle Chaussod***, Jean-François Falconnier***
David Corgier****, Florian Simon****

*Équipe Condillac, Laboratoire LISTIC, Université Savoie Mont-Blanc, France
francoise.deloule@univ-savoie.fr, yves.dumond@univ-savoie.fr, christophe.roche@univ-savoie.fr
**Haute École de Gestion de Genève, Suisse
philippe.dugerdil@hesge.ch, eric.harth@hesge.ch
***CIAT, Research, Development & Innovation, Culoz, France
i.chaussod@ciat.fr, jf.falconnier@ciat.fr
****CMDL, Le Bourget-du-Lac, France
david.corgier@cmdl.pro, florian.simon@cmdl.fr

Résumé. Le projet *Ontoreverse* est un projet européen Interreg IV France-Suisse portant sur la rétro-ingénierie de logiciels à base d'ontologies de domaine. Il repose sur l'idée que la compréhension de logiciels peut être vue comme un processus de mise en correspondance entre d'une part les concepts du domaine d'application concerné et d'autre part les structures logicielles qui s'y réfèrent. Les programmes sont ainsi considérés comme une forme d'expression des connaissances du domaine. Par exemple, un identifiant figurant dans un programme (nom de variable ou de classe) peut se référer à un ou plusieurs concepts métier. Une analyse des traces d'exécution permet de relier les parties du programme en rapport avec les connaissances du domaine d'application. La démarche est appliquée à la compréhension de codes logiciels relatifs à la conception d'échangeurs thermiques.

1. Introduction

Dans le cycle de vie des logiciels, la maintenance du code est de loin le poste le plus coûteux. Celle-ci peut représenter jusqu'à 90% du coût total (Erlikh 2000), dont près des deux tiers sont consacrés à la compréhension de logiciels (Müller *et al.* 1993). Par « compréhension de logiciels » on entend ici « mise en correspondance » entre les connaissances du domaine et les programmes, tant d'un point de vue descriptif que fonctionnel.

L'objectif de ce projet est de développer des outils informatiques destinés à améliorer la compréhension de logiciels à des fins de maintenance et d'évolution de logiciels métier, mais aussi de capitalisation de connaissances et de formation.

Cette étude est conduite dans le cadre du projet européen Interreg IV franco-suisse Ontoreverse associant deux partenaires universitaires : l'équipe Condillac du Laboratoire LISTIC de l'Université Savoie Mont Blanc et la Haute École de Gestion de l'Université de Genève, ainsi que deux industriels français : la société CIAT, spécialisée dans la production d'échangeurs thermiques et la société CMDL dont le domaine d'expertise est l'ingénierie et le conseil pour les performances énergétiques des bâtiments.

2. La démarche

Le travail présenté dans cet article repose sur deux idées principales :

- La première est que les ontologies constituent un formalisme idoine pour la représentation des connaissances propres aux domaines d'application traités et sur lesquelles les experts peuvent s'accorder dans la mesure où le consensus s'établit indépendamment de toute autre forme de représentation telle que des textes en langue naturelle ou des programmes.
- La seconde est que les logiciels sont une expression, dans un langage informatique, de tout ou partie des connaissances du domaine, comme peuvent l'être les documents scientifiques et techniques.

Nous sommes donc en présence de deux systèmes, l'un conceptuel correspondant à l'ontologie du domaine et l'autre « linguistique » correspondant aux programmes écrits dans un langage formel, en l'occurrence un langage de programmation. Ces deux systèmes ne sont pas isomorphes : on retrouve ici les résultats de l'extraction de connaissances à partir de corpus (Roche 2007). En effet, il existe des concepts nécessaires à la représentation ontologique des

connaissances qui sont sans contrepartie dans les programmes, de même qu'il existe des éléments de programmation, qui sont indispensables à la réalisation de certaines fonctions sans être liés à aucun concept de l'ontologie.

Partant des hypothèses que l'ontologie est au cœur de notre démarche de représentation, que les programmes sont une forme d'expression des connaissances représentées par l'ontologie et que ces deux systèmes ne se superposent pas, nous avons choisi de suivre une approche *ontoterminologique*. L'ontoterminologie est une terminologie dont le système notionnel est une ontologie formelle (Roche 2007, 2012). Ainsi, le nom d'un concept, en général explicite, ne se confond pas avec les « termes » qui s'y réfèrent (identifiants de variables, de classes…). La compréhension d'un logiciel est alors définie comme la mise en correspondance des éléments de programmes avec des structures de l'ontologie. Une analyse des traces d'exécution des programmes, sur la base de *scenarii* définis par les experts, permet de relier les parties du programme en rapport avec les connaissances du domaine.

3. La démarche ontoterminologique

3.1. Mise en œuvre

La spécification des ontologies est réalisée par la construction de graphes étiquetés. Les nœuds apparaissant dans ces graphes sont des concepts ou des instances de concepts relevant de l'expertise. Les concepts se réfèrent à des entités plurielles ayant une existence propre dans le domaine concerné. Ils peuvent être décrits par un ensemble d'attributs. Des termes peuvent être associés aux concepts et aux attributs au sein de l'ontologie. Les arcs des graphes constituant les ontologies sont étiquetés par des relations qui associent les concepts entre eux ou les concepts et leurs instances. Outre celles qui sont classiquement utilisées en représentation de connaissances, telles que *« is-a »*, *« part-of »* et *« instance-of »,* il est possible d'utiliser des relations propres à l'application.

Les ontologies étant spécifiées, on peut associer aux concepts, ainsi que le cas échéant à leurs attributs, des termes d'usage dans différentes langues (français, anglais, …) en incluant ceux associés aux langages de programmation et donc des noms de variables, de fonctions ou de méthodes, lorsque nous en disposons. L'analyse du code consistera ensuite à mettre en évidence les correspondances existant entre ces termes et des éléments de l'ontologie.

3.2. Exemple

Un extrait d'ontologie est représenté sur la partie gauche de la Figure 1. Trois concepts sont identifiés et les termes qui leur sont associés, ou qui le sont à leurs attributs, sont indiqués. En regard de cette ontoterminologie figure une portion de code écrite en langage C dans laquelle apparaissent certains des termes en question. Les correspondances existant entre les éléments de ces deux niveaux de description sont également représentées.

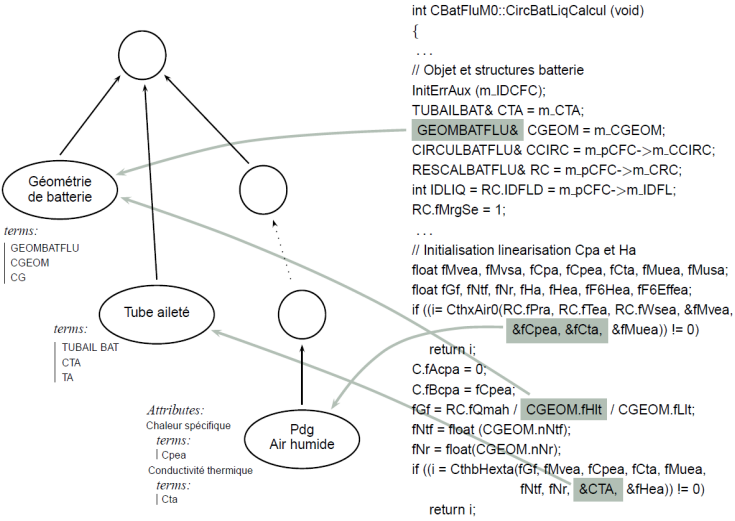


Figure 1 : exemple d'ontoterminologie

3.3. L'éditeur Ote 4 Soft

Afin de faciliter la saisie des ontoterminologies, un éditeur spécialisé appelé *« OTe 4 Soft »* a été réalisé. Celui-ci offre un haut niveau d'ergonomie et autorise en particulier des fonctionnalités de navigation dans les graphes ou d'affichage de ces derniers selon des vues spécifiques. Il est également possible d'assigner aux relations définies par les utilisateurs des propriétés spécifiques (transitivité, antisymétrie, etc…) et de procéder automatiquement à leur vérification. Les descriptions ontoterminologiques étant réalisées, il est possible de les exporter sous les formats d'échange JSON et RDFS/OWL pour permettre la phase d'analyse.

4. Analyse du code

Un outil d'analyse de code, appelé « *CSIM2* », a été implémenté afin de tester des stratégies de d'alignement terminologique (c'est-à-dire de mise en correspondance entre termes et concepts). Le processus de traitement et de représentation des informations associées aux codes sources et aux ontologies est le suivant (Figure 2) :

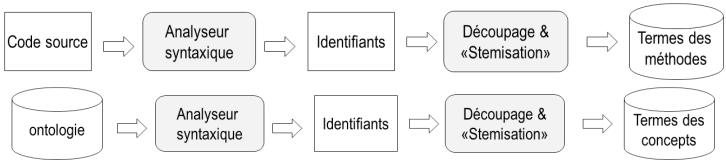


Figure 2 : processus de lemmatisation (stemisation)

Le découpage et la lemmatisation des identifiants respectent les étapes suivantes :

- élimination des standards de notations hongroises (Simonyi 1999) (cas du C/C++),
- découpage des termes composant les identifiants selon la notation *Camel Casing*,
- lemmatisation, c'est-à-dire recherche de la racine de chaque mot, à travers l'algorithme de Porter (Porter 1997).

Nous obtenons ainsi deux ensembles de termes comparables entre eux. La recherche des similitudes terminologiques consiste alors simplement à évaluer la distance entre chaque couple « terme-ontologique, terme-source-code », par exemple avec l'indice de Jaccard (Largeron *et al.* 2009) ou la distance de Levenshtein (Levenshtein 1966), et de pondérer celle-ci selon la fréquence des occurrences trouvées dans l'ensemble de tous les termes (Figure 3) :

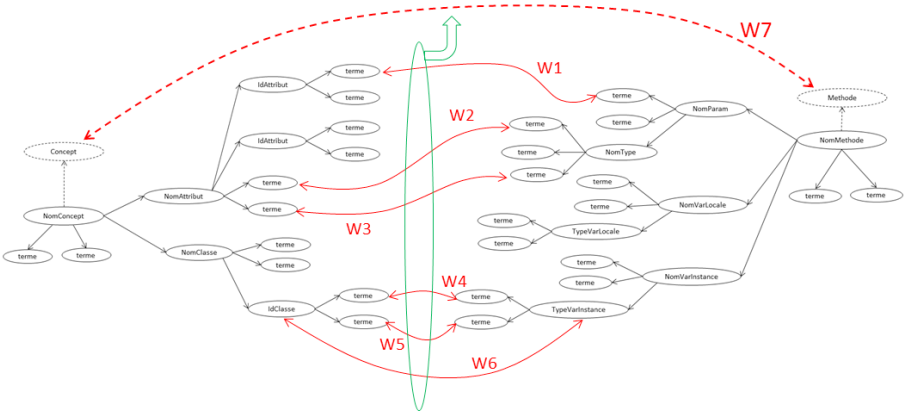


Figure 3 : graphe de correspondance pondérée

Nous avons ainsi établi un lien entre les connaissances du domaine et son implémentation concrète dans le code source du programme. Ceci apporte une information conceptuelle aidant à la compréhension du code à étudier.

5. Conclusion

Nous avons introduit dans cet article une démarche ontoterminologique pour l'aide à la compréhension de logiciels. Au-delà de son caractère prometteur, les premières expérimentations ont permis de mettre en évidence différents cas d'espèces. Si les développeurs ne sont pas impliqués dans le processus d'analyse, il est nécessaire d'avoir recours à des méthodes semi-automatiques de mise en correspondance des identifiants des programmes avec les dénominations apparaissant dans les ontologies. Cette mise en correspondance accroît la portée sémantique des termes apparaissant dans le code du programme.

Références

Belmonte J., Dugerdil Ph. & Agrawal A. (2014). A Three-Layer Model of Source Code Comprehension. In *proceedings of the 7th Indian Conference on Software Engineering (ISEC)*, Chennai, India, February 19-21.

Belmonte J. & Dugerdil Ph. (2010). Using Domain Ontologies in a Dynamic Analysis for Program Comprehension. In *proceedings of the 2nd International Workshop on Ontology-Driven Software Engineering. ACM Conference on Systems, Programming, Languages, and Applications (SPLASH)*, Reno, Nevada. October 17-21.

Biggersta T.J. & B. Mitbander B. (1994). Program understanding and the concept assignment problem. *Communications of the ACM*, 37(5):72-82.

Erlikh L. (2000). Leveraging legacy system dollars for e-business. *IT Professional*, 2(3):17-23.

Largeron C., Kaddour B. & Fernandez M. (2009). SoftJaccard : une mesure de similarité entre ensembles de chaînes de caractères pour l'unification d'entités nommées. *Actes de Extraction et Gestion des Connaissances.* Strasbourg, France, 27-30 Janvier.

Levenshtein V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, 10(8):707-710.

Müller H. A., Tilley S. R. & Wong K. (1993). Understanding software systems using reverse engineering technology perspectives from the Rigi project. In *proceedings of the 1993 conference of the Centre for Advanced Studies on Collaborative research : software engineering. (CASCON).* Toronto, Ontario, Canada, October 24-28. IBM Press.

Porter M.F. (1997). An algorithm for suffix stripping. In *Readings in Information Retrieval*, Morgan Kaufmann Publishers, p. 313-316.

Roche C. (2007). Saying is not modelling. In *Proceedings of the 4th International Workshop on Natural Language Processing and Cognitive Science (NLPCS) ; International Conference on Enterprise Information Systems (ICEIS).* Funchal, Portugal, June 12-16.

Roche C. (2012). Ontoterminology : How to unify terminology and ontology into a single paradigm. In *proceedings of the 8th International Conference on Language Resources and Evaluation (LREC).* Istanbul, Turkey, May 21-27.

Simonyi C. (1999). Hungarian Notation. *MSDN Library.* Microsoft Corporation.

Achevé d'imprimer en novembre 2020
Imprimerie Présence Graphique
2 rue de la Pinsonnière
F - 37260 MONTS