

Analyse et Représentation de la Conceptualisation Métaphorique en Langue de Spécialité à l'aide des Bases de Données Relationnelles

Sylvie Vandaele

Département de linguistique et de traduction

Université de Montréal

C.P. 6128, succ. Centre-ville

Montréal (Québec) H3C 3J7

Canada

sylvie.vandaele@umontreal.ca

Résumé

La mise en relation des données terminologiques, lexicographiques et conceptuelles constituent un des défis actuels de la représentation des connaissances dans les langues de spécialité. Nous présentons une base de connaissance permettant de modéliser la conceptualisation métaphorique en biologie cellulaire, plus spécifiquement la communication cellulaire, en relation avec les termes du domaine et la phraséologie décrite à l'aide des fonctions lexicales. La base est en conçue dans une perspective d'analyse des réalisations linguistiques des métaphores conceptuelles en français et en anglais, mais est également utilisable en mode de consultation, peut intégrer d'autres langues et est exportable pour la consultation en réseau ou sur Internet.

1. Introduction

Un des problèmes majeurs rencontrés dans l'analyse des langues de spécialité a trait aux rapports entretenus par le terme avec son environnement lexical en corpus, c'est-à-dire la phraséologie (par ex., Béjoint, 1992; L'Homme, 1992; Pavel, 1993; Meyer, 1994). Différents chercheurs en terminologie et en lexicographie ont ainsi examiné la possibilité de rendre compte des contraintes lexicales (collocation) en langue de spécialité, notamment au moyen des fonctions lexicales (L'Homme, 2002). Par ailleurs, les travaux issus de la sémantique cognitive ont permis de mettre en évidence l'importance fondamentale de la conceptualisation métaphorique en langue générale (Lakoff, 1980/2003), mais aussi dans les langues de spécialité (LSP) (Meyer, 1997), notamment la biomédecine (van Rijn-van Tongeren, 1997; Temmerman, 2001).

La *conceptualisation métaphorique* correspond, selon Lakoff (1993; 1980/2000) à la projection d'un domaine de conceptualisation sur un autre. On entend par *métaphore*, ou *métaphore conceptuelle*, la projection ainsi établie (par exemple, LA RECHERCHE BIOMEDICALE EST UNE ENQUETE.), qui est décrite par un certain nombre de correspondances (Le médecin/le chercheur est le détective. L'agent pathogène est le criminel. La maladie est le crime.) et se traduit par le recours à certains unités lexicales spécifiques du domaine

projeté. L'expression *métaphorique* correspond à la réalisation de surface de la conceptualisation métaphorique (« *Les chercheurs traquent le virus du SRAS, qui a fait des dizaines de victimes de par le monde.* »).

Selon notre hypothèse de travail, la conceptualisation métaphorique est le facteur essentiel sous-jacent à la phraséologie et à certaines dénominations terminologiques en LSP. Les résultats que nous avons déjà obtenus dans le cadre d'un projet de recherche portant sur la communication cellulaire (biologie cellulaire) appuient cette hypothèse (Vandaele, 2000; Vandaele, 2002a; Vandaele, 2002b). Par ailleurs, se pose la question de la représentation des données recueillies au sein des corpus, tant à des fins d'analyse fine que pour la constitution de bases de connaissance utilisables, *in fine*, par des usagers, traducteurs ou rédacteurs. La complexité des relations à établir entre termes, phraséologie et métaphore conceptuelle impose d'avoir recours à un outil autorisant l'établissement de multiples liens entre les données. La flexibilité est une condition impérative, étant donné la nécessité de faire évoluer le modèle en fonction des données observées. Dans ce contexte, le recours aux bases de données relationnelles s'impose, solution puissante déjà retenue par d'autres chercheurs (par ex., L'Homme, 1997).

2. Modélisation de la base de connaissance

Nous entendons ici par *base de connaissance* (BC) une base de données relationnelle réalisée au moyen de FileMakerPro, v.5.0/6.0. La BC comporte différents *fichiers* regroupés en *modules*. Les fichiers sont mis en relation les uns avec les autres au sein de la base au moyen de *liens*. Chaque fichier est constitué de *fiches*, comportant des *rubriques* (ou champs). Les rubriques peuvent s'afficher dans différents *modèles*, comme si les fiches pouvaient avoir plusieurs pages. Elles autorisent la saisie des données ou leur affichage par le biais de *tables externes*. Celles-ci affichent les données provenant soit d'un autre fichier (*lien externe*), soit du même fichier (*lien interne*).

L'objectif de la modélisation est double, car il vise à la fois l'analyse et la représentation des données, dans au moins deux langues (anglais/français). Le principe général gouvernant la structure de la BC repose sur l'atomisation maximale de l'information : une fiche ne peut mettre en rapport qu'une seule variable (ou qu'un seul bloc de variables uniques) pour un paramètre donné. La BC intègre trois types de données (terminologiques, lexicales, conceptuelles), correspondant à trois modules.

La conception du **module terminologique** (Figure 1) est fondée sur le rapport qu'entretient le terme avec la notion. Par convention, nous adoptons *notion* pour dénoter le signifié du terme, afin de réserver le vocabulaire dérivé de *concept* à ce qui relève de l'approche cognitive, c'est-à-dire du mode de conceptualisation du domaine. Le terme constitue le paramètre central des fichiers permettant de colliger indépendamment différentes variables : définitions, contextes, renseignements linguistiques (catégorie lexicale, genre, niveau de langue).

La notion, quant à elle, est le paramètre central des fichiers indiquant les domaines d'appartenance, les relations hyperonymiques, les relations méronymiques, ainsi que la terminologie latine (nomenclatures internationales d'anatomie, par exemple). Présente dans tous les fichiers du module, la rubrique 'Notion' a pour fonction d'établir les liens : son contenu est en fait purement conventionnel, mais on utilise généralement le terme recommandé, auquel on adjoint un numéro dans les cas de polysémie. Un module

terminologique de structure identique est réalisé pour l'anglais. Enfin, un fichier synthétique (*FICHES TERMINO*, Figure 2) regroupe, grâce à des tables externes, l'ensemble des des données colligées dans chacun des fichiers du module.

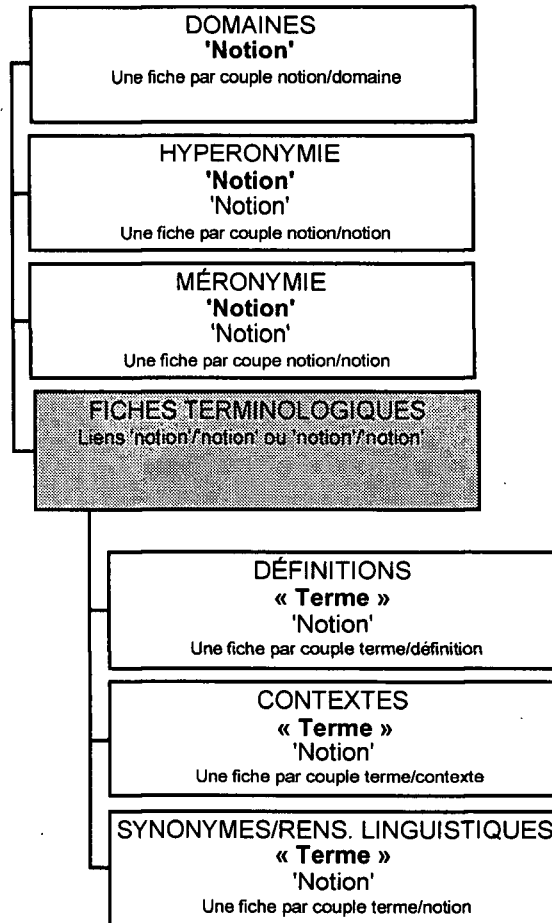


Figure 1 – Structure du module terminologique

Notion : protéine		Domaines : Biologie moléculaire Biologie cellulaire Biochimie	
Catégorie(s) : Nom		Hyperonymie : macromolécule Structure	
Termes F	Genre	Sigle	Référence
protéine	Féminin		tous niveaux
peptide (moins de 20 AA)	Masculin		spécialisé
polypeptide (plus de 100 AA)	Masculin		spécialisé
protéide	Masculin		désuet
Termes A	Genre	Référence	Usage
protein			1. Terme recommandé
protein molecule		(3.280)	2. Synonyme
peptide (< 20 AA)		(3.280)	2. Synonyme
polypeptide (> 100 AA)		(3.280)	2. Synonyme
protide (désuet)			4. Synonyme
Latin	Genre	Nomenclature	Référence
Définition française			
Macromolécule polypeptidique, c'est-à-dire formée d'une succession d'acides aminés (AA) qui commence par un groupement amine (-NH ₂) et se termine par un groupement carboxyle (-COOH). Elle remplit diverses fonctions biologiques selon sa séquence, sa localisation (intracellulaire, membranaire ou extracellulaire) et sa composition, simple (albumine) ou complexe (hémoglobine).			
Définition anglaise			
Macromolecule made up of one or more chains of amino acids joined covalently through peptide bonds. Their functional architecture, conferred by disulfide bridges, hydrogen bonds and hydrophobic interactions. Proteins vary greatly in molecular size from a few thousand to several million daltons.			
Hyperonymie		Dimension utilisée dans la relation	
récepteur (1)		Structure	
protéine membranaire		Localisation	
protéine transmembranaire		Localisation	
protéine intrinsèque		Localisation	
protéine fibreuse		Structure	
protéine globulaire		Structure	
protéine de transport		Fonction	
hydrolytique		Structure	
hémaprotéine		Structure	
Méronymie (esp)			
acide aminé			
Méronymie (inf)			

Figure 2 – Modèle principal du fichier regroupant les informations du module terminologique.

Le **module lexical** (Figure 3) vise à colliger les collocations relevées dans les corpus, à l'aide des fonctions lexicales décrites dans la Théorie Sens-Texte (Mel'cuk, 1996). Le point de départ de la description des collocations est le repérage d'unités lexicales prédictives ayant un statut de terme clair (par ex., *récepteur*, *domaine*) ou ambigü (*message*), ou encore faisant partie de la phraséologie repérable en corpus (par ex., *résider [dans la membrane]*). La structure actantielle des prédicats et les collocations significatives repérées en corpus (Tableau 1 : exemple de *message*) sont colligées dans des fichiers indépendants et encodées à l'aide des fonctions lexicales. À cette fin, deux corpus ont été constitués à partir de documents scientifiques sur la communication cellulaire en anglais et en français, respectivement de plus de 300 000 et 500 000 mots. Les données sont acquises après examen de contextes extraits des corpus et recueillis dans le fichier contexte déjà utilisé dans le module terminologique.

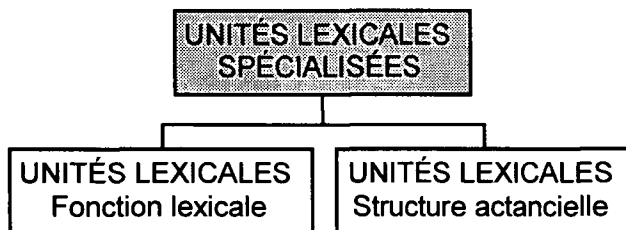


Figure 3 – Structure du module lexical

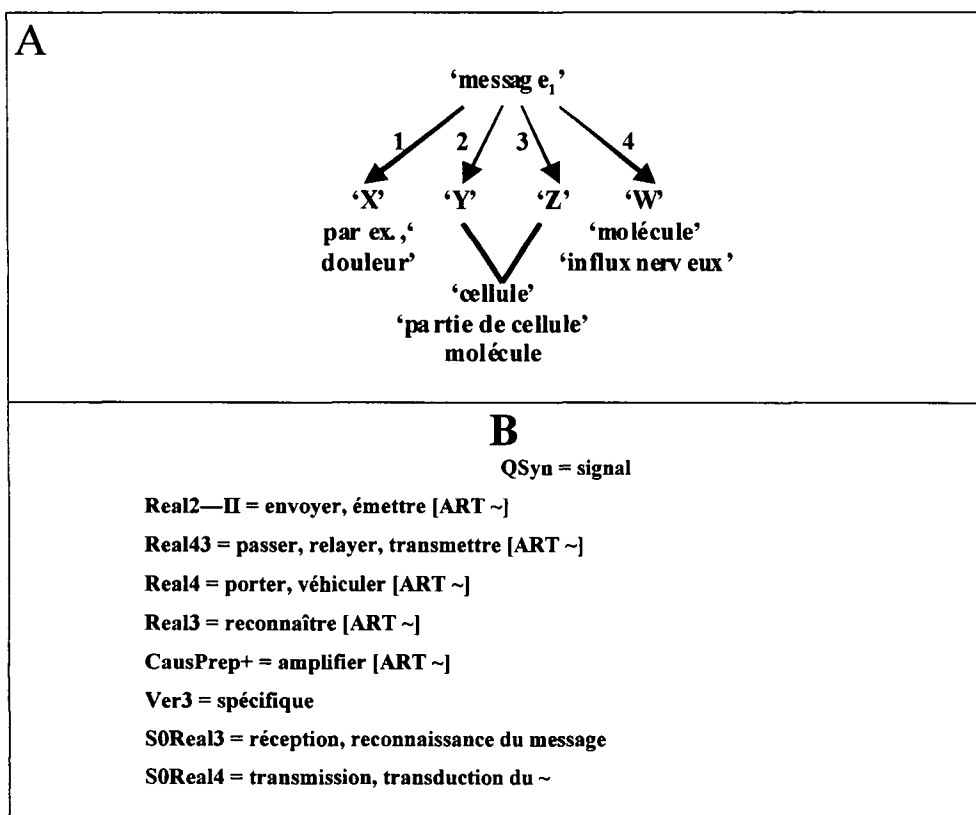


Tableau 1 – Structure actancielle (A) et fonctions lexicales (B) de l'unité lexicale *message* colligées dans le module de phraséologie.

Le module conceptuel a pour objet l'analyse des métaphores conceptuelles parcourant le domaine (Figure 4). Le paramètre principal est le prédicat indice de métaphore (indice métaphorique) : le modèle principal du fichier (Figure 5) affiche dans une table externe les contextes contenant l'unité lexicale en question, les actants ainsi que les fonctions lexicales (provenant du module lexical) dans la zone CIBLE, tandis que s'affichent en parallèle (zone SOURCE) des données lexicographiques provenant soit de la base DICO (Polguère, 2000), soit d'une analyse menée selon les règles de la TST en langue générale et compatible avec un format DICO (Polguère, 2003). Cette mise en parallèle permet d'évaluer la convergence d'actants et de fonctions lexicales entre la lexie cible et la lexie source. Par exemple, les contextes contenant des expressions du type « [...]cette protéine résidant dans les microdomaines occupe une position centrale dans l'activation... » suggèrent la présence d'une conceptualisation métaphorique sous-jacente du type UNE MOLÉCULE EST UN INDIVIDU HABITANT UN LIEU, parce que la structure actancielle de l'unité lexicale *résider* est telle que X réside dans Y, avec X = individu, Y = lieu d'habitation (maison, quartier, région, pays...). Le nom de la métaphore conceptuelle est saisi dans la rubrique *ad hoc*. Celle-ci établit un lien avec un fichier indépendant permettant de hiérarchiser les métaphores conceptuelles, qui est mis à jour au fur et à mesure de la progression de l'analyse. La hiérarchisation permet d'afficher les métaphores conceptuelles plus générales dans le fichier synthèse du module conceptuel, d'en vérifier la cohérence et d'observer le jeu des projections multiples de plusieurs domaines sources sur le domaine cible considéré. Ainsi, dans l'exemple cité, UNE MOLÉCULE EST UN INDIVIDU HABITANT UN LIEU est un cas particulier de UNE MOLÉCULE EST UN INDIVIDU (ou un personnage, [Vandaele, 2000]), intégrée avec LES PARTIES DE LA CELLULE SONT DES ESPACES GÉOGRAPHIQUES, ce qui rappelle l'idée de dualité (Lakoff, 1993) ou d'intégration conceptuelle (Fauconnier, 1998).

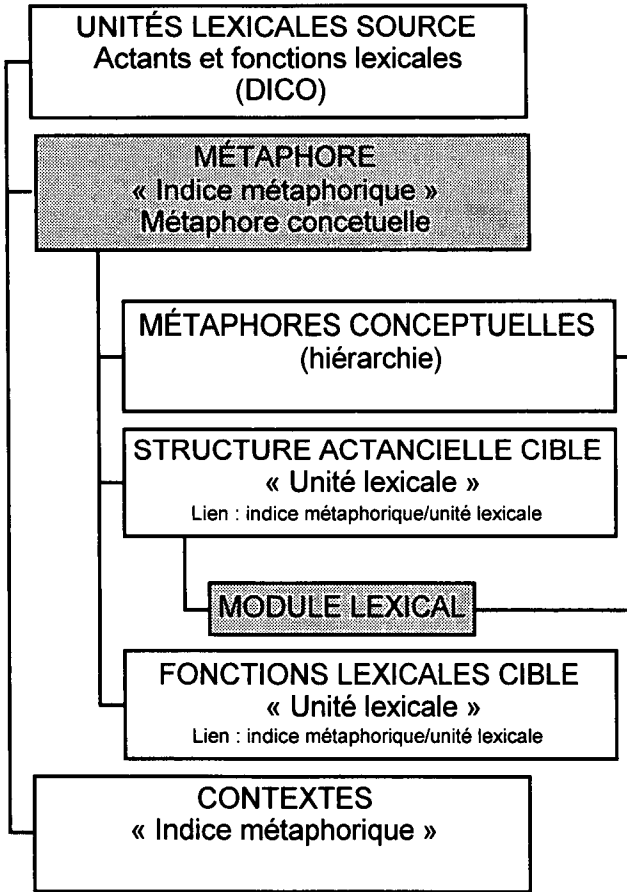


Figure 4 – Structure du module conceptuel

Métaphore conceptuelle
L'ORGANISME EST UN SYSTÈME DE COMMUNICATION.

Une entité est une autre entité.

Prédicat message (V)

Étiquette haut niveau	Étiquette bas niveau	Actant Cible
entité		cellule, partie de cellule
entité		molécule, partie de
entité		cellule, partie de cellule
entité		molécule, partie de
entité		molécule, infir, parveus

(S5yn) signal
(S1) information?
(S3) cellule cible
(S3) molécule cible
(S4) message
(S4) molécule signal
(S4) molécule message
(Fact1-1) envoyer [ART ?]
(Fact2-1) émettre [ART ?]
(Fact42) passer [ART ? à N-?]

Bien entendu, chaque cellule possède, à un moment donné, un assortiment particulier de différents récepteurs, de différents protéines G et de différents types d'adénylate-cyclase et présentera un contexte quasi-unique d'intégration des messages. Apparemment pour la production d'AMPc.

Toute cellule est en effet dotée de la capacité de détecter les signaux de son environnement. Ces signaux, essentiellement transmis sous la forme de molécules chimiques que la cellule a appris à reconnaître, sont ensuite décodés, c'est-à-dire organisés en messages auxquels la cellule peut donner une réponse tirée de son répertoire génétique, par exemple une contraction pour une cellule musculaire ou la sécrétion d'une hormone pour une glande. C'est en ce sens que l'on peut parler d'un « langage des cellules ».

Dans les êtres multicellulaires, la croissance, la division et l'activité des cellules sont régulées par des molécules de communication (hormones, neurotransmetteurs, facteurs de croissance, cytokines...). Si on excepte les hormones thyroïdiennes et stéroïdes, les récepteurs de ces molécules de communication et de ceux des messages venant du milieu environnant (odeur, lumière...) sont exposés à la surface membranaire. La famille des récepteurs couplés aux protéines G est la plus probable, permettant de reconnaître des messages de nature

Cible EST **Source**

Actant source	Étiquette bas niveau	Étiquette haut niveau
ventouse		fait
émetteur	Personne ou Artéfact	Entité
récepteur	Personne ou Artéfact	Entité
moyen	Personne ou Artéfact	Entité

Unité d'information : ~ CONTENANT X ET ENVOYÉ PAR Y À Z PAR W

Fonction lexicale source

Fact2	envoyer [ART ? à N-?]
Fact1	parvenir [à N-?]
Fact1	arriver [à N-?]

Figure 5 – Modèle principal du fichier d'analyse conceptuelle. Il autorise l'étude des métaphores conceptuelles et affiche des données d'autres fichiers.

3. Conclusion et perspectives

À notre connaissance, il s'agit de la première tentative visant à intégrer des données terminologiques, conceptuelles et lexicales. Au moment de la rédaction du présent article (novembre 2003), 1506 contextes et 938 définitions en français, 1419 contextes et 766 définitions servent de base à l'analyse. Le module terminologique comporte 276 notions, 485 termes français et 407 termes anglais. Les modules phraséologiques et conceptuels sont en cours de développement, une trentaine de métaphores conceptuelles ayant été identifiées à travers l'analyse d'une cinquantaine de lexies prédicatives. L'atomisation des données et la flexibilité de la BC permet de faire évoluer la modélisation au fur et à mesure des difficultés rencontrées, sans pour autant perdre les données déjà colligées : il suffit d'exporter ou d'importer des données dans de nouveaux fichiers, de créer des rubriques supplémentaires et de remanier les liens. Ce travail de modélisation permet d'analyser finement les relations entretenues par les termes et les autres unités lexicales du discours de spécialité, d'une part, et de mettre à l'épreuve différentes hypothèses quant à la réalisation linguistique des métaphores conceptuelles parcourant le domaine. La structure des modules phraséologiques et conceptuels, actuellement développés pour le français, est aisément transposable pour l'anglais et d'autres langues. La BC représente donc un outil unique permettant la mise en correspondance non seulement des termes mais aussi de la phraséologie, par le truchement des fonctions lexicales, d'une part, et des métaphores conceptuelles, d'autre part. Enfin, le passage d'outil d'analyse à outil de consultation est relativement aisé à réaliser, FileMakerPro autorisant notamment la publication des données sur Internet.

Remerciements

628

La recherche a été menée dans le cadre d'un projet de recherche appuyé par le Conseil de recherches en sciences humaines du Canada. Nous remercions nos collègues Marie-Claude L'Homme pour les discussions fructueuses, Alain Polguère pour sa disponibilité et l'accès à la base de données DICO, Igor Mel'Čuk pour sa générosité à transmettre ses connaissances, ainsi que Sylvie Boudreau, Magali Cape, Elizabeth Marshman et Mélanie Proulx pour leur contribution technique au projet.

Références

- Béjoint, H. et Thoiron P.** 1992. 'Macrostructure et microstructure dans un dictionnaire de collocations en langue de spécialité.' *Terminologie et traduction*, n° 2/3-1992: 513-522.
- Fauconnier, G. et Turner M.** 1998. 'Conceptual Integration Networks.' *Cognitive Science* 22(2): 133-187.
- L'Homme, M.-C.** 1992. 'Les unités phraséologiques verbales et leur représentation en terminographie.' *Terminologie et traduction*, n° 2/3-1992: 493-503.
- L'Homme, M.-C. et Bertrand C.** 2000. 'Specialized Lexical Combinations: Should they be described as Collocations or in Terms of Selectional Restrictions?' *Proceedings Euralex'00*, Ninth Euralex International Congress, Stuttgart:497-506.
- L'Homme, M.-C. et Gemme R.** 1997. 'Modèle d'accès informatisé aux combinaisons lexicales spécialisées verbe + nom (terme) et extension au nom (déverbal) + préposition + nom (terme).' *Mélanges de linguistique offerts à Rostislav Kocourek*, Université Dalhousie (Halifax), Les Presses Alfa.
- Lakoff, G.** 1993. 'The contemporary theory of metaphor.' In A. Ortony (ed.), *Metaphor and Thought*. Cambridge, Cambridge University Press: 202-251.
- Lakoff, G. et Johnson M.** 1980/2003. *Metaphors we live by - With a new afterwords*. Chicago, The University of Chicago Press.
- Mel'Čuk, I. A.** 1996. 'Lexical Functions: A tool for the description for lexical relations in the lexicon.' In L. Wanner. (ed) *Lexical Functions in the Lexicography and Natural Language Processing*. Amsterdam/Philadelphia, Benjamins: 37-102.
- Meyer, I. et Mackintosh K.** 1994. 'Phraseme Analysis and Concept Analysis: Exploring a Symbiotic Relationship in the Specialized Lexicon.' *Proceedings of Euralex'94*, Sixth Euralex International Congress, Amsterdam:339-348.
- Meyer, I., Zaluski, V. et al.** 1997. 'Metaphorical Internet Terms: A Conceptual and Structural Analysis.' *Terminology* 4(1): 1-33.
- Pavel, S.** 1993. 'La phraséologie en langue de spécialité. Méthodologie de consignation dans les vocabulaires terminologiques.' *Terminologies Nouvelle*, n° 10: 67-82.
- Polguère, A.** 2000. 'Towards a Theoretically-Motivated General Public Dictionary of Semantic Derivations and Collocations in French.' *Proceedings of EURALEX'00*, Ninth Euralex International Congress, Stuttgart:527-527.
- Polguère, A.** 2003. *Lexicologie et sémantique lexicale*. Montréal, Les Presses de l'Université de Montréal.
- Temmerman, R.** 2001. 'Sociocognitive Terminology Theory.' *II simposio Internacional de Verano de Terminologia*, Verano, Institut Universitari de Lingüística Aplicada, Universitat Peompeu Fabra.

- van Rijn-van Tongeren, G. W.** 1997. 'Metaphors in medical texts.' Amsterdam / Atlanta, Editions Rodopi B.V.
- Vandaele, S.** 2000. 'Métaphores conceptuelles et traduction biomédicale' In S. Méjri, T. Baccouche, A. Clas, G. Gross (éd.), *La traduction : théorie et pratiques*, actes du colloque *Traduction humaine, traduction automatique, interprétation*, Tunis, 28-29 septembre 2000, Publications de l'ENS, pp. 393-404.
- Vandaele, S.** 2002a. Métaphores conceptuelles en traduction médicale. TTR, volume XV, n °1:223-239
- Vandaele, S.** 2002b. 'Metaphorical Conceptualization in Cell Biology' *Proceedings of EURALEX 2004*, The Tenth Euralex International Congress, Copenhagen, Denmark, 13-17 août 2002. Proceedings, vol. II:649-655.