

Annotation XML et interrogation de corpus pour l'étude de la conceptualisation métaphorique

Sylvie Vandaele, Sylvie Boudreau

Université de Montréal

Abstract

In metaphorical conceptualization, the structure of one conceptual system is projected onto another. Our previous work suggest that idiomaticity in specialized languages is based in part on this phenomenon. However, since metaphorical conceptualizations are cognitive in nature, they are not easy to identify and characterize. Hypothesizing that it is possible to describe these conceptualizations based on their expression in discourse, we propose a semantic corpus annotation method using *XML*. The system we have developed allows for the systematic study of metaphorical conceptualizations and constitutes an elegant and efficient means of extracting qualitative and quantitative data that could not otherwise be obtained.

Résumé

La conceptualisation métaphorique consiste à projeter la structure d'un système conceptuel sur une autre. Nos travaux antérieurs suggèrent que l'idiomaticité repose en grande partie sur ce phénomène, qui se manifeste, entre autres, dans les langues de spécialité. Toutefois, comme elle se situe au plan de la pensée, la conceptualisation métaphorique ne se laisse pas appréhender facilement. Faisant l'hypothèse qu'il est possible de la caractériser par l'intermédiaire de ses réalisations en discours, nous proposons une méthode d'annotation sémantique de corpus faisant appel au langage *XML*. Le système mis en oeuvre permet une systématisation de l'étude de la conceptualisation métaphorique et autorise, de façon élégante, l'obtention de données qualitatives et quantitatives autrement impossibles à obtenir.

Mots-clés : conceptualisation métaphorique, linguistique de corpus, annotation sémantique, langage de balisage *XML*

1. Introduction

À la suite de Lakoff (1980/2003), nous situons la **conceptualisation métaphorique** (*CM*) au plan de la pensée, l'expression métaphorique étant sa réalisation linguistique. Il s'agit d'avoir recours à la structure d'un système (ou cadre) conceptuel connu, pour conceptualiser un autre système. L'exemple courant de *CM* est LA DISCUSSION, C'EST LA GUERRE, qui s'exprime dans la langue par des expressions telle que *j'ai attaqué ses arguments*. La conceptualisation métaphorique, phénomène de pensée universel, est très présente en sciences, notamment dans le domaine biomédical : par exemple, en biologie cellulaire, la *CM* LES CELLULES SONT DES DISPOSITIFS EMETTEURS/RECEPTEURS se réalise dans la construction *Ces deux actions permettent à la cellule d'émettre ou de recevoir un signal*¹. Bien que les *CM* aient fait l'objet de nombreuses études dans différents domaines, ce n'est que relativement récemment qu'elles font l'objet d'investigations approfondies dans les langues de spécialité, par exemple, en

¹Alfandari, D. et coll. (1999) : Les protéines de la famille ADAM protéolyse, adhérence et signalisation. *ms/médecine sciences*, 15-10 : 1148-1151.

informatique (Meyer, 1997, 1998) et dans la médecine et les sciences (Vandaele, 2000, 2003 ; Temmerman, 2000, 2002 ; Vandaele et Lubin 2005).

D'ores et déjà, les études que nous avons menées vont dans le sens de notre hypothèse de travail, à savoir que la *CM* revêt une importance capitale pour comprendre les facteurs sous-jacents à la phraséologie des langues de spécialité, autrement dit à l'idiomaticité. De nombreux travaux se sont penchés sur la question des termes (voir Cabré, 2003, pour une synthèse des différents courants théoriques en terminologie) ; toutefois, la pratique de la traduction met en relief la nécessité de s'attarder aux éléments phraséologiques, relativement peu décrits, mais qui constituent en fait la pierre d'achoppement principale du processus de traduction. Ainsi en est-il de *message* ou de *signal* en biologie cellulaire, ou de la question de l'usage de verbes tels que *se jeter*, *s'aboucher*, *contourner*, etc. dans les descriptions anatomiques de veines et d'artères (*La grande veine saphène perce ensuite la gaine fémorale et se jette dans la veine fémorale*²).

La conclusion des études que nous avons menées est que la conceptualisation métaphorique opère notamment via une unité lexicale prédicative particulière, dont le repérage est lié à une impression de dissonance cognitive, que nous avons appelée **indice de conceptualisation** (Vandaele et Lubin, 2005). Ainsi, en anatomie, l'expression *la veine naît de la réunion de la veine digitale dorsale médiale du gros orteil et de l'arcade veineuse dorsale* évoque deux systèmes conceptuels : d'une part, l'utilisation prototypique du verbe *naître*, dans son sens premier, prend la forme suivante : *X naît de Y*, où *X* et *Y* sont des entités animées (*Y* n'étant pas toujours exprimée). D'autre part, dans une acception métaphorique de *naître*, *X* est exprimée par des noms de cours d'eau, *Y* étant exprimée de façon variable, le plus souvent par des noms de lieux géographiques (*Le Salat naît de neuf sources [...]*). L'impression de dissonance cognitive résulte de l'évocation mentale de la réalisation habituelle de *X* et de *Y*, ce qui conduit à conceptualiser les veines en premier lieu comme des cours d'eau, en second lieu comme des entités animées. Ces phénomènes cognitifs sont à rapprocher de l'intégration conceptuelle de Fauconnier et Turner (1998) et des représentations fictives (« *fictivity* ») de Talmy (2001).

Bien que ces manifestations de dissonances cognitives, observées isolément, puissent sembler arbitraires, elles tendent à former des sous-systèmes cohérents d'expressions plus ou moins prévisibles. En effet, certains termes dénotent des notions conceptualisées de façons semblables ou apparentées, différentes unités prédicatives induisant un effet de dissonance cognitive cohérent. Ainsi, en biologie cellulaire, plusieurs indices renforcent la conceptualisation des molécules (peu importe leur type) comme des êtres humains : *communauté*, *coloniser*, *mort*, *suicide*, *parenté*, *partenaire*, *famille*, *population*... À l'inverse, une même notion peut être conceptualisée simultanément de façon différente : les cellules peuvent être conceptualisées comme des personnes, mais aussi comme des usines ou des machines. Par conséquent, il semble possible de classer et de nommer des sous-systèmes de modes de conceptualisation et d'extraire les réseaux lexicaux qui leur sont associés, grâce au repérage et à la caractérisation des indices de conceptualisation et de leurs actants. Pour ce faire, nous avons fait appel à la méthode d'analyse actancielle reposant sur les principes de la composante de lexicologie explicative et combinatoire de la Théorie Sens-Texte (Mel'cuk et col. 1995).

² Garnier, E., traduit par Bossy, J. (1979). *Anatomie*, Paris, Doin : 190-207.

Le présent travail vise à fournir une méthode de repérage des indices de conceptualisation et de leurs actants, et à extraire les réseaux lexicaux caractérisant les *CM* parcourant la biologie et la médecine. La méthode est parfaitement généralisable à d'autres domaines. Appliquée à un corpus parallèle bilingue (anglais et français), elle permet de plus d'effectuer une comparaison des *CM* à l'oeuvre dans les deux langues et de schématiser les différences phraséologiques correspondantes. Dans la section suivante, nous discuterons du repérage et de l'annotation de ces indices. Cette étape fait appel à un sujet locuteur natif annotant manuellement les corpus. Enfin, nous présenterons, dans la section 3, la caractérisation des structures actanciennes, des combinaisons rencontrées et des différents modes de conceptualisation.

2. Repérage et annotation des indices de conceptualisation

Notre étude repose essentiellement sur le repérage des indices de conceptualisation et de leurs actants : cette étape est fondamentale et vise une analyse qualitative et quantitative des données. Notre corpus est constitué de textes spécialisés de médecine, en anglais et en français. Il s'agit d'un corpus parallèle, l'objectif étant de rendre compte des *CM* spécifiques dans chacune des langues envisagées. Nous avons constaté que la compilation des contextes contenant les éléments à repérer dans une base de données relationnelle (*BDR* ; Vandaele, 2002) comportait un certain nombre d'inconvénients : en effet, les indices sont dispersés dans le corpus, plusieurs indices peuvent être présents dans une même phrase, les anaphores forment un obstacle de taille et, enfin, la mise à jour des données dans une *BDR* est d'une grande complexité. Ces limites ont conduit à adopter une stratégie consistant plutôt à annoter directement les éléments à analyser au sein du corpus à l'aide du langage de balisage *XML*. *XML* est généralement envisagé comme un moyen de structurer les données en arborescence. Toutefois, il présente l'immense avantage de pouvoir établir des liens entre les noeuds de différentes branches (par exemple, des pointeurs vers des attributs d'identification *id*), ce qui autorise une représentation des interactions complexes repérables au sein d'un corpus. Étant donné que le principe consiste à repérer les indices de conceptualisation et les syntagmes exprimant leurs actants, l'annotation se limite essentiellement à la phrase. Le seul paramètre imposant de prendre en compte un échantillon de texte plus important est l'anaphore, bien que ce processus se situe au stade de l'analyse, l'annotation elle-même restant intraphrastique. Un ensemble de balises comprenant différents jeux d'attributs a été mis au point, et formalisé grâce à *XML Schema*, de manière à pouvoir annoter les différents éléments intervenant dans les expressions témoignant de conceptualisation métaphorique (Vandaele et coll., en préparation). En résumé, l'indice de conceptualisation est cerné par la balise `<concInd>` contenant les attributs suivants : *id* (numéro de l'élément dans la phrase) ; *lem* (forme lemmatisée de l'élément annoté ou de son antécédent dans les cas d'anaphore) ; *act1*, *act2*, etc. (attribut dont la valeur est le numéro de l'actant correspondant dans la phrase) ; *met1*, *met2*, etc. (attribut rendant compte du ou des actants prototypiques évoqués par l'indice de conceptualisation). L'expression de l'actant de l'indice de conceptualisation est, quant à elle, cernée par la balise `<lingEl>`, qui contient également les attributs *id* et *lem*. Dans les deux cas, ont également été prévus un attribut nommé *com*, dans lequel le sujet annotateur peut écrire librement un commentaire, et un attribut *jug*, permettant de repérer des éléments jugés étranges (*jug*="?") ou fautifs (*jug*="fautif"). L'exemple suivant³ illustre une phrase annotée selon cette convention.

³Alfandari, D. et coll. (1999). Les protéines de la famille ADAM protéolyse, adhérence et signalisation. *ms/médecine sciences*, 15-10 : 1148-1151.

```

<phr pos="113">
  Les <lingEl id="2" lem="cellule">cellules</lingEl>
  des crêtes neurales céphaliques sont issues de l'épithélium neural, et
  <concInd id="1" act1="2" act3="3"
  met1="personne,animal" met3="région géographique"
  lem="migrer">migrent</concInd> vers la
  <lingEl id="3">région ventrale de l'embryon</lingEl> où
  <lingEl id="5" lem="cellule">elles</lingEl>
  <concInd id="4" act1="5" act2="6"
  met1="personne,colon" met2="région géographique"
  lem="coloniser">colonisent</concInd> différents
  <concInd id="6" lem="territoire">territoires</concInd>
  pour former, entre autres, les structures de la face (muscles et cartilages).
</phr>

```

L'annotation du corpus fait appel à Oxygen⁴, un éditeur commercial XML multi-plateforme d'un usage relativement aisé, qui présente à l'annotateur, selon le contexte, différents choix pour les balises, les attributs et leurs valeurs.

3. Extraction des données

3.1. Fonctionnement général

La machinerie informatique qui permet de faire tourner le système repose sur un ensemble d'outils plus ou moins usuels liés à XML (*Extensible Markup Language*) et à HTML (*HyperText Markup Language*).

- Une transformation XSLT (*Extensible Stylesheet Language Transformation*) et une feuille de style CSS (*Cascading Style Sheets*) permettent de visualiser, dans un texte complet en format HTML, les éléments annotés repérés par un surlignage coloré. La transformation est réutilisée dans l'affichage des contextes lors de l'interrogation du corpus.
- XPath (*XML Path Language*), une composante de XSLT, permet de générer automatiquement des tables en format MySQL à partir des noeuds constitués par les éléments étiquetés afin de transformer périodiquement, selon l'état de l'annotation, le corpus en base de données. L'avantage de l'utilisation d'une base de données de type SQL (*Structured Query Language*) facilite l'interrogation du corpus et augmente la vitesse des requêtes. La base de données liée à l'étude des CM est constituée de deux tables : l'une pour les indices de conceptualisation, l'autre pour les actants⁵. Les tables contiennent également des méta-informations : nom du fichier, référence

⁴<http://www.oxygenxml.com/>.

⁵La structure des tables n'est pas exactement parallèle à l'annotation des fichiers XML. L'organisation des tables vise la facilité d'interrogation, tandis que les conventions d'annotation visent au confort du sujet annotateur. Ceci explique que les attributs met sont transportés dans la table des actants.

bibliographique du texte analysé, mots-clés, langue, ainsi que la phrase contenant le noeud indexé.

- Le serveur Web Apache coordonne la base de données *MySQL* à des pages Web par l'intermédiaire de scripts en *PHP*, autorisant à distance un accès multi-utilisateurs aux données.
- Les scripts *PHP* (*PHP : Hypertext Preprocessor*) permettent à Apache d'interroger la base *MySQL* et créent un formulaire dynamique tenant compte de la structure de la base. Le formulaire est conçu en vue d'un affichage sur Internet, de telle façon que son utilisation soit intuitive pour un usager non informaticien.

Ce système présente l'avantage d'une grande flexibilité, tant pour les corpus annotés (ajout de textes dans différentes langues sans perturbation du travail déjà effectué, processus de révision de l'annotation), que pour les jeux de balises et d'attributs (déterminés en fonction des caractéristiques à étudier). La même structure peut ainsi être transposée à différents projets. Enfin, outre un navigateur Internet tel que Firefox ou Explorer, la seule composante que le sujet annotateur doit apprendre à utiliser est le logiciel Oxygen.

3.2. Les requêtes SQL

La syntaxe d'une requête *SQL* a la forme suivante :

```
SELECT x1, x2, ..., xn
FROM [prédicat | actants]
WHERE y1 [AND|OR] y2 [AND|OR] ... [AND|OR] yn
GROUP BY z1, z2, ..., zn
ORDER BY w1, w2, ..., wn;
```

Dans le formulaire (un par table : prédicat ou actants), la requête se traduit par des champs dans lesquels s'inscrivent les paramètres faisant l'objet de la requête (*WHERE*), des cases à cocher qui sélectionnent les attributs à afficher (*SELECT*), ainsi que par des menus déroulants (*GROUP BY* et *ORDER BY*). Cette forme générale est appliquée spécifiquement à différentes requêtes : recherche de contextes, modes de conceptualisation, structures actanciennes des indices de conceptualisation, données quantitatives.

3.3. Les indices de conceptualisation et leurs contextes

La requête présentée ci-dessous correspond à un concordancier amélioré, permettant d'afficher les contextes contenant les indices de conceptualisation en fonction de différents critères. Par exemple, pour rechercher les contextes exprimant des modes de conceptualisation de type personnification (valeur de l'attribut *met*="personne"), elle prendra la forme suivante :

```
SELECT act_lemme, contexte
FROM actants
WHERE act_met="personne" AND lang="fr";
```

Voici quelques contextes obtenus :

- (1) *Dans un organisme, les cellules forment une **communauté** au sein de laquelle les échanges sont permanents.*
- (2) *Cependant, ces trois autres **membres** de la **famille** ne partagent pas le **rôle** de VLDLR et d'ApoER2 au cours du développement cérébral.*
- (3) *Cette revue décrit une nouvelle **famille** de protéines, la **famille** ADAM, qui pourrait **intervenir** dans ces deux types d'**interactions**.*

Dans l'exemple (1), *cellule* est l'actant de *communauté*, qui est l'indice de conceptualisation des cellules en tant que personnes. L'exemple (2) montre trois indices de conceptualisation cohérents témoignant de la CM de personnification (*membre, famille, rôle*). *VLDLR* et *ApoER2* désignent des molécules regroupées dans une « famille » partageant certaines caractéristiques, mais qui ne portent pas de nom particulier, contrairement à l'exemple (3) : *ADAM* est un sigle, fruit de l'humour des chercheurs, se comportant comme un nom propre. Ce type d'affichage permet ainsi de visualiser en contexte les réseaux lexicaux exprimant un mode de conceptualisation.

3.4. Les modes de conceptualisation

Un des objectifs de l'étude est d'obtenir l'ensemble des modes de conceptualisation spécifiques du domaine. Par exemple, comment sont conceptualisées les protéines ? Le formulaire permet d'ajouter à la requête un critère de classement des résultats par ordre alphabétique (ORDER BY) ainsi qu'un critère de groupement (GROUP BY), ce qui permet de générer des données quantitatives.

```
SELECT act_met, count(*)
FROM actants
WHERE act_lemme="protéine", lang="fr"
GROUP BY act_met;
```

L'application de la requête au sous-ensemble annoté du corpus⁶ génère le résultat suivant :

52	personne
5	personne, criminel
1	bateau
15	(vide)
7	?

Tableau 1 : La conceptualisation des protéines

⁶L'annotation du corpus est en cours, les résultats chiffrés ne sont donc pas définitifs.

Dans 57 occurrences, les protéines sont conceptualisées comme des personnes, plus spécifiquement un « criminel impliqué dans un délit » dans 5 occurrences (*D'autres protéines impliquées dans ces phénomènes de résistance aux chimiothérapies*⁷ ; voir Vandaele, 2003). Dans une occurrence, le cadre évoqué est celui d'un « bateau » à l'ancre (*la protéine est ancrée dans la membrane*). Enfin, 15 occurrences du terme n'ont pas encore reçu de valeur de leur attribut *met* et sept autres doivent être revues (?). Cet exemple est représentatif d'un mode de travail itératif dans lequel une réflexion globale sur les valeurs des attributs peut être entreprise à partir de données déjà annotées, ce qui permet notamment la correction d'erreurs et l'intervention éventuelle de plusieurs sujets annotateurs s'accordant sur un consensus.

3.5. La structure actancielle de l'indice de conceptualisation

Pour l'affichage de la structure actancielle de l'indice de conceptualisation associé à une *CM* particulière, nous utilisons une requête du type suivant :

```
SELECT act1_lemme, predicat_lemme, act2_lemme, contexte
FROM predicats
WHERE act2_met="message", lang="en"
ORDER BY predicat_lemme;
```

Le résultat de la requête affiche ainsi un ensemble de verbes cohérents avec la métaphore du message :

(4) *gene to encode type II receptor/type I receptor*⁸ :

*Molecular cloning has yielded genes encoding two type II receptors (ActRII and ActRIIB) and at least two type I receptors (ALK2 and ALK4)*⁹.

(5) *rhombomere to express molecule* :

*We suggest that during regional specification of the early hindbrain, presumptive rhombomeres express distinct signalling molecules that regulate cell identity*¹⁰.

3.6. Données quantitatives

L'expression `count(*)`, en association avec `GROUP BY`, permet de compter le nombre d'occurrences de combinaisons spécifiées par un critère de regroupement `GROUP BY`. L'exemple suivant fait état des agencements de modes de conceptualisation associés à *to migrate (X migrate from Y to Z)* :

⁷Alliet J. et Lalégérie P. (1997) *Cytobiologie*, Ellipses, Paris : 468.

⁸Dans ce cas, l'actant sémantique de *to encode* se réalise syntaxiquement par deux éléments, *type II receptor* et *type I receptor*.

⁹Zimmerman, C.M. (1995). *Activin receptors : cellular signalling by receptor serine kinases* », *Extracellular Regulators of Differentiation and Development*, sous la direction de K. Chapman, S.P. Jackson, D.G. Wilkinson, G.G. Lunt, London, Portland Press Ltd : 25-38.

¹⁰Irving, C. (1996). *Cell-cell interactions and segmentation in the developing vertebrate hindbrain. Extracellular Regulators of Differentiation and Development*, sous la direction de K. Chapman, S.P. Jackson, D.G. Wilkinson, G.G. Lunt, London, Portland Press Ltd : 85-95.

```

SELECT act1_met, act2_met, act3_met, COUNT(*)
FROM predicats
WHERE predicat_lemme="migrate, to" AND lang="en"
GROUP BY act1_met, act2_met, act3_met
ORDER BY act1_met, act2_met, act3_met;

```

Ceci donne les résultats suivants :

act1_met	act2_met	act3_met	total
animal ; person			1
animal ; person	geographical area		4
animal ; person		geographical area	4
animal ; person	geographical area	geographical area	1
person	geographical area	geographical area	1

Tableau 2 : Nombre d'occurrences des combinaisons de *migrer*

Ce type de requête permet, notamment, d'analyser la cohérence de l'annotation¹¹ et l'expression des différents actants.

4. Conclusion et perspectives

L'analyse des tableaux obtenus à l'aide des requêtes permet d'établir les réseaux lexicaux mis en oeuvre dans la conceptualisation métaphorique en biologie cellulaire (Vandaele et col. 2006, soumis). Par exemple, la métaphore de l'enquête est mise en évidence par les verbes *impliquer*, *élucider*, *identifier*. Les actants de ces verbes sont conceptualisés en tant que personne, criminel, délit, ... Il devient ainsi possible d'examiner de façon synthétique l'existence de *CM* dans les deux langues, avec leurs réalisations linguistiques respectives. Les travaux actuels visent à raffiner l'attribution des valeurs de l'attribut *met*, notamment la hiérarchisation des projections (UNE MOLECULE EST UNE PERSONNE est plus générale que UN AGENT PATHOGENE EST UN CRIMINEL).

La relative lourdeur imposée par l'annotation manuelle est largement compensée par la souplesse des formulaires d'interrogation, par la possibilité des corrections et des mises à jour du corpus, ainsi que par l'adaptabilité du contenu des balises. Il est en effet aisé de concevoir des attributs *ad hoc*, en fonction des caractéristiques textuelles à analyser. Par conséquent, cette méthode de travail dépasse largement l'utilisation proposée, et constitue un outil extrêmement puissant d'analyse textuelle, notamment pour le repérage terminologique (nous pensons notamment au repérage des néologismes), la constitution d'ontologie et l'extraction de données permettant la constitution de dictionnaires ou de bases de données terminologiques. Par ailleurs, le résultat de l'annotation de corpus parallèles, outre l'analyse

¹¹Il faudrait, ici, vérifier la raison pour laquelle le sujet annotateur a utilisé *animal ;person* dans 10 occurrences, et *person* une fois.

fine des phénomènes textuels dans plusieurs langues, se veut un outil d'aide à la traduction et pédagogique. Enfin, il faut rappeler que l'un des problèmes du traitement automatique de la langue, surtout en français, est lié à l'absence de corpus « étalons » de référence, ces derniers étant cependant indispensables au calcul du rappel. Des corpus annotés finement par la méthode que nous proposons pourraient donc constituer l'outil indispensable à la mise au point de méthodes automatisées de repérage des indices de conceptualisation métaphorique, et, plus généralement, de tout autre phénomène repérable en corpus.

Références

- Cabré, M.T. (2003). Theories of Terminology : Their Description, Prescription and Explanation. *Terminology* 9 (2) : 163-199.
- Fauconnier, G. et Turner M. (1998). Conceptual Integration Networks. *Cognitive Science* 22 (2) : 133-187.
- Lakoff G. et M. Johnson (1980/2003). *Metaphors We Live By – With a New Afterwords*. Chicago, University of Chicago Press.
- Mel'cuk, I.A., Clas, A. et Polguère, A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve (Belgique), Duculot / Aupelf – UREF.
- Meyer, I. et coll. (1997). Metaphorical Internet Terms : A Conceptual and Structural Analysis. *Terminology* 4 (1) : 1-33.
- Meyer, I. et coll. (1998). Metaphorical Internet Terms in English and French. In *Proceedings of the 9th EURALEX International Congress*, Liège (Belgique), Université de Liège.
- Temmerman, R. (2000). *Towards New ways of terminology Description : The sociocognitive Approach*. Amsterdam, John Benjamins.
- Temmerman, R. (2002). *Metaphorical models and the translator's approach to scientific texts*. *Linguistica Antverpiensa* 1 : 211-226.
- Talmy, L. (2001). *Toward a cognitive semantics*. Volume I, *Concept structuring systems*, Cambridge, MIT Press.
- Vandaele, S. (2000). Métaphores conceptuelles et traduction biomédicale. *La traduction : théorie et pratiques, actes du colloque international Traduction humaine, traduction automatique, interprétation*, sous la direction de S. Mějri, T. Baccouche, A. Clas. G. Gross, Tunis, 28-29 septembre 2000, Publications de l'ENS : 393-404.
- Vandaele, S. (2003). Métaphores conceptuelles et traduction médicale. *TTR*, (XV) 1 : 223- 239.
- Vandaele, S. et L. Lubin (2005). Approche cognitive de la traduction dans les langues de spécialité : vers une systématisation de la description de la conceptualisation métaphorique. *META*, numéro spécial dirigé par H. Lee-Jahnke, (20) 2 : 415-431.
- Vandaele, S., S. Boudreau, L. Lubin et E. Marshman (2006). La conceptualisation métaphorique en biomédecine : vers une typologie des indices de conceptualisation. soumis à *Glottopol*.
- Vandaele, S., S. Boudreau, L. Lubin et E. Marshman (en préparation). Semantic Annotation in XML : A Method for the Description of Metaphorical Conceptualizations in Specialized Languages.

