

---

# RÉSEAU SÉMANTIQUE ET APPROCHE COMPONENTIELLE DES BASES DE DONNÉES LEXICALES MULTILINGUES

Marc Van Campenhoudt: *Centre de recherche TERMISTI, Institut supérieur de traducteurs  
et interprètes (34, rue Joseph Hazard B-1180 Bruxelles)*  
([marc.van.campenhoudt@euronet.be](mailto:marc.van.campenhoudt@euronet.be))

---

## Abstract

The author posits that a componential approach seems, at the moment, more appropriate in specialized lexicography or in terminography than in general lexicography. Although the use of lattices has its merits, the author observes that trees appear to be more effective to find an equivalent in case of lexical gaps.

## 1. Introduction

On a assisté au cours de la dernière décennie à l'émergence de plusieurs modèles de gestion de l'équivalence fondés sur l'exploitation d'un réseau sémantique. Ces différentes approches sont marquées par un point commun : l'identification d'un lien pertinent entre la relation d'hyponymie et la recherche d'équivalents traductifs en cas d'absence d'isomorphisme entre les langues (Lyons 1970 : 45). L'idée d'identifier l'équivalent traductif à travers une approche componentielle qui permette de choisir un hyperonyme est un point commun entre l'approche de Simullda et celle que nous avons déjà présentée ici même (Van Campenhoudt 2001). Nous nous bornerons à évoquer brièvement dans les lignes qui suivent quelques points qui n'ont pas été approfondis dans une comparaison de ces deux approches (Janssen et Van Campenhoudt à paraître).

## 2. Le modèle de Termisti

Le modèle de Termisti est né dans le cadre d'une tentative de modélisation d'un célèbre dictionnaire de marine trilingue, *De la quille à la pomme de mâât* (Paasch 1901), dont l'organisation macrostructurelle s'est avérée particulièrement novatrice (Van Campenhoudt 1994). Le **principe d'équivalence** appliqué dans ce dictionnaire peut se résumer comme suit :

Si A de  $L_1$  (langue 1) est équivalent à  $\alpha$  de  $L_2$  (langue 2) et si A de la  $L_1$  est équivalent à  $\beta$  de  $L_2$  alors que  $\alpha$  de  $L_2$  n'est pas synonyme de  $\beta$  de  $L_2$ , c'est que probablement A de  $L_1$  possède deux sens qui devraient être différenciés par deux entrées distinctes au sein du dictionnaire.

Ce principe, qui est très proche d'un test permettant d'isoler les nœuds au sein d'un réseau sémantique monolingue (Levrat et Sabah 1990 : 93), a montré sa grande efficacité dans le cadre du projet européen Dhydro, s'agissant de fusionner au sein d'une même base de

données multilingue trois dictionnaires monolingues consacrés à l'hydrographie (Descotte *et al.* 2001).

Nous avons déjà montré que l'application du principe d'équivalence recoupe très souvent un mécanisme de recours à l'hyperonyme - l'hyperonymase (Van Campenhout 2001) - voire de recours à certaines formes de méronymies (Van Campenhout 1996). Ce principe permet d'expliquer le grand nombre de dégroupements homonymiques opérés dans un dictionnaire multilingue classique, imprimé ou affiché sur un écran<sup>1</sup>. Dans un tel dictionnaire, les hyperonymes disparaissent dès lors que leur extension est entièrement reproduite à un niveau inférieur pour combler des trous lexicaux (figure 1). Ce mécanisme de suppression est propre à une approche lexicographique fondée sur l'homonymie et regroupant ensemble les termes considérés comme équivalents. Le dictionnaire qui suit une semblable approche peut être qualifié de « fusionnel » dans la mesure où il rassemble au sein d'une même entrée (ou fiche) tous les découpages du sens référentiel qui lui semblent compatibles.

### 3. Vers une meilleure exploitation des attributs et du réseau

Le respect du principe d'équivalence implique, bien entendu, une révision du découpage des entrées à chaque fois que l'on ajoute une nouvelle langue. Le non-respect de ce principe lors de l'ajout d'idiomes supplémentaires dans certaines grandes bases terminologiques est regrettable. Il s'explique par la masse de travail – et donc le coût - que représenterait la révision de toutes les équivalences déjà établies et de toutes les définitions proposées.

Il est donc très tentant d'envisager, comme le fait Maarten Janssen, un système de dictionnaire électronique multilingue qui soit tout à la fois plus économique et plus respectueux de la réalité de chacune des langues envisagées. Il s'agit de concevoir un dictionnaire davantage « satellitaire », où chaque langue est décrite de manière autonome et

reliée par une forme d'interlingua. Les équivalences y sont soit déclarées par le biais d'un lien *ad hoc* – tel est le cas dans *Eurowordnet* - soit calculées au départ d'une description de traits sémantiques minimaux. Simullda relève de cette dernière perspective, de même que le modèle de Termisti. Même si ce dernier modèle n'a pas fait l'objet d'un développement informatique permettant de calculer l'équivalence, celle-ci peut-être entièrement déduite du système de matrice proposé (Van Campenhoudt 2001 : 192-193), que ce soit par l'exploitation du lien d'hyponymie déduit de la combinatoire des traits [+], [-] et [0]<sup>2</sup> ou par la lexicalisation de l'équivalent manquant sur la base des traits distinctifs de l'hyponyme (tableau 1 et figure 2).

#### 4. Treillis vs arborescences : quel interlingua ?

Simullda est fondé sur l'analyse formelle de concepts : le système génère non pas des arborescences, mais des treillis<sup>3</sup>. Les treillis envisagent toutes les possibilités combinatoires et donc toutes les acceptions interlinguistiques imaginables au sein de l'interlingua, y compris les plus aberrantes. À la différence de Simullda, l'exploitation du modèle componentiel (et arborescent) de Termisti ne suppose pas de passer par un treillis où chaque nœud correspond à une combinaison d'attributs définitionnels susceptible d'être désignée. Le seul interlingua du modèle de Termisti concerne les traits sémantiques qui réunissent ou distinguent des termes : un terme n'y est pas perçu comme l'étiquette d'un nœud du réseau, mais comme entretenant un réseau de relations arborescentes avec d'autres termes.

Pour des raisons qui demeurent à explorer (Janssen et Van Campenhoudt à paraître), les treillis ne permettent pas, pour l'heure, de résoudre divers cas de non-isomorphismes que nous avons déjà décrits et qui sont correctement traités par notre système (voir, p.ex., Van Campenhoudt 2001 : 202-203). Il s'agit plus précisément de problèmes de chevauchement partiel ressemblant à celui décrit dans la figure 15 de la communication de Maarten Janssen.

## 5. Modèle componentiel : principales difficultés

Les possibilités de calcul offertes par Simullda et par Termisti plaident pour la réalisation de bases de données lexicales multilingues fondées sur une description componentielle des acceptions qui permette de lier des sèmes de langues différentes<sup>4</sup>. Toutefois, la réalisation de tels systèmes se heurte encore à diverses difficultés qui compliquent leur mise en œuvre.

5.1. Les attributs définitionnels de Simullda comme les traits sémantiques de Termisti sont déduits de définitions figurant dans les dictionnaires. Une intervention humaine est indispensable pour les importer dans l'interlingua, lequel permet d'identifier les traits communs aux lexies relevant de langues différentes. Si cette intervention coûteuse en ressources demeure nécessaire lors de l'ajout d'une nouvelle langue, elle permet toutefois d'éviter un redécoupage manuel complet de l'ensemble des entrées équivalentes et de leurs définitions au sein du dictionnaire. L'idéal serait de développer un système d'interprétation de définitions multilingues, capable de générer les traits de l'interlingua.

5.2. Les grilles de traits permettent d'isoler « automatiquement » des relations d'hyponymie mais il est beaucoup plus ardu d'en extraire d'autres types de liens (causal, spatial, temporel...), qui n'échapperont pas, à ce stade, à une déclaration humaine.

5.3. À la différence de Simullda, le modèle de Termisti porte uniquement sur des vocabulaires spécialisés. La possibilité de travailler par microdomaines distincts permet d'envisager un nombre raisonnable de traits inhérents pour un ensemble limité et cohérent de taxèmes. La description de la langue générale requiert bien davantage de sèmes ; elle pose beaucoup plus souvent le problème de la connotation (les sèmes afférents) et du prototype (un fleuve est une grande rivière) par rapport à l'usage en langue spécialisée, moins connoté et plus attentif aux conditions nécessaires et suffisantes (un fleuve se jette dans la mer). Le travail sur des vocabulaires spécialisés simplifie également le choix des traits distinctifs

pertinents. Ce constat général doit toutefois être relativisé dans le cas de terminologies marquées par une forte diversification lexicale ou sémantique.

5.4. L'attribution de traits sémantiques distinctifs ne suffit assurément pas à créer un dictionnaire, lequel doit s'accompagner de tous les champs indispensables à une fine description lexicale, syntaxique et sémantique. Les grilles de traits ne constituent donc qu'un aspect d'un ensemble plus vaste d'informations appelé à constituer un « métadictionnaire » (Van Campenhoudt 2000).

5.5. Lorsque, confronté à un cas de trou lexical, le système recherche une équivalence, il doit être capable de proposer – en fonction du contexte - un choix entre un simple hyperonyme et une lexicalisation des traits distinctifs.

5.6. La lexicalisation des traits sémantiques pour générer des équivalents traductifs suppose que soit effectué un encodage manuel des expressions *ad hoc* dans chaque langue et que celles-ci s'appliquent correctement à tout hyperonyme concerné.

## 6. Perspectives

Le modèle de Simullda arrive indubitablement à son heure. Les temps sont mûrs pour qu'émerge un projet de plus grande ampleur visant à construire une base de données terminologique multilingue intégrant tout à la fois une véritable description lexicale et une gestion de l'équivalence fondée sur un modèle componentiel. Disant cela, nous ne faisons que plaider, avec beaucoup d'autres, pour le développement d'une véritable base de connaissances terminologique.

## Bibliographie

**Descotte, S., Husson, J.-L., Romary, L., Van Campenhoudt, M. et Viscogliosi, N.** 2001.

« Dhydro: a generic environment developed to edit and access multilingual terminological data on the Internet », in Vainio, J. (éd.). *Maritime Terminology: Dictionaries and Education, Proceedings of the Second Conference on Maritime Terminology, 11-12 May 2000, Turku, Finland*. (Publications from the Centre for Maritime Studies, University of Turku, A36.) Turku: University of Turku, 47-61.

**Levrat, B. et Sabah, G.** 1990. « "Sorte de", une façon de rendre compte de la relation d'hyponymie/hyperonymie dans les réseaux sémantiques » *Langages* 98 : 87-102.

**Lyons, J.** 1970. *Linguistique générale. Introduction à la linguistique théorique*. (Langue et langage.) Paris : Larousse.

**Janssen, M. et Van Campenhoudt, M.** à paraître : « Terminologie traductive et représentation des connaissances : l'usage des relations hyponymiques », soumis en novembre 2003 à Depecker, L. (dir.), *La terminologie : nature et enjeux*. in *Langages*. à paraître.

**Paasch, H.** 1901. *De la quille à la pomme de mât. Dictionnaire de marine en anglais, français et allemand illustré de nombreux dessins explicatifs*. 3<sup>e</sup> édit. Anvers : H. Paasch et Hamburg : Eckardt & Messtorff.

**Sowa, J.F.** 2001. *Building, Sharing, and Merging Ontologies*.

<http://users.bestweb.net/~sowa/ontology/ontoshar.htm>.

**Van Campenhoudt, M.** 1994. *Un apport du monde maritime à la terminologie notionnelle multilingue : étude du dictionnaire du capitaine Heinrich Paasch « De la quille à la*

*pomme de mâ* » (1885-1901). Paris : Université Paris XIII (thèse de doctorat en sciences du langage).

**Van Campenhoudt, M.** 1996. « Recherche d'équivalences et structuration des réseaux notionnels : le cas des relations méronymiques ». *Terminology*. 3:1 : 53-83.

**Van Campenhoudt, M.** 2000. « De la lexicographie spécialisée à la terminographie : vers un “ métadictionnaire ” ? » in Ph. Thoiron et H. Béjoint (dir.), *Le sens en terminologie*. (Travaux du C.R.T.T.) Lyon : Presses universitaires de Lyon, 127-52.

**Van Campenhoudt, M.** 2001. « Pour une approche sémantique du terme et de ses équivalents ». *International Journal of Lexicography*. 14.3 : 181-209.



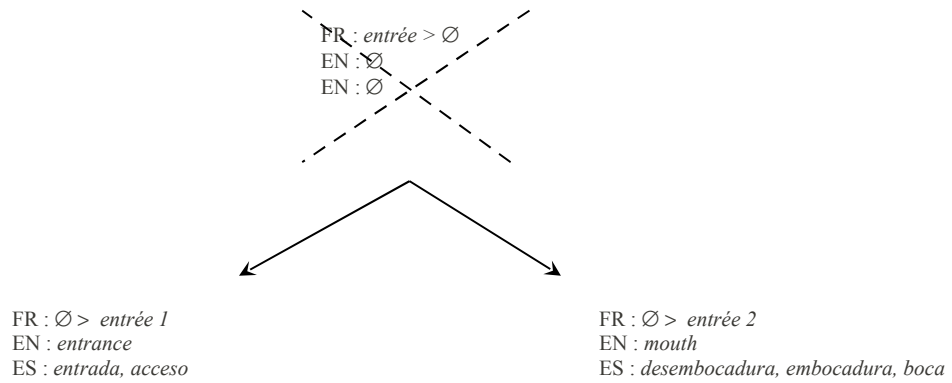
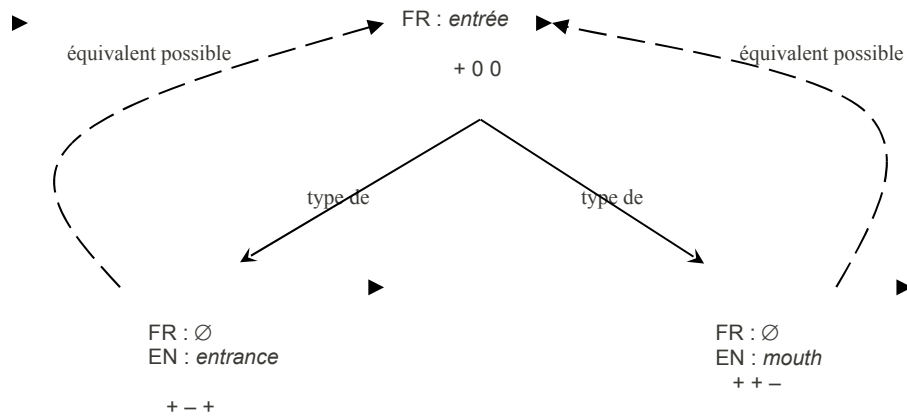


Figure 1. Dans le dictionnaire multilingue classique, le terme *entrée* perd sa valeur d'hyperonyme lors de l'établissement de l'équivalence par hyperonymase (d'après Van Campenhoudt 2001 : 197).

Figure 2. Un système fondé sur une description des traits sémantiques permet de récupérer un



hyperonyme comme équivalent traductif, voire de générer une désignation plus précise lorsque le contexte l'exige (*entrée d'une baie vs entrée d'un port*)

**Tableau 1.** Grille sémique des équivalents anglais et espagnols du terme *entrée* (Van Campenhoudt 2001 : 195)

	<b>Passage maritime</b>	<b>Entre baie et mer</b>	<b>vers infrastructure portuaire</b>
<i>entrée</i>	+	0	0
<i>entrance</i>	+	–	+
<i>entrada, acceso</i>	+	–	+
<i>mouth</i>	+	+	–
<i>desembocadura, embocadura, boca</i>	+	+	–

## <sup>1</sup>Notes

. Nous opposons ici le dictionnaire multilingue au dictionnaire bilingue. Dans un dictionnaire bilingue classique, le dégroupement n'est pas aussi apparent dans la mesure où l'ouvrage est découpé en deux dictionnaires distincts qui inversent les langues source et cible et adoptent, le plus souvent, une perspective polysémique. Il reste pourtant vrai que le découpage de la langue cible y préside à la diversification sémantique en autant d'acceptions qu'il y a de traductions différentes.

<sup>2</sup>. L'hyperonyme possède les traits de l'hyponyme ou les remplace par une valeur [0].

<sup>3</sup>. Concernant la distinction entre treillis et arborescence, lire Sowa (2001).

<sup>4</sup>. S'agissant de ne traiter que de la problématique de l'équivalence, on n'abordera pas ici un avantage fondamental de l'approche componentielle qui est de permettre d'éviter tous les errements d'une déclaration humaine de la relation d'hyponymie. Sous cet angle, on déplorera que la rigueur initiale de *Wordnet* ait été dévalorisée par un travail parfois trop peu consciencieux (consulter, p.ex., l'incroyable liste des hyponymes de *ship*).