

Marc Van Campenhoudt
Centre de recherche Termisti
Institut supérieur de traducteurs et interprètes
Bruxelles
www.termisti.refer.org

Lexicographie vs terminographie : quelques implications théoriques du projet DHYDRO

1 Introduction

Le projet DHYDRO - comme *Dictionnaire hydrographique* - a été financé par le programme européen *Multilingual Information Society (MLIS)*. Il avait pour objectif d'informatiser le dictionnaire de l'Organisation hydrographique internationale (O.H.I., Monaco) et de créer une interface client-serveur permettant à une équipe internationale de rédiger cet ouvrage à distance, par le biais des autoroutes de l'information.

Le détail du projet a déjà donné lieu à plusieurs communications, consultables sur le site internet du consortium DHYDRO¹. Notre objectif sera ici de montrer que les nouvelles techniques de balisage de l'information mises en œuvre dans ce projet conduisent à reconsidérer les différences, traditionnellement jugées fondamentales, entre la lexicographie et la terminographie².

2 Le Dictionnaire hydrographique

Le *Dictionnaire hydrographique* de l'O.H.I. présente au départ toutes les apparences d'une lexicographie classique, avec des entrées polysémiques classées selon l'ordre alphabétique. Il s'agit plus précisément d'une lexicographie spécialisée monolingue, où les entrées sont souvent des syntagmes et dans laquelle les définitions intègrent des développements encyclopédiques.

1. www.loria.fr/projets/MLIS/DHYDRO/. Les partenaires du projet étaient l'Institut für deutsche Sprache (Mannheim), le LORIA (Nancy), l'Organisation hydrographique internationale (Monaco), le Service hydrographique de la Marine (Paris), le Centre de recherche TERMISTI (Bruxelles).

2. Nous publierons prochainement dans l'*International Journal of Lexicography* une analyse théorique plus argumentée montrant les nombreuses convergences méthodologiques entre la terminographie et la lexicographie et remettant en cause la primauté du concept dans la première de ces disciplines (Van Campenhoudt, à paraître).

Avant que ne débute le projet européen, le dictionnaire était publié en anglais, espagnol et français sous la forme de trois volumes distincts. Les éditions française et espagnole paraissaient toujours après la publication de la nouvelle édition anglaise, dont elles s'inspiraient. Ce n'est qu'accessoirement que le dictionnaire pouvait servir à des fins de traduction, grâce à un système de numérotation fondé sur les entrées de la version anglaise et qui servait de clé de consultation pour pénétrer dans un autre volume.

Alignement. m

- 85-2696 **a)** Ligne droite définie par deux ou plusieurs AMERS clairement indiqués sur une CARTE, le long de laquelle un navire peut faire ROUTE en toute sécurité, pour entrer dans une PASSE, parer un danger, etc. Voir aussi RELÈVEMENT D'UN ALIGNEMENT.
- 4211 **b)** Ligne droite définie par deux ou plusieurs objets (dont on dit alors qu'ils constituent un alignement). On dit d'un observateur qu'il est situé sur l'alignement quand il voit ces objets dans le même plan vertical. Voir aussi PASSAGE.
- 5596 **c)** En NAVIGATION, position de deux objets fixes et distants l'un de l'autre, quand ils sont vus en ligne, (on dit : l'un par l'autre) pour l'observateur. La ligne passant par ces deux objets et l'observateur est alors une LIGNE DE POSITION. C'est aussi le relèvement de cette ligne tel qu'il est vu par un observateur venant de la mer.
- 4217 **Alignement m de contrôle des compas magnétiques.**
ALIGNEMENT dont la direction est connue et qu'on utilise pour la détermination de la DÉVIATION d'un COMPAS MAGNÉTIQUE.

Figure 1 : extrait du volume français du Dictionnaire hydrographique

3 Du livre à la base de données

S'agissant d'informatiser un dictionnaire multilingue appelé à prendre en compte un nombre croissant de langues (l'allemand, le mandarin et le japonais devraient bientôt s'adjoindre), il est rapidement apparu qu'il était impossible de concevoir une base de données qui calque le fonctionnement d'un dictionnaire de traduction imprimé. Rappelons quelques caractéristiques fondamentales d'un semblable dictionnaire :

- travail par couples de langues source et cible, consistant en autant de volumes ou de parties (3 langues = 6 couples de langues³) ;
- indexation alphabétique des entrées de la langue source ;
- équivalents en langue cible fournis en fonction des acceptions en langue source.

Une étude préliminaire (Bessero *et al.* 1999) proposait de suivre les prescriptions de la *Text Encoding Initiative*⁴ (T.E.I.) et d'utiliser un système de pointeurs de sens à sens pour gérer l'équivalence. Dans le cadre du projet, nous avons rapidement démontré la complexité d'une telle approche :

3. Rappelons la formule : Nombre de couples de langues = (Nombre de langues)² - Nombre de langues.

4. <http://www.tei-c.org/>.

- la moindre modification d'une entrée ou d'une acception dans une langue suppose la mise à jour des pointeurs dans chacune des autres langues.
- l'ajout d'une nouvelle langue suppose de revoir l'ensemble du mécanisme de pointage, chaque acception des entrées de chaque langue devant en permanence pointer vers les acceptions des entrées de chacune des autres langues.

Il est intéressant de constater que les prescriptions de la T.E.I. concernent des livres imprimés et considèrent les dictionnaires sous ce seul angle. Or, il nous paraît évident qu'à la différence du livre, le dictionnaire est avant tout une base de données :

- l'inventaire des catégories de données d'un dictionnaire est nettement plus élevé que celui des types d'informations structurant les romans, pièces de théâtre et autres recueils de poèmes ;
- un dictionnaire n'est pas lu linéairement, mais se consulte : les mécanismes de renvoi y sont nettement plus complexes ;
- certaines catégories de données sont interdépendantes et la structure ainsi formée n'est pas nécessairement explicite.

Revenant sur leur travail, les concepteurs du chapitre 12 de la T.E.I., consacré aux dictionnaires, ont reconnu cet état de fait. Comme le notent très bien Nancy Ide et Jean Véronis (1996 : 174) : « Les dictionnaires ont bien évidemment l'apparence de textes et possèdent de nombreuses caractéristiques communes à tous les types de textes. Néanmoins, les utilisateurs ne lisent pas en principe les dictionnaires de manière linéaire de A à Z comme ils le font pour la plupart des textes, mais accèdent à des entrées à partir d'une clé (la vedette) dans le but de récupérer divers champs d'information associés à cette clé (prononciation, information grammaticale, étymologie, définitions, etc.). Cet accès non linéaire est typique de l'accès aux bases de données. »

De ce point de vue, la T.E.I. peut juste rendre compte d'une structure superficielle, de nature éditoriale et pleine d'implicite, ce qui n'offre guère d'intérêt pour qui veut exploiter toute la puissance d'un véritable dictionnaire électronique.

Il est d'ailleurs intéressant de constater que dans le chapitre 12 de la T.E.I., le dictionnaire bilingue est vu comme un ensemble de deux livres :

- le livre 1 permet de traduire de la langue A vers la langue B ;
- le livre 2 permet de traduire de la langue B vers la langue A.

Les équivalents sont mentionnés sous chaque acception balisée <sense> à l'aide d'une balise <tr>. On imagine aisément la difficulté qu'il y aurait à suivre les prescriptions de la T.E.I. pour baliser une version en 6 langues (DE, EN, ES, FR, JA, ZH) du *Dictionnaire hydrographique*. Cela supposerait la création de 30 livres (ou fichiers balisés) proposant sous chaque entrée des traductions vers 5 langues différentes...

La conclusion semble logique : le modèle polysémique ne permet pas une gestion efficace des dictionnaires polyglottes. La seule manière rationnelle de traiter un grand nombre de langues semble, dès lors, d'adopter une logique de base de données relationnelle et de regrouper dans une même fiche toutes les informations qui concernent une même acception, c.-à-d. de fonder sa démarche sur la monosémie et l'homonymie. Il apparaît donc très paradoxal que le chapitre 12 de la T.E.I. prétende s'appliquer aussi bien aux dictionnaires polyglottes qu'aux dictionnaires monolingues, alors même que le traitement proposé n'est guère adéquat au-delà de deux langues.

Force est d'ailleurs de remarquer qu'on ne connaît guère d'exemples de dictionnaires de la langue générale qui couvrent plus de deux langues : à partir de trois langues, l'ouvrage est habituellement un dictionnaire de la langue spécialisée (Van Campenhoudt 2000). Or, le balisage des dictionnaires spécialisés, relève du chapitre 13 de la T.E.I., intitulé - en toute logique - « Terminological Databases ».

Malheureusement, le seul traitement en base de données proposé dans ce chapitre de la T.E.I. est entièrement pris en charge par les tenants de l'École de Vienne⁵, très actifs à l'ISO, lesquels proposent une approche conceptuelle, ce qui - en bonne épistémologie - n'équivaut pas à une approche fondée sur la monosémie : le concept n'est pas le sens !

4 Les choix du projet DHYDRO

4.1 Hypothèse forte : une granularité maximale

L'ambition de créer un dictionnaire électronique multilingue qui puisse être rédigé en concertation, grâce à internet, par une équipe internationale de rédacteurs impliquait obligatoirement de privilégier une approche monosémique. Il ne pouvait toutefois être question de recommencer le dictionnaire *ex nihilo* sous prétexte qu'en bonne doctrine terminologique, il convenait de d'abord se mettre d'accord sur un inventaire de concepts.

Les partenaires ont préféré formuler l'hypothèse forte que s'ils arrivaient à identifier et baliser correctement tous les types d'informations - implicites ou explicites - présentes dans les fichiers de traitement de texte des anciennes versions imprimées du dictionnaire, ils devraient ensuite pouvoir les réorganiser conformément aux exigences d'un traitement multilingue. Leur idée était qu'au départ d'un même balisage X.M.L. des données, il est possible de générer indifféremment :

5. Le préambule du chapitre 13 précise : « this chapter was first published in December, 1993, as part of TEI P2. Since that publication, it has been revised slightly for the sake of consistency with the rest of the Guidelines and with the work of Technical Committee 37 of the International Organization for Standardization (ISO) on ISO DIS 12 200. »

- ❖ une présentation « lexicographique » caractérisée par :
 - des entrées polysémiques
 - un classement alphabétique
 - un éclatement des langues
- ❖ une présentation « terminographique » caractérisée par :
 - des entrées monosémiques (ce qui implique l’homonymie)
 - un regroupement des synonymes
 - un regroupement des langues

Voici un schéma simplifié de la différence de présentation entre la version monolingue et polysémique du *Dictionnaire hydrographique* et sa nouvelle version multilingue et monosémique (d’après Descotte *et al.* 2001a : 70) :

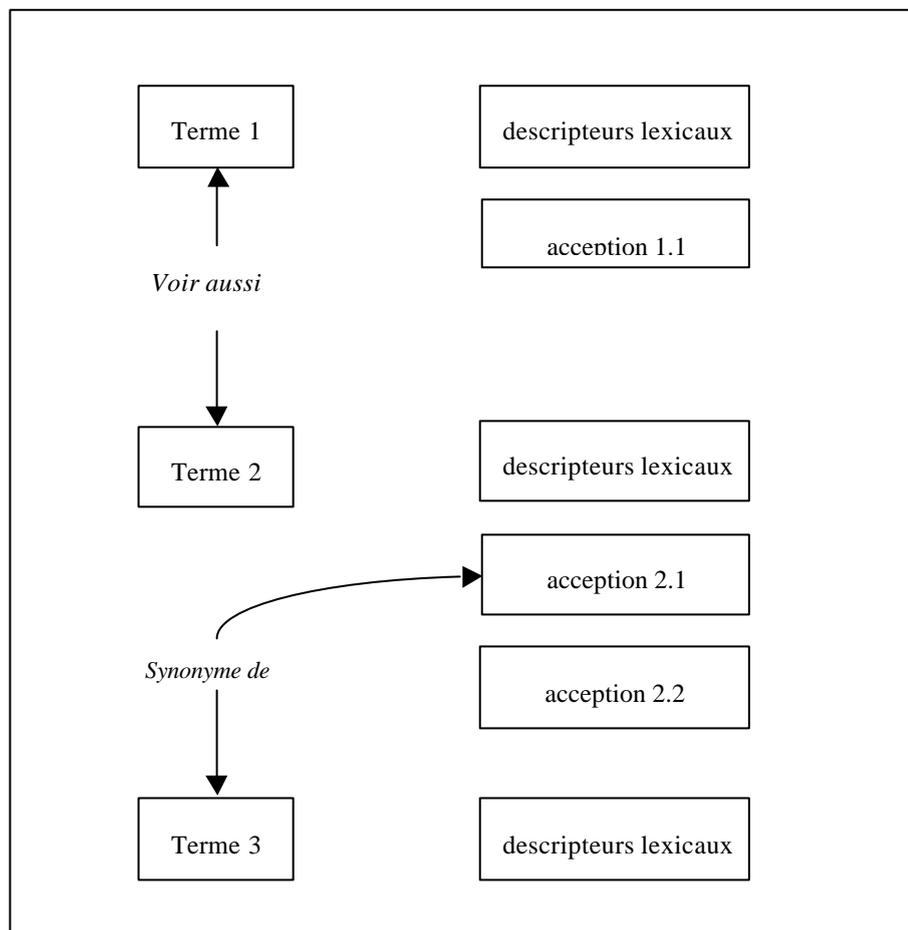


Figure 2 : présentation dite « lexicographique »

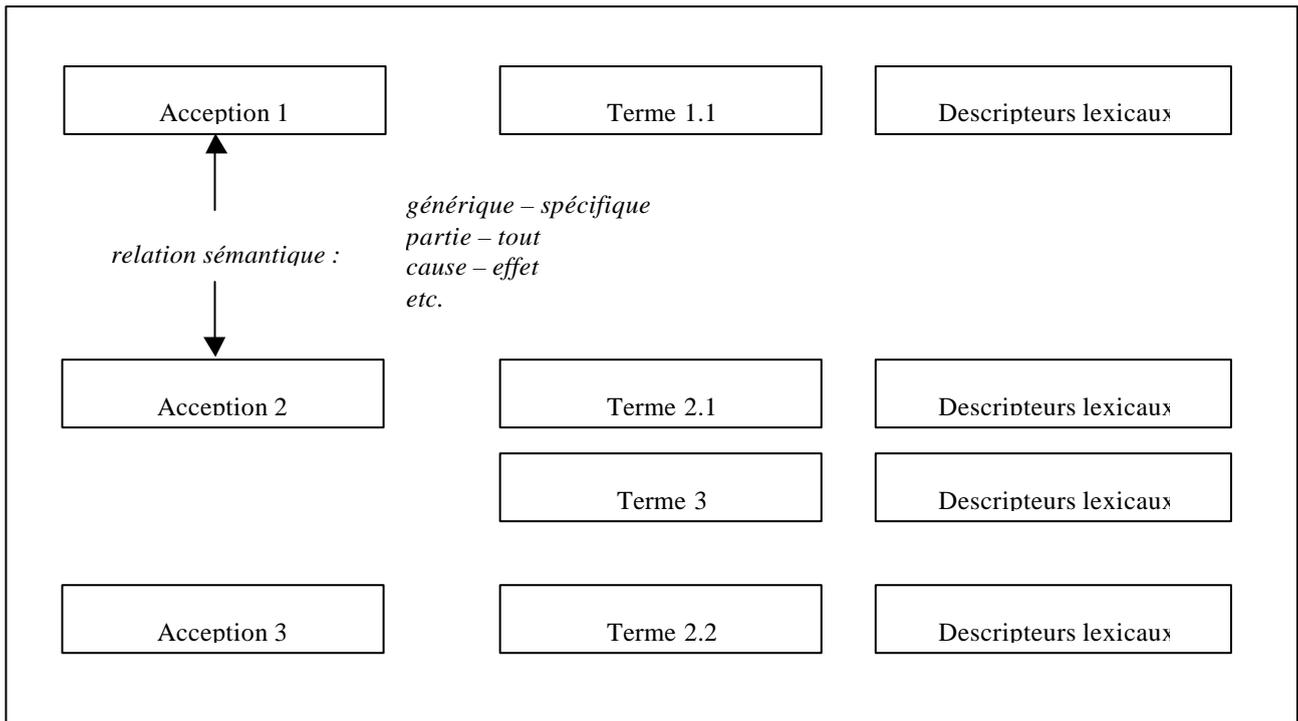


Figure 3 : passage à une présentation dite « terminographique »

4.2 De la théorie à la pratique

Grâce aux grandes compétences du LORIA, partenaire du projet en charge des développements informatiques, le passage d'une présentation à l'autre a été géré entièrement par un mécanisme de transformation au départ d'un même fichier X.M.L. Les extraits qui suivent montrent tour à tour une présentation monolingue et polysémique et une présentation trilingue et monosémique générée à partir du même fichier.

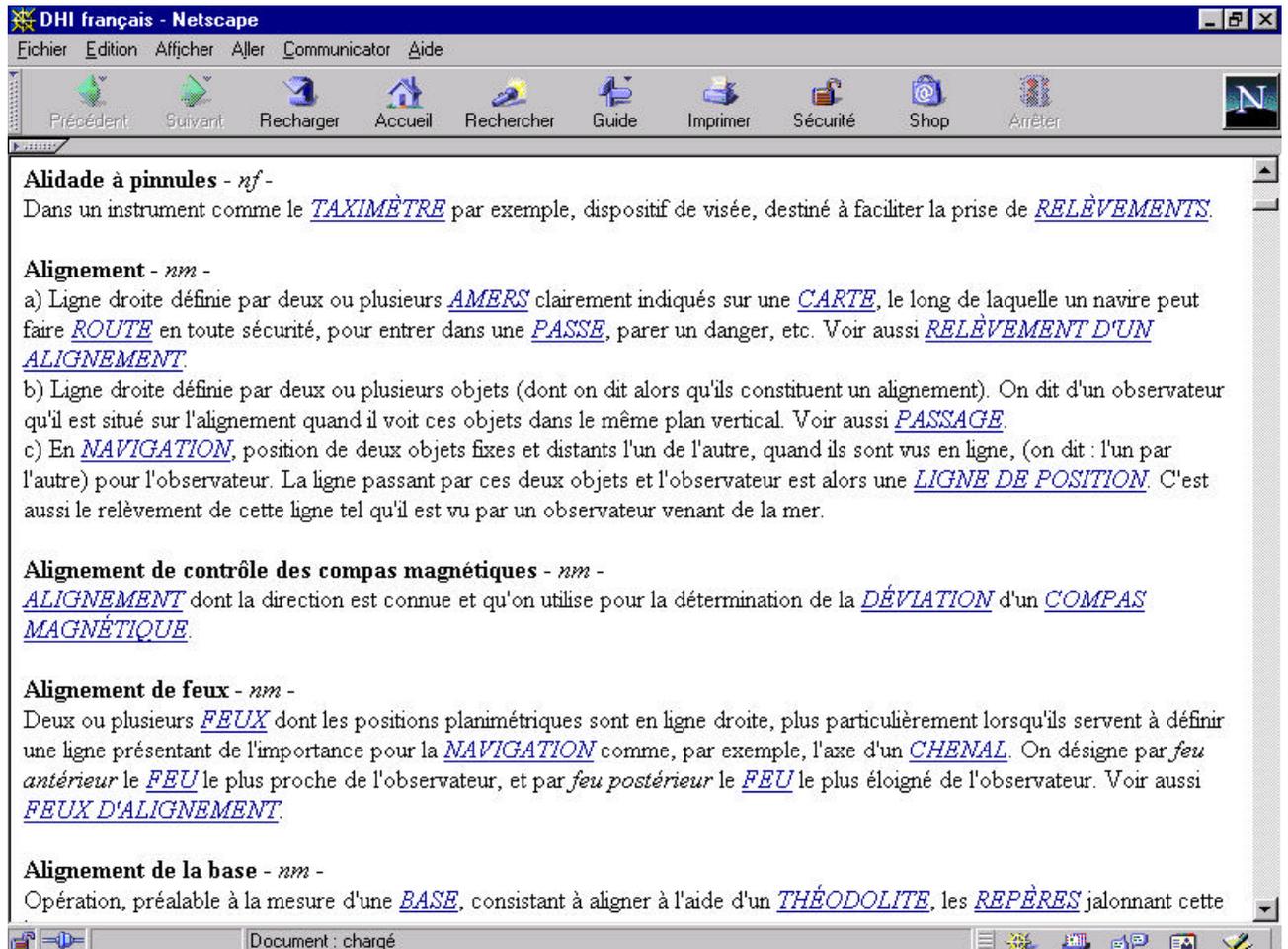


Figure 4 : *présentation monolingue et polysémique*



Figure 5 : présentation trilingue, monosémique et homonymique

5 Une expérience qui n'est pas sans conséquences

La démonstration du passage d'une présentation à l'autre au départ d'un même balisage X.M.L. permet de valider l'hypothèse de départ, décrite en 4.1. Ce constat nous amène à formuler différentes observations :

5.1 Les limites méthodologiques du dictionnaire imprimé

- Tout dictionnaire - général ou spécialisé - peut adopter une perspective polysémique ou monosémique (homonymique).
- Un dictionnaire bilingue imprimé ne peut être polysémique qu'à la condition de contenir deux parties :
 - langue A \Leftrightarrow langue B
 - langue B \Leftrightarrow langue A
- Un dictionnaire imprimé contenant plus de 3 langues devient ingérable si l'on veut séparer les couples de langues.

Partant, nous sommes aujourd'hui tenté d'affirmer que dans un dictionnaire, le choix de la présentation dite *terminographique* s'impose naturellement lorsqu'on décrit plus de deux langues. Un tel choix ne semble pas nécessairement lié à la description des vocabulaires spécialisés ou à l'adoption d'une démarche conceptuelle onomasiologique. S'il s'observe surtout dans les dictionnaires de traduction spécialisée, c'est parce que ceux-ci n'hésitent pas à prendre en compte un grand nombre de langues, au contraire des dictionnaires de la langue générale.

5.2 Le dictionnaire électronique s'affranchit des impératifs du papier

Lorsqu'on conçoit un dictionnaire électronique comme une base de données, la représentation monosémique précède la représentation polysémique dans la mesure où seule une granularité maximale de l'information permet d'ensuite varier les modes de représentation. Il est patent que du point de vue du balisage de l'information, la monosémie précède la polysémie, qui consiste en un assemblage d'informations minimales (regroupement des acceptions et établissement de liens entre celles-ci). On observe d'ailleurs que la monosémie est *de facto* le mode privilégié de représentation du sens dans les dictionnaires électroniques développés par l'ingénierie linguistique, laquelle accorde une grande attention au lien entre sémantique et syntaxe.

Les conseils minimaux que l'on pourrait donner à toute personne désireuse de concevoir un dictionnaire ouvert à la modernité sont :

- chaque type d'information doit être identifié (balisé) avec précision ;
- la base de données doit pouvoir servir à générer des présentations variées (une diversité de produits dérivés) ;
- la base de données doit tenir compte des exigences d'autres applications au sein des industries de la langue, particulièrement les applications qui tiennent compte de l'axe syntagmatique.

5.3 Le concept n'est pas nécessaire à la monosémie

La réussite évidente de l'expérience du passage de la monosémie à la polysémie au départ d'un fichier unique nous permet de défendre l'idée que dans le cadre d'un dictionnaire électronique bien conçu, la différence entre lexicographie spécialisée et terminographie relève de l'ordonnancement des informations et non de la méthodologie de constitution des données, qui peut-être commune (Van Campenhoudt à paraître).

La prétention de la terminologie wüstérienne (École de Vienne, ISO) à être une discipline autonome semble battue en brèche par les constats précédents : la terminologie apparaît comme une lexicologie de la langue spécialisée, la terminographie, un mode particulier de présentation des dictionnaires.

Le concept de Wüster appartient à l'héritage aristotélicien. Il véhicule l'idée d'une langue-nomenclature évacuée par une majorité de linguistes depuis Saussure et qui n'est en rien nécessaire à la rédaction des dictionnaires spécialisés. La description du sens des termes et l'établissement de l'équivalence peuvent très bien s'effectuer sans référence au concept. Les acquis des différentes écoles de la sémantique (réfèrent, signification, sciences cognitives) suffisent amplement à établir les frontières sémantiques du terme.

La terminologie a assurément d'autres originalités méthodologiques majeures à mettre en avant. Elle considère de longue date les dictionnaires comme des bases de données et a insisté très tôt sur l'intérêt de méthodes de classement non alphabétiques fondées sur les relations sémantiques. Par ailleurs, elle a toujours veillé à ne pas limiter le signe aux frontières du mot et à octroyer des entrées particulières aux syntagmes dès lors qu'ils sont figés et véhiculent un signifié propre. Il s'agit là d'autant de pas fondamentaux que la tradition lexicographique répugne à franchir définitivement, si l'on excepte quelques tentatives novatrices, hélas privées de succès.

Bibliographie

Bessero (G.), Bonhomme (P.) et Romary (L.), 2000 : « L'informatisation du *Dictionnaire hydrographique international* : normalisation et utilisation », dans Newman (D.L.) et Van Campenhoudt (M.), éd., 1999 : *Terminologie maritime : traduire et communiquer. Actes du 1^{er} colloque international de terminologie maritime (Bruxelles, 15 et 16 mai 1998)*, Bruxelles, Éditions du Hazard, p. 178-198.

Blampain (D.), Descotte (S.), Husson (J.-L.), Rohde (H.), Romary (L.), Van Campenhoudt (M.) et Viscogliosi (N.), 2000 : « Le projet européen DHYDRO : la normalisation à l'épreuve d'un forum terminologique », dans *Actes de la Conférence sur la coopération dans le domaine de la terminologie en Europe (Paris, 17-19 mai 1999)*, Association européenne de terminologie.

Descotte (S.), Husson (J.-L.), Romary (L.), Van Campenhoudt (M.), Viscogliosi (N.), 2001a : « Specialized lexicography by means of a conceptual data base: establishing the format for a multilingual marine dictionary », dans Vainio (J.), éd., *Maritime Terminology : Dictionaries and Education, Proceedings of the Second Conference on Maritime Terminology, 11-12 May 2000, Turku*, Finland, Turku : University of Turku, p. 63-81 (Publications from the Centre for Maritime Studies, University of Turku, A36).

Descotte (S.), Husson (J.-L.), Romary (L.), Van Campenhoudt (M.), Viscogliosi (N.), 2001b : « Dhydro: a generic environment developed to edit and access multilingual terminological data on the Internet », dans Vainio (J.), éd., *Maritime Terminology : Dictionaries and Education, Proceedings of the Second Conference on Maritime Terminology, 11-12 May 2000, Turku*, Finland, Turku : University of Turku, p. 47-61 (Publications from the Centre for Maritime Studies, University of Turku, A36).

Ide (N.) et VÉRONIS (J.), 1996 : « Codage TEI des dictionnaires électroniques », dans *Cahier GUTenberg, Numéro spécial : TEI - Text Encoding Initiative*, juin 1996, n° 24, p. 170-176.

International Hydrographic Organization, 1994 : *Hydrographic Dictionary*, 5th ed. Monaco, Bureau hydrographique international.

ISO 12 200 (1999) : *Applications informatiques en terminologie - Format d'échange de données terminologiques exploitables par la machine (MARTIF) - Transfert négocié*, Genève, Organisation internationale de normalisation (ISO/TC 37).

Organisation hydrographique internationale, 1998 : *Dictionnaire hydrographique*, 5^e édit., Monaco, Bureau hydrographique international.

Organización Hidrográfica Internacional, 1996 : *Diccionario Hidrográfico*, 5^e édit., Monaco, Bureau hydrographique international.

Sperberg-McQueen (C.M.) et Burnard (L.), éd., 1999 : *Guidelines for Electronic Text Encoding and Interchange*, Revised edition, Chicago, TEI P3 Text Encoding Initiative. Consultable sur le site <http://www.tei-c.org/>.

Van Campenhoudt (M.), 2000 : « De la lexicographie spécialisée à la terminographie : vers un "métadictionnaire" ? », dans Thoiron (Ph.) et Béjoint (H.), dir., *Le sens en terminologie*, Lyon, Presses universitaires de Lyon (Travaux du C.R.T.T.), p. 127-152.

Van Campenhoudt (M.), à paraître : « Pour une approche sémantique du terme et de ses équivalents », article à paraître dans *l'International Journal of Lexicography*, Oxford University Press.
