

Maarten Janssen
Instituto de Linguística Teórica e Computacional
Lisboa (Portugal)

Marc Van Campenhoudt
Centre de recherche TERMISTI
Institut supérieur de traducteurs et interprètes
Haute École de Bruxelles (Belgique)

TERMINOLOGIE TRADUCTIVE ET REPRÉSENTATION DES CONNAISSANCES : L'USAGE DES RELATIONS HYPONYMIQUES

1. Introduction

Identifier le sens des mots est sans doute une tâche aussi ardue en langue spécialisée qu'en langue générale. Ainsi, n'est-il pas évident de déterminer ce qui distingue précisément des termes d'hydrographie comme *bras, canal, chenal, détroit, entrée, goulet, ouverture, passage, passe* ou *raz*¹. La nuance peut être à ce point ténue qu'il est difficile de se prononcer sur leur éventuelle synonymie. Mais il peut être encore plus compliqué de déterminer des équivalences entre langues : *arm of the sea, reach, canal, channel, entrance, fairway, gat, gut, inlet, lane, lead, mouth, narrows, pass, opening, passage, race, reach, slough, sound, strait, swash* constituent autant de traductions envisageables en anglais. Pour désigner les cas de non-isomorphisme entre les langues (Lyons 1970 : 45) où il n'est pas possible d'identifier un candidat équivalent, on parlera ici de « trou lexical ».

Les trous lexicaux sont liés aux divergences sémantiques qui séparent les termes des langues envisagées. Leur existence constitue une véritable problématique de recherche en terminologie multilingue, car la représentation de telles divergences est une tâche cruciale pour la traduction. À ce jour, on n'a guère creusé la voie, déjà explorée par les projets CODE (Meyer *et al.* 1997 : 108-117) et OMBI (Martin *et al.* 1998), d'une modélisation de ces divergences sémantiques sous la forme de traits définitionnels. Il s'agit pourtant d'un terrain familier pour les lexicographes et terminographes, celui des *differentiae specificae* (différences spécifiques), connues depuis Aristote et qui jouent un rôle essentiel dans les dictionnaires. La définition y utilise communément le modèle aristotélicien du genre prochain : elle désigne l'hyperonyme du référent, puis énumère les différences spécifiques qui distinguent le référent de ses cohyponymes.

Dans un système de base de données plus élaboré, les différences spécifiques peuvent être conçues de deux manières différentes : une approche taxinomique où les traits définitionnels étiquettent les nœuds d'une arborescence et une approche où les traits constituent eux-mêmes la structure taxinomique, qui prend alors la forme d'un treillis. Cet article compare ces deux approches en les illustrant chacune par un modèle concret : celui du Centre de recherche TERMISTI (Van Campenhoudt 2001) et celui du projet SIMuLLDA, élaboré à l'Université d'Utrecht (Janssen 2002),

1. Les exemples relevant de l'hydrographie sont empruntés au *Dictionnaire hydrographique* de l'Organisation hydrographique internationale (O.H.I. 1994-1998). La plupart ont déjà été présentés dans *l'International Journal of Lexicography* (Van Campenhoudt 2001) et ont servi de corpus de travail aux auteurs de la présente communication.

2. Le principe d'équivalence

Généralement polysémiques, les dictionnaires de traduction de la langue générale énumèrent les acceptions des lexèmes retenus en entrée. Se revendiquant souvent comme « conceptuelle », l'approche classique en terminographie multilingue est plutôt monosémique et aboutit donc à créer des entrées homonymes en cas de polysémie avérée. Un dépouillement superficiel de quelques terminographies de qualité suffirait à montrer que les termes spécialisés tendent à être plus spécifiques et nettement moins polysémiques que ceux de la langue générale.

Une analyse plus minutieuse conduit toutefois à reconnaître qu'une monosémie stricte ne peut être systématiquement identifiée. Les mécanismes de diversification du sens affectent aussi les langues spécialisées, surtout lorsqu'elles sont marquées par la diachronie. Au sein d'un même domaine, des communautés distinctes utilisent parfois les mêmes termes d'une manière quelque peu différente. Par ailleurs, l'interdisciplinarité conduit, elle aussi, à une diversification des sens. Ces constats plaident en faveur d'une intégration du fait polysémique en terminologie, une discipline qui mérite d'être perçue comme une lexicologie appliquée optant pour une approche homonymique fondée sur des critères rigoureux lorsqu'elle œuvre dans un cadre monolingue (Van Campenhoudt 2001 : 186-188).

Aborder le multilinguisme et le problème de l'équivalence des termes conduit à ressentir davantage le besoin d'une prise en compte du fait polysémique. Les dictionnaires de traduction spécialisée qui se contentent d'établir un lien entre un terme source et un terme cible sont souvent le fait de terminographes « aux pieds nus » qui n'ont guère idée des problèmes que pose la terminologie traductive. Marqués par une grande indigence de l'information sémantique, ces « lexiques » sont fondés sur l'illusion d'une équivalence parfaite, comme si le fait que *A* est toujours égal à *B* lorsqu'on traduit de la langue 1 (L_1) vers la langue 2 (L_2) implique que *B* soit toujours égal à *A* en cas de changement du sens de la traduction.

Lorsque le dictionnaire bilingue classique prévoit une inversion des langues source et cible, il se compose *de facto* de deux dictionnaires : de L_1 vers L_2 et de L_2 vers L_1 . Chaque volume envisage chacune des manières de traduire l'entrée en langue cible et la discrimination des sens est précisément fondée sur ce questionnement. Ainsi, le nombre des acceptions envisagées pour le mot français *banane*, varie-t-il logiquement entre le *Nouveau Petit Robert* et le *Robert et Collins* (Van Campenhoudt 2000). Par contre, dans les terminographies, et plus encore dans les bases de données terminographiques, l'inversion des langues est censée se réaliser sous une entrée unique, ce qui pose un problème méthodologique évident, qui est d'arriver à spécifier les conditions sémantiques de l'équivalence. Ce choix de ne travailler qu'avec un seul « volume » (ou ensemble de fiches) multilingue permettant l'interversion des langues source et cible implique nécessairement d'adopter une démarche homonymique.

Le principe d'équivalence énonce que « Si *A* de L_1 (langue 1) est équivalent à α de L_2 (langue 2) et si *A* de la L_1 est équivalent à β de L_2 alors que α de L_2 n'est pas synonyme de β 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. » (Van Campenhoudt 1994 : 66.)

Ainsi, en hydrographie, le terme *entrée* pourrait aisément faire l'objet d'une définition unique, mais il reçoit deux acceptions dans le *Dictionnaire hydrographique* de l'O.H.I. (1994-1998) dès lors que le choix des équivalents anglais et espagnol varie en fonction du sens :

Entrée. Passage par lequel on pénètre dans un PORT, un DOCK, un CHENAL.

Entrance. The SEAWARD end of a CHANNEL, HARBOUR, DOCK, etc.

Entrada, acceso. SECTOR HACIA EL MAR de la finalización de un CANAL, PUERTO, DÁRSENA, etc.

Entrée. Ouverture faisant communiquer une BAIE avec la MER.

Mouth. The place of discharge of a STREAM into the OCEAN or ENTRANCE to a BAY from the OCEAN.

Desembocadura, embocadura, boca. Lugar de descarga de un CURSO DE AGUA en el OCEANO o ENTRADA a una BAHIA desde el OCEANO.

Le principe d'équivalence est la simple transposition dans un cadre multilingue du principe d'exploitation de la relation de synonymie utilisé pour délimiter les nœuds au sein d'un réseau sémantique monolingue : « si A est synonyme de B et si A est synonyme de C alors que B n'est pas synonyme de C, c'est que probablement A possède deux sens qui devraient être différenciés par deux nœuds du réseau. » (Levrat et Sabah 1990 : 93.) En d'autres termes, si *entrée* équivaut à *entrance* et si *entrée* équivaut à *mouth*, mais que *entrance* n'équivaut pas à *mouth*, il convient de créer deux vedettes *entrées* au sein du dictionnaire.

L'application de ce principe conduit à créer autant d'entrées que nécessaire dans le dictionnaire multilingue en un volume – ou dans chaque fiche de la base de données – pour garantir une validité de l'équivalence quel que soit le couple de langues retenu. Son usage systématique dans le cadre du projet européen DHYDRO a montré sa robustesse s'agissant de concevoir une base de données terminologique multilingue au départ des versions française, anglaise et espagnole du *Dictionnaire hydrographique* (Descotte et al. 2001).

Aide pragmatique à la création de dictionnaires multilingues, le principe d'équivalence aide le terminographe à maîtriser le nombre d'entrées homonymes à envisager, mais ne permet aucune prédiction sur les liens sémantiques qui unissent les différentes acceptions isolées. Il résout des cas relativement complexes d'équivalences multiples, mais ne peut rendre compte de tous les cas de non-isomorphisme comme nous le verrons plus loin.

3. Approche arborescente et componentielle de l'équivalence

Le lien entre le principe d'équivalence et la relation d'hyponymie a été établi de longue date (Van Campenhoudt 1994 : 65-116 et 1996). L'approche arborescente de l'équivalence a pour ambition de formaliser la découverte des trous lexicaux à travers l'exploitation de l'hyponymie et de traits sémantiques ou sèmes². Elle permet de proposer dans le même temps une application « automatique » du principe d'équivalence.

On considère à la base que dans les vocabulaires spécialisés, tout sème peut obtenir trois valeurs différentes : [+] lorsque le trait se réalise toujours, [-] lorsque le trait ne se réalise jamais et [0] lorsque le trait peut ou non se réaliser sans changer la nature du référent. Ce sème [0] caractérise l'archisémème de l'hyperonyme par rapport aux sémèmes des hyponymes, qui réalisent un tel trait sous une valeur [+] ou [-]. Le modèle nomme *sème potentiel* ce trait sémantique de l'hyperonyme qui se réalise nécessairement dans le cadre de la relation d'hyponymie. Il est assez classique que les termes cohyponymes lexicalisent ce trait distinctif pour mieux se différencier de l'archilexème (tableau 1).

2. On trouvera une présentation beaucoup plus détaillée chez Van Campenhoudt (2001).

	déplacement d'eau	périodique	dans le sens du vent	dans le sens inverse du vent
Courant	+	0	0	0
Courant de marée	+	+	0	0
Courant de marée portant au vent	+	+	-	+
Courant de marée portant sous le vent	+	+	+	-

Tableau 1 : typologie des courants

La méthode terminographique propose, de longue date, de répertorier les termes d'un domaine et d'établir dans chaque langue un réseau sémantique des liens qui unissent les termes non-synonymes. L'idée fondamentale est que la confrontation de chaque réseau monolingue permet de mieux dépister les trous lexicaux. L'apport du modèle de TERMISTI consiste à confronter les sémèmes de termes de langues différentes qui sont des candidats équivalents. Seront ainsi réputés équivalents les termes qui possèdent les mêmes sémèmes. *A priori*, la démarche ne semble pas orthodoxe puisqu'elle conduit à mélanger des traits sémantiques propres à différentes langues. Pourtant, elle s'avère logique dès lors que l'ambition est précisément de rapprocher les termes de ces langues et de résoudre les problèmes de non-isomorphisme au travers d'un réseau sémantique multilingue.

La confrontation de sémèmes relevant de langues distinctes multiplie inévitablement le nombre de traits pris en compte et, partant, le nombre de sèmes potentiels - obtenant une valeur [0] - dès lors qu'ils ne sont pas pertinents dans une langue particulière. Ainsi est-il parfaitement logique qu'*entrée* obtienne la valeur [0] pour les sèmes 'entre baie et mer' et 'vers infrastructure portuaire' qui jouent un rôle différenciateur en anglais comme en espagnol (tableau 2).

	passage maritime	entre baie et mer	vers infrastructure portuaire
<i>entrée</i>	+	0	0
<i>entrance</i>	+	-	+
<i>entrada, acceso</i>	+	-	+
<i>mouth</i>	+	+	-
<i>desembocadura, embocadura, boca</i>	+	+	-

Tableau 2 : comparaison des sémèmes d'*entrée* et de ses équivalents anglais et espagnols

Comme on l'a vu, les sémèmes qui contiennent un sème potentiel peuvent être classés de manière très aisée : un sémème *A* est un hyponyme d'un sémème *B* si *A* spécifie au moins un sème potentiel de *B*. Ceci peut être aisément résumé à travers la règle suivante :

hyponyme = hyperonyme + actualisation d'un sème potentiel

Selon cette règle, *entrance* et *mouth* apparaissent comme des hyponymes de *entrée* (figure 1) - à défaut d'en être les équivalents parfaits - dès lors qu'ils actualisent ses sèmes potentiels 'entre baie et mer' et 'vers infrastructure portuaire'.

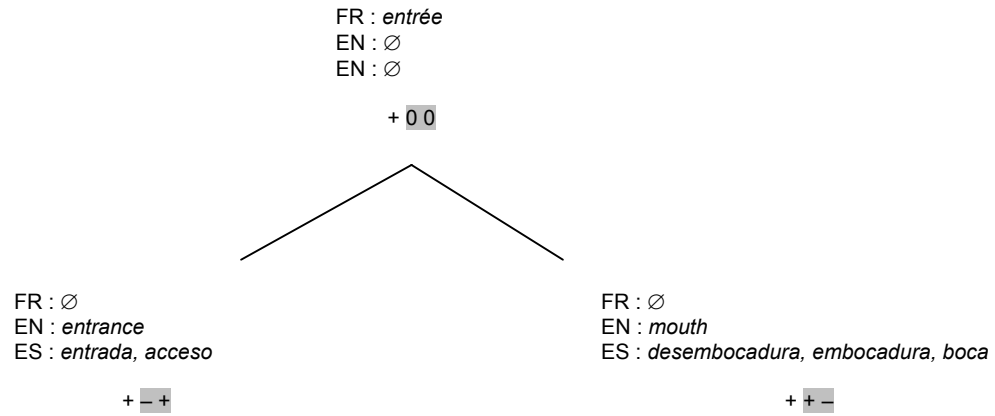


Figure 1 : le réseau sémantique multilingue

Les dictionnaires imprimés présentent ces termes comme équivalents en s'appuyant sur un procédé anaphorique courant : le recours à l'hyperonyme pour corréler à l'hyponyme (p.ex., *souris* est remplacé par *rongeur* ou *courant de marée* par *courant*). Ce mécanisme, que le modèle appelle *hyperonomase*, est précisément celui qui permet de traduire *entrée* tantôt par *entrance*, tantôt par *mouth*. Monosémique en français, *entrée* devient polysémique dans le cadre d'une traduction vers l'anglais et chacun des deux sémèmes nécessaires à l'établissement de l'équivalence se réalise au niveau hyponymique dans le réseau sémantique multilingue. Ceux-ci actualisent sous la forme de valeurs inverses les sèmes potentiels du terme français appréhendé dans un cadre monolingue.

	passage maritime	entre baie et mer	vers infrastructure portuaire
<i>entrée</i>	+	-	+
<i>entrance</i>	+	-	+
<i>entrada, acceso</i>	+	-	+
<i>entrée</i>	+	+	-
<i>mouth</i>	+	+	-
<i>desembocadura, embocadura, boca</i>	+	+	-

Tableau 3 : les sémèmes d'*entrée* et de ses équivalents dans le dictionnaire de traduction

Dans un premier temps, l'approche arborescente propose d'expliquer le mécanisme de l'équivalence partielle par une modification logique du réseau sémantique qu'il nomme *phagocytose* : l'hyperonyme *entrée*, entièrement désigné par ses hyponymes, n'est plus nécessaire et est en quelque sorte « phagocyté » par ceux-ci. La figure 2 tente d'illustrer ce processus : le nœud supérieur est supprimé dès lors que toute l'extension du terme *entrée* est couverte par les sémèmes des deux homonymes *entrée* transférés au niveau inférieur comme équivalents de *mouth* et de *entrance*.

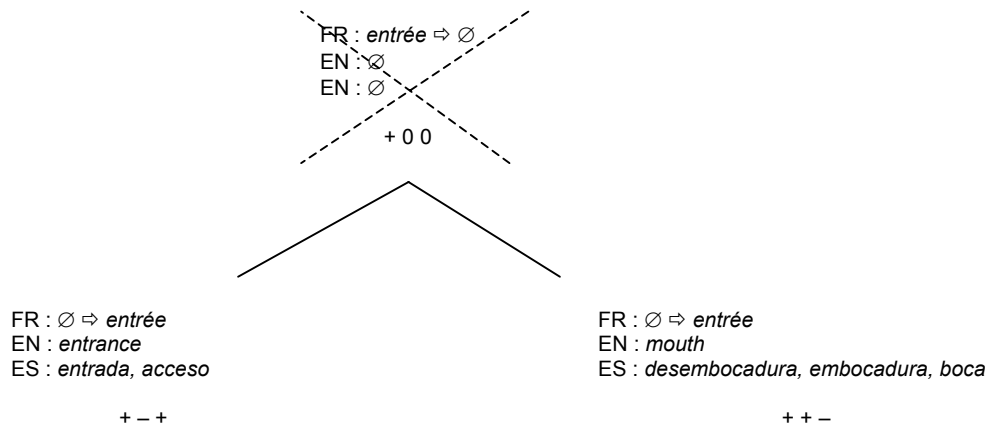


Figure 2 : le réseau sémantique multilingue après hypernomase et phagocytose

Dans un second temps, ce modèle – qui n’a jamais fait l’objet d’une expérimentation informatique – devrait permettre de concevoir un outil de recherche des équivalences plus respectueux de chaque univers linguistique. La phagocytose est clairement un mécanisme nécessaire à la rédaction du dictionnaire de traduction. Elle préside à la distinction des acceptions dans le dictionnaire bilingue en deux volumes, fondé sur la polysémie, et à la distinction des entrées homonymes dans le dictionnaire multilingue en un volume ou dans les bases de données terminographiques multilingues. Elle n’est, par contre, nullement indispensable à un système informatique qui exploiterait un réseau sémantique multilingue pour établir les équivalences. En effet, la simple identification d’hyponymes actualisant les sèmes potentiels dans une langue étrangère devrait permettre de les proposer comme équivalents partiels de l’hyperonyme.

Un réseau sémantique multilingue permet de résoudre des cas beaucoup plus complexes comme nous le verrons au point 5. Ainsi, l’absence d’isomorphisme peut affecter bien plus qu’un terme isolé et concerner de véritables hiérarchies au sein d’un champ particulier. Un exemple extrême de ce type de cas a déjà été discuté par Sowa (1993), qui a montré les profondes divergences qui s’observent entre la typologie des véhicules en anglais et en chinois. Nous en trouverons d’autres dans le domaine de l’hydrographie.

4. Gérer une approche componentielle de l’équivalence à travers un treillis

Dans un arbre, chaque nœud est relié à un superordonné (le genre). Dans les treillis, un simple nœud peut posséder plusieurs superordonnés. Dans les treillis à intersections, pertinents pour la terminologie, les nœuds dictent eux-mêmes la structure : chacun représente un ensemble de traits et deux nœuds possédant une intersection de leurs traits auront cette intersection pour concept superordonné. Ainsi, l’ensemble de la structure du treillis simple de la figure 3, résulte de l’intersection des trois nœuds {a,b}, {b,c} et {a,c}³.

3. Pour une brève initiation aux structures qui fondent les ontologies, lire Sowa (2001).

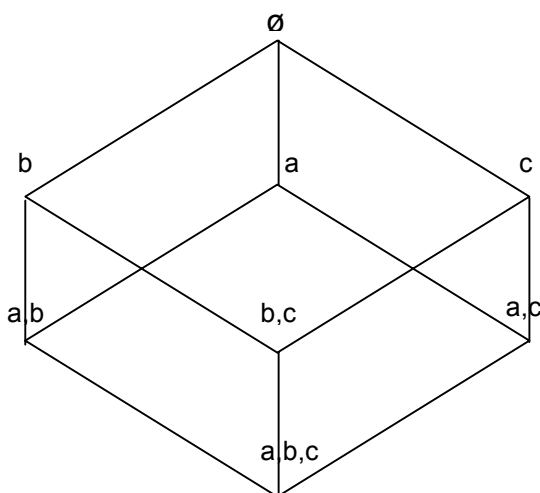


Figure 3 : exemple de treillis à intersections

Le système SIMuLLDA (Janssen 2002), qui servira à illustrer cette approche, a été conçu pour gérer une base de données lexicales multilingues (B.D.L.M.) capable de générer des dictionnaires bilingues. Il a moins été conçu dans une perspective terminologique que dans celle des dictionnaires en langue générale. La B.D.L.M. de SIMuLLDA est pilotée par une structure logique appelée *analyse formelle de concepts* (A.F.C.⁴). Ce modèle algébrique sert précisément à exploiter des treillis.

L'originalité de l'A.F.C. est de produire un treillis qui prend uniquement en compte les combinaisons de traits qui définissent un « concept ». Dans un treillis booléen, par contre, toutes les possibilités combinatoires sont envisagées, y compris la plus absurde, qui apparaît au bas du treillis (*infimum*). Tout ajout d'un nouvel attribut conduit à une nouvelle version mise à jour du treillis (*refinement*).

Comme on le perçoit, cette perspective diffère de celle fondée sur les arbres par son arrière-fond théorique particulier. Toutefois, ces approches présentent également d'importantes ressemblances : toutes deux se fondent sur une décomposition du sens, utilisant des traits sémantiques comparables aux sèmes de la sémantique interprétative. Toutes deux tentent d'établir par ce biais un appariement translinguistique de mots ou de termes réputés équivalents et toutes deux placent la question des trous lexicaux au centre de leurs préoccupations. Leurs méthodes divergent par l'utilisation des traits sémantiques, appelés *attributs définitionnels* dans le cadre du modèle SIMuLLDA.

Dans l'approche arborescente, les sèmes permettent de distinguer des sémèmes qui, au sein d'un champ sémantique clairement délimité par un domaine de spécialité, appartiennent au même archisémème. Leur valeur est précisée par rapport à cet archisémème. L'existence même de cet archisémème est propre aux arbres, puisque ceux-ci comportent des sémèmes ne possédant qu'un seul archisémème. Les sèmes permettent d'y établir la relation d'hyponymie.

L'idée qui sous-tend les attributs définitionnels des treillis est différente. Toutes les acceptions sont définies au sein de la B.D.L.M. à l'aide du même (large) ensemble d'attributs. Chaque acception est différenciée à l'aide d'un sous-ensemble qui est considéré *être* sa définition lexicographique. Ceci

4. Plus souvent désignée à l'aide du sigle *F.C.A.* (de l'allemand *Formalen Concept Analyse*). On trouvera une présentation de ce formalisme logique et de son application à une B.D.L.M. chez Ganter et Wille (1996) et chez Janssen (2002).

ne signifie pas que les attributs définitionnels définissent chaque aspect du sens du mot, mais qu'ensemble, ils en fournissent la même caractérisation que des dictionnaires monolingues. Les attributs définitionnels représentent les *differenciae specifica*e que l'on trouve dans ces ouvrages. Ils ne peuvent recevoir une valeur neutre comme dans le modèle arborescent : un attribut définit ou non une acception. La structure du treillis est établie à l'aide du principe décrit ci-dessus, c.-à-d. en considérant les intersections récursives des ensembles d'attributs définitionnels caractérisant les différentes acceptations.

Il importe de souligner que dans un treillis géré par SIMuLLDA, les attributs définitionnels ne se rapportent pas à des mots ou à des termes, mais à des *accep*tions *interlinguistiques* constituées d'un ensemble d'attributs définitionnels. Chaque mot de chaque langue exprime une ou plusieurs acceptations interlinguistiques. Ainsi, le tableau 4 contient-il une acception interlinguistique à laquelle se rapportent le mot anglais *entrance* et le mot espagnol *entrada*. Chaque acception interlinguistique est représentée à l'aide de petites capitales et est arbitrairement dénommée d'après le mot qui la désigne dans une des langues qui l'expriment. Les attributs définitionnels, représentés en gras, sont, eux aussi, interlinguistiques et peuvent être lexicalisés dans chaque langue. Par exemple, **between bay and sea** peut se dire « *to a bay from the ocean* » dans la définition anglaise.

	coastal passage	between bay and sea	to harbour
ENTRÉE	+		
ENTRANCE	+		+
MOUTH	+	+	

Tableau 4 : les attributs définitionnels des acceptations linguistiques dans SIMuLLDA

Si les acceptations sont définies sous la forme d'un ensemble d'attributs définitionnels, il est très aisé de les hiérarchiser à l'aide des deux règles suivantes :

1. Deux acceptations sont identiques (synonymes) si elles sont définies à l'aide du même ensemble d'attributs définitionnels.
2. Un hyperonyme est défini par un sous-ensemble commun à tous ses hyponymes, les attributs additionnels des hyponymes constituant les *differentiae specifica*e.

Ces deux principes ne sont pas implémentés en tant que tels dans la structure du système, puisqu'ils sont pris en charge par le mécanisme sous-jacent de l'analyse formelle de concepts. Au départ d'une table représentant la distribution des attributs définitionnels, l'A.F.C. génère automatiquement une structure hiérarchique qui envisage toutes les intersections récursives : le treillis⁵. L'ajout d'un nouvel attribut débouche automatiquement sur une mise à jour du treillis (*refinement*). S'appuyant sur ce principe, l'A.F.C. crée de nouvelles acceptations d'une manière assez simple : si deux acceptations partagent un ensemble commun d'attributs définitionnels (correspondant à l'intersection des deux ensembles), ceux-ci seront réputés correspondre à une acception, même s'il n'y a pas de mot ou de terme qui lui corresponde. Par exemple, quand bien même il n'existerait pas de mot pour exprimer l'acception interlinguistique ENTRÉE, celle-ci serait présente dans le treillis

5. On trouvera une bonne illustration de ce processus au travers de l'interface *JaLaBA*, active sur le site internet de SIMuLLDA (<http://janssenweb.net/simullda>). Des exemples concrets des résultats produits par *JaLaBA* seront présentés dans la section 5.2.

puisqu'elle correspond à l'intersection entre MOUTH et ENTRANCE. La figure 4 représente le treillis correspondant au tableau 4 et inclut les termes français, anglais et espagnols exprimant les trois acceptions interlinguistiques dudit tableau.

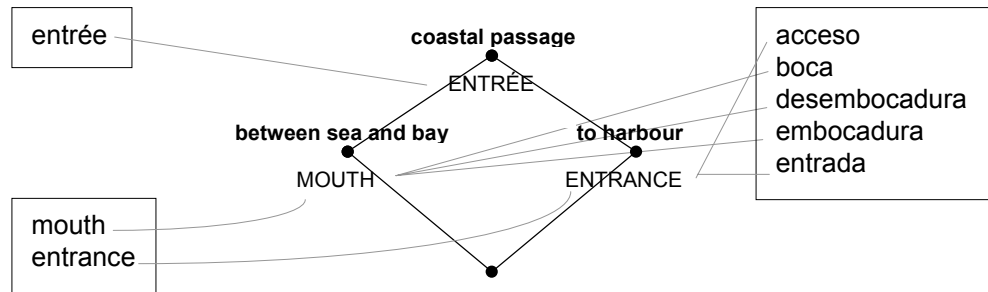


Figure 4 : treillis des passages côtiers représenté dans SIMuLLDA

Visant d'abord à expliquer l'établissement des équivalences dans les dictionnaires, le modèle arborescent proposé par TERMISTI ne respecte le réseau sémantique que dans une première phase de l'analyse. En effet, à travers l'hyponomase et la phagocytose, les dictionnaires font descendre l'hyperonyme *entrée* à un niveau subordonné après en avoir subdivisé les acceptions. Dans l'approche sous forme de treillis, le réseau interlinguistique est directement exploitable en lui-même. Il est conçu pour représenter des relations concrètes entre acceptions. Ainsi, SIMuLLDA ne crée pas de nouvelles acceptions pour *entrée* (hyponomase) dans la mesure où, du point de vue d'un francophone, il n'y a pas de raison d'attribuer deux sens différents à *entrée*. L'acception interlinguistique ENTRÉE demeure un hyperonyme de MOUTH et de ENTRANCE et elle ne sera pas exprimée en anglais. Toutefois, le treillis contient deux hyponymes pouvant servir d'équivalents traductifs (au sens du *translational equivalent* de Zgusta [1971 : 320]) à choisir en fonction du contexte : *mouth* et *entrance*.

Le treillis permet également de dériver un « équivalent explicatif » (*explanatory equivalent*) au sens où l'entend Zgusta (1971 : 320). En l'absence de terme français pour exprimer MOUTH (c.-à-d. en cas de trou lexical), le correspondant français de l'hyperonyme ENTRÉE ne peut servir d'équivalent car MOUTH possède davantage d'attributs définitionnels : MOUTH = ENTRÉE + **between sea and bay**. Une traduction adéquate en français de *mouth* suppose de compléter la lexicalisation de ENTRÉE (qui est *entrée*) par celle de **between sea and bay** (qui est *d'une baie*) pour produire un équivalent qui serait *entrée d'une baie*. Ceci explique pourquoi le principe d'équivalence n'est pas activé dans un treillis : le terme français *entrée* n'y recevra qu'une seule acception, même s'il peut se traduire de deux manières différentes.

5. Cas concrets

À la différence du principe d'équivalence, destiné à aider le terminographe, les modèles théoriques de l'arbre et du treillis exploitent un raisonnement pour résoudre le problème posé par l'établissement de l'équivalence translinguistique. Des terminographies multilingues de qualité - tel le *Dictionnaire hydrographique* - offrent des exemples concrets de difficultés de traduction qui peuvent servir de banc d'essai à ces deux modèles. Dans la suite de cet article, nous nous servons de cette source pour comparer la manière dont les arbres et les treillis résolvent les problèmes d'équivalences

liés à trois variétés de trous lexicaux : l'inclusion hyponymique simple, l'inclusion hyponymique sur plusieurs niveaux et l'intersection partielle. Les résultats de la modélisation arborescente de TERMISTI (Van Campenhoudt 1996 et 2001) seront comparés avec l'analyse sous forme de treillis appliquée par SIMuLLDA aux mêmes cas.

5.1. L'inclusion hyponymique simple

Le cas de *entrée*, *entrance* et *mouth* est de loin le plus simple et le plus fréquent : un mot de L_1 est l'hyponymie de ses équivalents potentiels en L_2 . Comme on l'a vu, les deux modélisations ici envisagées permettent de résoudre ce type de trou lexical. Deux points méritent toutefois d'être soulignés à ce propos.

Le premier est que dans un semblable cas, on observe un vide dans chaque langue cible. Plus simple est la situation du trou observé dans une seule langue. Par exemple, dans le *Dictionnaire hydrographique*, la définition de *ooze* permet de percevoir ce terme comme hyponyme de *mud*. En français, *boue* est présenté comme équivalent de *mud* mais sert aussi pour combler un trou lexical observé au niveau hyponymique et traduire *ooze*. Il n'est pas rare d'observer de tels trous pour des notions spécifiques à une culture, y compris dans un domaine spécialisé (songeons aux abers et aux calanques).

Boue. SÉDIMENT des grandes profondeurs, à grains fins, contenant dans une proportion de 30% ou plus du SABLE ou des débris calcaires ou siliceux, provenant de petits organismes marins, dont les dimensions granulométriques sont du même ordre que celles du LIMON, le reste étant constitué de matériaux amorphes dont les dimensions granulométriques sont du même ordre que celles de l'ARGILE.

Ooze. A fine-grained pelagic SEDIMENT containing undissolved SAND- or silt-sized, calcareous or siliceous skeletal remains of small marine organisms in proportion of 30 percent or more, the remainder being amorphous clay-sized material.

Le deuxième point qui doit retenir l'attention est que les deux approches traitent le trou lexical différemment. Exploitant un arbre, TERMISTI applique le principe de l'hyponomase à *boue* de manière à lui donner une acception plus spécifique, celle de *ooze*. Mais, à la différence du cas de *mouth/entrance* par rapport à *entrée*, *boue* ne subit pas de phagocytose après l'hyponomase puisque *boue* est toujours nécessaire comme équivalent du générique *mud*, en sorte que la phagocytose ne peut avoir lieu. *Boue* reçoit donc deux équivalents traductifs : *mud* et *ooze* et, à l'inverse, *mud* et *ooze* ne possèdent qu'un seul équivalent traductif : *boue*.

La situation ainsi créée peut paraître étrange : *boue* reçoit deux acceptions, dont l'une est hyponyme de l'autre. Bien entendu, si l'on adopte une perspective traductive, cette situation n'est en rien erronée puisque tant *ooze* que *mud* doivent être rendus par *boue*. Il est aussi vrai que tout *ooze* est une boue, ce qui fonde l'hyponomase⁶. En revanche, on se rappellera que dans un système comme SIMuLLDA, le tri des concepts à l'aide d'un treillis interlinguistique est une fin en soi. Ce système respectera donc davantage la perspective originelle de la langue française, qui ne semble pas considérer qu'une boue contenant au moins 30% de sable ou des débris calcaires ou siliceux mérite une dénomination particulière.

6. On observera que ce type de situation n'est pas inconnu dans les dictionnaires de la langue générale (voir, par exemple, les abréviations « spécialt » du *Nouveau Petit Robert*). Il reste que l'on est plus habitué à voir le phénomène unir un mot de la langue générale et un terme de la langue spécialisée.

5.2. L'inclusion hyponymique sur plusieurs niveaux

Comme nous l'avons déjà précisé, il est des cas où les langues organisent d'une manière fort différente les relations d'hyponymie au sein d'un même champ sémantique. Un bon exemple d'une telle situation de non-isomorphisme est fourni par la typologie des vagues déferlantes telle qu'elle ressort des définitions anglaises et française du *Dictionnaire hydrographique*. Le tableau 5 en présente l'analyse sémique selon la modélisation arborescente de TERMISTI (la représentation de SIMuLLDA s'obtient en retirant les [-] et les [0]). Là où le français n'utilise que deux termes (*vague déferlante* et *brisant*), l'anglais en possède cinq (*comber*, *breaker*, *spilling breaker*, *plunging breaker* et *surging breaker*).

	vague	déferlement	eau profonde	côte, rivage	haut-fond ou récif	déferlement glissant	déferlement plongeant	déferlement frontal
<i>vague déferlante</i>	+	+	0	0	0	0	0	0
<i>brisant</i>	+	+	-	+	+	0	0	0
<i>comber</i>	+	+	+	-	0	0	0	0
<i>breaker</i>	+	+	-	+	0	0	0	0
<i>spilling breaker</i>	+	+	-	+	0	+	-	-
<i>plunging breaker</i>	+	+	-	+	0	-	+	-
<i>surging breaker</i>	+	+	-	+	0	-	-	+

Tableau 5 : les traits sémantiques des désignations des vagues déferlantes

Le seul principe d'équivalence ne permet pas d'isoler les six équivalences proposées par le *Dictionnaire hydrographique*, à savoir : *vague déferlante* = *comber*, *vague déferlante* = *breaker*, *vague déferlante* = *spilling breaker*, *vague déferlante* = *plunging breaker*, *vague déferlante* = *surging breaker*, *brisant* = *breaker*

Le réseau sémantique dégagé selon l'approche arborescente de TERMISTI (figure 5) permet de mieux comprendre les équivalences retenues par le dictionnaire. La solution envisage une hyperonomase sur plusieurs niveaux et l'existence de nœuds sémantiques vides. En effet, les trois équivalences proposées au niveau inférieur, après hyperonomase, héritent des traits d'un nœud intermédiaire qui n'est pas pris en compte dans le dictionnaire multilingue (il est « phagocyté ») dès lors que l'ensemble de ses hyponymes couvrent toute son extension et qu'il n'est désigné dans aucune des langues considérées. De fait, ni l'anglais ni le français ne désignent une vague [+] qui déferlerait [+] non pas en eau profonde [-] mais sur une côte ou un rivage [+], sans affecter un haut-fond ou un récif [-] et quel que soit son mode de déferlement [0] [0] [0].

L'analyse en treillis réalisée par SIMuLLDA offre un résultat comparable, mais avec quelques différences. Ainsi, l'A.F.C., qui sous-tend ce système, permet de faire l'économie des règles de l'hyperonomase et de la phagocytose. Le treillis interlinguistique est généré sur la seule base des valeurs [+] contenues dans le tableau 5. La figure 6 montre le treillis créé par l'interface *JaLaBA* lorsqu'on lui fournit ces valeurs.

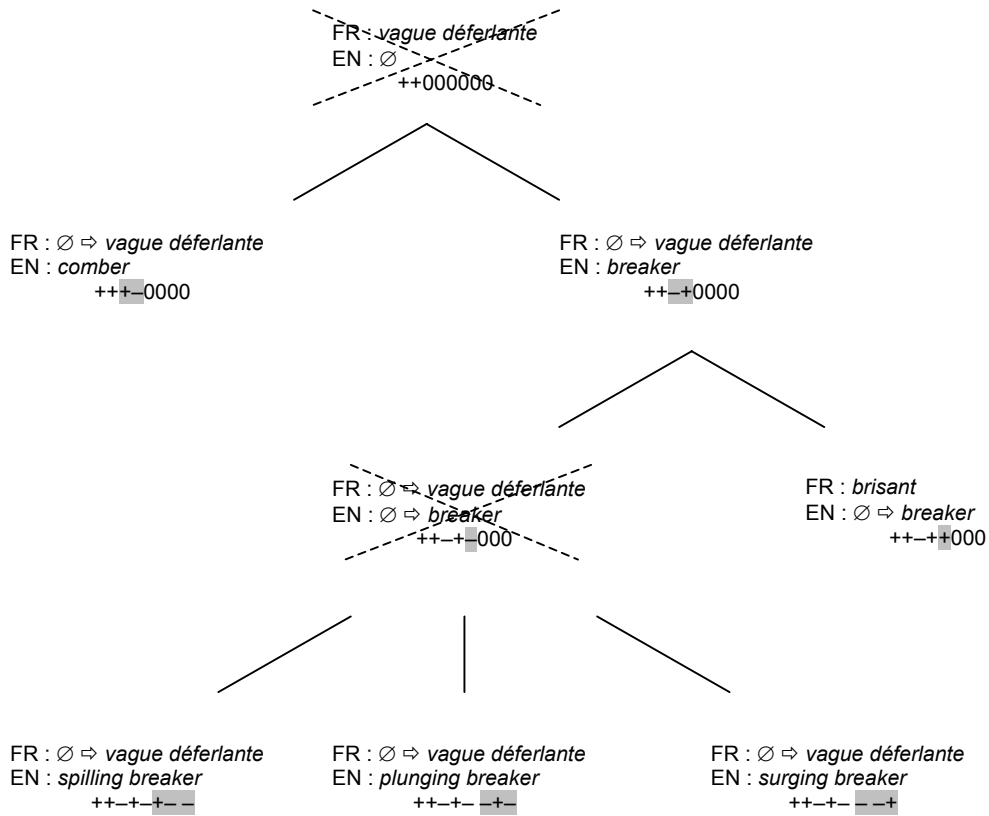


Figure 5 : résolution des équivalences de *vague déferlante* et de *brisant* en anglais selon le modèle de TERMISTI

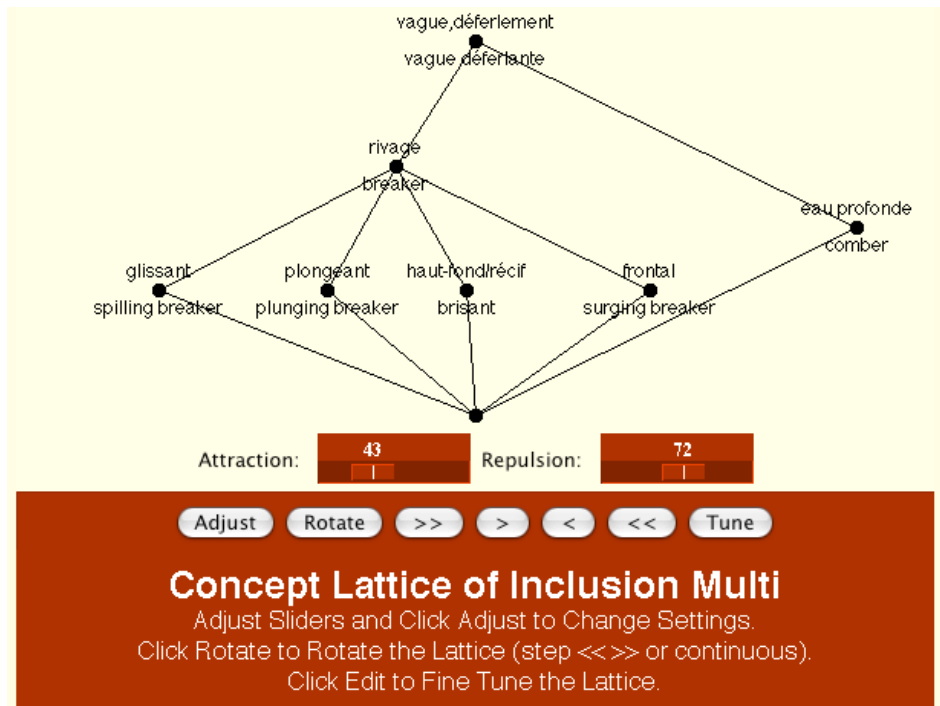


Figure 6 : SIMuLLDA : treillis des vagues déferlantes créé par l'interface *JaLaBA*

Une différence plus marquante est la position subordonnée de *brisant*, conséquence de la distinction à opérer entre sèmes d'un arbre et attributs définitionnels d'un treillis. Selon la représentation arborescente, rien ne semble devoir empêcher un brisant de déferler de manière glissante, plongeante ou frontale. Et puisqu'un *breaker* peut se situer sur un récif, rien n'empêche un brisant d'être un *plunging breaker*. *Brisant* et *plunging breaker* sont des hyponymes du *breaker* en vertu de sèmes distincts : le trait 'haut fond ou récif' dans le cas de *brisant* et les trois traits relatifs au déferlement dans le cas de *plunging breaker*.

Dans un treillis, BRISANT et PLUNGING BREAKER sont subordonnés à BREAKER parce qu'ils s'en distinguent par un seul attribut : respectivement **haut fond/récif** et **plongeant**. Le critère de la localisation sur un haut fond ou un récif est absent de la définition de *plunging breaker* comme celui du déferlement dans la définition du *brisant*.

Le statut de l'équivalence produite différencie également les deux modèles. Dans l'arborescence, *vague déferlante* possède une acception identique à celle de *plunging breaker*, ce qui justifie une équivalence mutuelle. Dans le treillis (figure 6), les deux acceptions interlinguistiques VAGUE DÉFERLANTE et PLUNGING BREAKER ne sont pas identiques : la première est l'hyperonyme de la seconde. Par ailleurs, chaque langue cible contient un trou lexical. Toutefois, on peut dire que VAGUE DÉFERLANTE est identique à PLUNGING BREAKER si on ajoute les deux attributs définitionnels **rivage** et **plongeant**. L'équivalent explicatif pour le trou lexical peut être obtenu en lexicalisant tous les éléments.

plunging breaker : *vague déferlante adoptant un déferlement plongeant à proximité d'une côte*

Une perspective inverse consisterait à « soustraire » des attributs définitionnels : VAGUE DÉFERLANTE = PLUNGING BREAKER – **rivage** – **plongeant**. Un tel mécanisme n'est pas rare en lexicographie : les attributs « superflus » de la langue source sont parfois énumérés devant l'équivalent en langue cible.

vague déferlante : (adoptant un déferlement plongeant à proximité d'une côte) *plunging breaker*

La précision des conditions de l'équivalence ne suppose généralement qu'un attribut et l'exemple qui précède est plutôt extrême. Face à un tel cas, SIMuLLDA optera pour une solution plus commune : générer un équivalent explicatif en utilisant, pour traduire, un hyperonyme de *vague déferlante*, comme par exemple WAVE + **déferlement** : *wave which breaks*.

5.3. L'intersection partielle

Dans les deux situations déjà examinées, les termes entretiennent une relation d'hyponymie et le trou lexical peut indifféremment affecter le subordonné ou le superordonné. Les cas où deux langues possèdent des taxinomies distinctes mais qui se chevauchent partiellement sont beaucoup plus complexes. Un exemple bien connu, quoique incorrect (Janssen 2001 : 129-137), est celui de la distinction opérée en français entre les mots *fleuve* et *rivière* selon qu'ils désignent ou non des cours d'eau se jetant dans la mer et de la distinction opérée en anglais entre les mots *river* et *stream* en fonction de la taille du cours d'eau. Comme les gros cours d'eau ont tendance à se jeter dans les mers, ces deux distinctions ne sont pas sans liaison.

5.3.1. L'analyse sous forme d'arborescence

L'intersection entre deux arborescences peut correspondre à un sémème que le dictionnaire désigne pour établir une équivalence recouvrant toutes les situations envisagées par chaque langue (Van Campenhoudt 1996 : 296-300). L'étude de la désignation des bras de mer fournit une bonne illustration du phénomène (Van Campenhoudt 2001 : 202-203). Le *Dictionnaire hydrographique* établit quatre équivalences entre les termes français *bras de mer* et *section rectiligne*, d'une part, et les termes anglais *reach*, *arm of the sea* et *inlet*, d'autre part. L'interprétation des traits sémantiques contenus dans les définitions (tableau 6) permet de proposer le réseau qui suit (figure 7).

	étendue d'eau	partie de mer	partie d'un fleuve	allongé	resserré	pénétrant les terres	entre port et mer
<i>bras de mer</i> ⁰	+	+	-	0	0	0	0
<i>bras de mer</i> ¹	+	+	-	0	+	+	0
<i>bras de mer</i> ²	+	-	-	+	+	-	-
<i>goulet</i>	+	+	-	0	+	+	+
<i>section rectiligne</i>	+	-	+	0	+	0	0
<i>reach</i> ⁰	+	0	0	0	0	0	0
<i>reach</i> ¹	+	-	+	0	+	0	0
<i>reach</i> ²	+	+	-	0	+	+	0
<i>arm of the sea</i>	+	+	-	+	+	-	-
<i>inlet</i>	+	+	-	0	+	+	+

Tableau 6 : les traits sémantiques des désignations des bras de mer

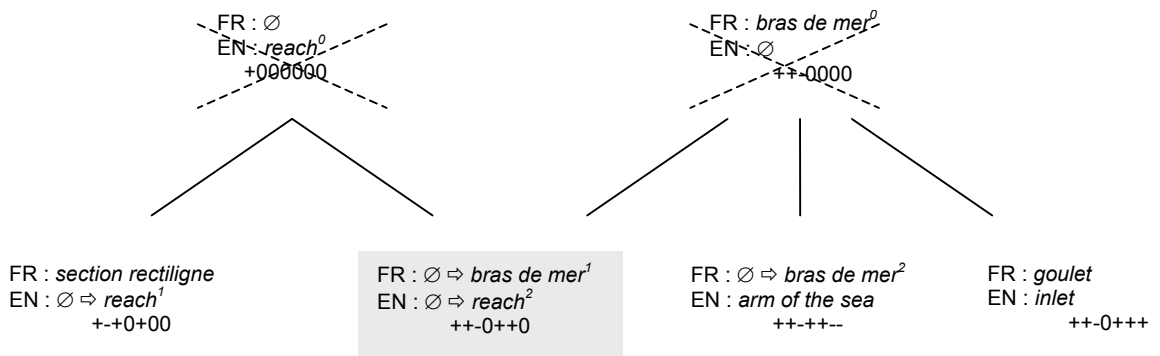


Figure 7 : *bras de mer* et *reach* : résolution de l'intersection partielle selon le modèle de TERMISTI

Le principe d'équivalence permet de résoudre ce cas. Il conduit toutefois à déclarer deux fois la même équivalence entre *reach* et *bras de mer* selon qu'on considère l'anglais ou le français comme langue source (tableau 7).

<p><i>reach</i> = <i>section rectiligne</i> <i>reach</i> = <i>bras de mer</i> et <i>section rectiligne</i> ≠ <i>bras de mer</i></p>	<p><i>bras de mer</i> = <i>reach</i> <i>bras de mer</i> = <i>arm of the sea</i> <i>bras de mer</i> = <i>inlet</i> et <i>reach</i> ≠ <i>arm of the sea</i> et <i>reach</i> ≠ <i>inlet</i></p>
<p><i>section rectiligne</i> = <i>reach</i></p>	

Tableau 7 : application du principe d'équivalence

Cette double déclaration de l'équivalence *bras de mer* = *reach* correspond exactement à un cas d'intersection partielle, comme le montre la figure 7. Son examen approfondi montre que les traductions de *reach* en français (partie gauche du graphe) et de *bras de mer* en anglais (partie droite) sont obtenues après cinq hyperonomases dès lors que les équivalents se situent à un niveau subordonné de la relation hyponymique. On ne compte pourtant que quatre équivalences dans le dictionnaire. En effet, l'adoption d'un réseau multilingue commun au français et à l'anglais conduit à considérer que le sémème {'mer', 'partie étroite', 'intérieur des terres'} est obtenu par deux hyperonomases. Il se situe à l'intersection de deux typologies établies en fonction de critères distincts induits par chaque langue cible : partie de mer ou de cours d'eau en français, partie étroite de la mer ou pénétration de la mer à l'intérieur des terres en anglais. Le nœud qui correspond à l'intersection partielle possède deux hyperonymes (ou archisémèmes) et l'on n'a donc plus affaire à un arbre. Force est ainsi de reconnaître que si le réseau hyponymique identifiable dans une approche monolingue est bien un arbre, le réseau multilingue créé par l'hyperonomase correspond à un graphe acyclique, dont l'arbre et le treillis constituent deux variétés.

5.3.2. L'analyse sous forme de treillis

Le traitement dans un treillis du cas de l'intersection partielle diffère de celui proposé ci-dessus. En effet, cette variété de trou lexical ne constitue pas une catégorie particulière dans l'organisation de SIMuLLDA ou, tout au moins, ne reçoit pas une importance équivalente à celle accordée par l'approche précédente. Deux raisons expliquent cette différence.

D'abord, beaucoup d'exemples d'intersection partielle - sinon la plupart - ne sont pas considérés comme tels dans les treillis mis en œuvre. Ainsi, dans SIMuLLDA, le terme *reach* sera considéré comme polysémique au vu de sa définition dans le *Dictionnaire hydrographique* :

reach. A straight section of a RIVER, especially a navigable RIVER between two bends. See also SEA REACH.
An ARM OF THE SEA extending into the LAND. (O.H.I. 1994-1998)

Celle-ci sera interprétée comme contenant deux acceptions interlinguistiques distinctes, l'une possédant les attributs définitionnels **straight** et **part of a river** et l'autre, les attributs **extending inland** et **part of sea**⁷. Le fait que de tels « semi-homonymes » sont appréhendés comme monosémiques dans les arborescences de TERMISTI est à l'origine de la plupart des cas d'intersection considérés par ce modèle.

Ensuite, et plus fondamentalement, un système de treillis comme celui de SIMuLLDA privilégie les équivalents explicatifs et ne considère pas les cas d'intersection partielle comme relevant d'une catégorie distincte : *reach* et *bras de mer* y seront définis à l'aide d'attributs les distinguant de leur hyperonyme et traités comme de simples trous lexicaux (cf. 4 et 5.2). Ils n'appartiennent à une catégorie distincte que si l'on considère qu'un cohyponyme constitue un meilleur équivalent qu'un hyperonyme ou un hyponyme⁸. On peut penser que les trous lexicaux correspondant à des

7. Une définition comme celle de *The Oxford Companion to Ship and the Sea*, « A straight, or nearly straight, stretch of navigable river or estuary. », (Kemp 1976) pourrait conduire à n'accorder qu'une seule acception interlinguistique à *reach* en utilisant l'attribut définitionnel **part of a river or a estuary**. Toutefois, nous ne discuterons pas ici du traitement des attributs disjonctifs dans l'A.F.C.

8. L'idéal serait que le treillis permette de produire et un équivalent traductif et un équivalent explicatif. L'explication du fait que dans les cas d'intersection partielle, l'équivalent traductif n'est ni un simple hyperonyme ni un unique hyponyme, mais un

intersections partielles n'émergent que lorsqu'on privilégie la recherche d'un équivalent traductif. En effet, ce qui conduit à considérer que *reach* et *bras de mer* se chevauchent partiellement est le fait qu'aucun de ces deux termes ne possède un équivalent parfait et que, dans certaines circonstances, l'un constitue « la meilleure traduction » de l'autre.

6. En guise de conclusion

Les deux approches présentées sont basées sur l'interprétation de définitions existantes. La piste des treillis et de l'A.F.C. a été explorée par SIMuLLDA avec pour objectif de créer une B.D.L.M. respectueuse du réseau sémantique. Le modèle arborescent exploité par le centre TERMISTI est moins respectueux des réalités de chaque langue conçue comme manifestation autonome d'une conceptualisation du monde. Ceci semble lié à sa visée première qui est d'expliquer le fonctionnement de terminographies multilingues existantes, lesquelles « manipulent » la vision du monde de chaque communauté linguistique dans le but louable de permettre une traduction. Cette « manipulation » transparaît très bien dans le cas de l'intersection partielle.

Une équivalence doit fonctionner quelles que soient les langues source et cible et quel que soit l'emplacement du trou lexical. Les modélisations proposées permettent toutes deux de résoudre, à leur manière et sur plusieurs niveaux, l'équivalence entre un superordonné de langue source et un subordonné de langue cible. L'exploitation des sèmes ou des attributs définitionnels permet aussi, dans les deux modèles, d'envisager une lexicalisation du trait distinctif pour proposer une équivalence entre un subordonné de langue source et un superordonné de langue cible.

Il est entendu que les quelques exemples envisagés au fil de ces lignes ne suffisent en aucun cas à rendre compte de la diversité des problèmes d'équivalence qui se présentent à l'intérieur d'un réseau sémantique multilingue et qui ne peuvent être tous résolus dans le cadre de l'hyponymie. En ce sens, cette brève étude n'épuise en rien un vaste sujet d'investigation, qui conduit à situer la terminologie tout à la fois au sein des sciences du langage et au sein de l'ingénierie de la connaissance.

cohyponyme ne peut trouver ici sa place. Une difficulté fondamentale est que le choix de ce cohyponyme ne dépend pas uniquement des attributs définitionnels. De ce fait, la création de l'équivalent traductif souhaité constitue un problème de taille dans des approches sous forme de treillis.

Références bibliographiques

Descotte, S., Husson, J.-L., Romary, L., Van Campenhoudt, M. et Viscogliosi, N., 2001, « 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, pp. 63-81 (Publications from the Centre for Maritime Studies, University of Turku, A36).

Ganter, B. et Wille, R., 1996, *Formale Begriffsanalyse : mathematische Grundlagen*, Berlin : Springer Verlag.

Janssen, M., 2002, *SIMuLLDA: a Multilingual Lexical Database Application using a Structured Interlingua*, Ph.D. thesis, Universiteit Utrecht : Utrecht.

Kemp, P., éd., 1976 : *The Oxford Companion to Ship and the Sea*, London, New York, Melbourne : Oxford University Press.

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

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

Martin, W., Heid, U., Schuurman, I., Beeken, J. et Laureys, G., 1998, *On the Construction of Bilingual Dictionaries. Feasibility Study carried out by order of the European Commission*, The Hague/Stuttgart.

Meyer, I., Eck, K. et Skuce, D., 1997, « Systematic Concept Analysis within a Knowledge-Based Approach to Terminology », Wright, S.E. et Budin, G. (comp.), *Handbook of Terminology Management, Volume I : Basic Aspects of Terminology Management*, Amsterdam and Philadelphia : John Benjamins Publishing Company.

O.H.I., 1994-1998, *Hydrographic Dictionary = Diccionario Hidrográfico = Dictionnaire hydrographique*, 5^e édit., 3 vol., Monaco : Bureau hydrographique international.

Sowa, J.F., 1993, « Lexical Structures and Conceptual Structures », dans Pustejovsky, J., éd., *Semantics and the Lexicon*, Dordrecht : Kluwer.

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 du mât » (1885-1901)*, Université Paris XIII, 2 vol., 431 pp. + annexes (thèse de doctorat en sciences du langage).

Van Campenhoudt, M., 1996, « Réseau notionnel, intelligence artificielle et équivalence en terminologie multilingue : essai de modélisation », dans Clas, A., Thoiron, Ph. et Béjoint, H., éd., *Lexicomatique et dictionnaires, IV^{es} journées scientifiques du réseau thématique "Lexicologie, terminologie, traduction"*, Université Lumière (Lyon II), 28-30 septembre 1995, Montréal : AUPELF-UREF et Beyrouth : F.M.A., pp. 281-306.

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.), pp. 127-152.

Van Campenhoudt, M., 2001, « Pour une approche sémantique du terme et de ses équivalents », *International Journal of Lexicography*, XIV, 3, 181-209.

Vossen, P., 1998, « EuroWordNet: building a multilingual database with wordnets for European languages », *The ELRA Newsletter*, February 1998, III, 1, 7-10.

Zgusta, L., 1971, *Manual of Lexicography*, Den Haag : Mouton.

Résumé – Abstract

One of the major problems in multilingual terminology is the existence of lexical gaps. Amongst the models that have been proposed to deal with this problem are the TERMISTI model, using a tree, and SIMuLLDa, using a lattice. The TERMISTI model is prompted by a practical need; the SIMuLLDA system is a practical implementation of a formal framework. This article gives a brief overview and a comparison of these two models, and how they deal with lexical gaps. A central question is whether the concrete cases raised by the TERMISTI project can be correctly dealt with in the more formalised SIMuLLDA system. The major obstacle for the SIMuLLDA system are the cases of partial overlap, where two terms have overlapping, but non-identical meanings.
