

Sonia Szramek-Karcz

*Université de Silésie,
Katowice, Pologne*

L'Approche Orientée Objets ou l'EuroWordNet ? Quelle approche pour la traduction automatique ? Partie II : hiérarchie, héritage et désambiguïsation

Abstract

The presentation of research results concerning applicability of AOO and EWN in machine translation is divided into two parts. In the first part, the author presents the influence of databases design, their theoretical aspects and the categorization of lexical items on the process and result of machine translation. The second part is devoted to the internal hierarchy of databases including the established categorization of lexical items as well as semantic inheritance comprising multiple inheritance and disambiguation accomplished by the above mentioned databases. The presented results coupled with the capability of contemporary computer technology give grounds for claiming that a bilingual approach in lexical description for machine translation is the best choice.

Key words

Object Oriented Approach (AOO), EuroWordNet (EWN), machine translation, lexical databases, ontology, object class, object, synset, semantic inheritance, disambiguation.

1. Introduction

Après avoir analysé les origines, les principes et l'organisation de l'AOO et de l'EWN en vu de leur utilité pour la TA dans la première partie de notre propos, passons à l'évaluation de ces deux bases de données lexicales en regardant de près leur hiérarchie, héritage sémantique et la façon dont elles permettent de désambiguïser leur contenu lexical.

2. Hiérarchie et héritage

Les classes d'objets de l'AOO et les synsets du EWN dont l'efficacité pour la TA a été montrée dans la première partie de notre propos, ne sont pas suspendus dans le vide, bien au contraire, ils ont leur place bien déterminée dans la hiérarchie des bases de données. De subtiles différences entre l'AOO et l'EWN influent sur les résultats des traductions automatiques par ailleurs surprenants.

Les classes hypo et hyperonymes dans l'AOO (comme les synsets dans le WN) sont désignées suivant la relation X EST-UNE SORTE DE Y. Les hiérarchies lexicales de type IS-A-KIND-OF sont largement utilisées à représenter le savoir (Sowa, 2000). Elles permettent d'éviter des cercles vicieux, *vacuous circles* (Miller, 1990 : 247), définir *idem per idem* n'est plus possible. L'avantage de l'application des hiérarchies lexicales (dans le WN comme dans l'AOO) réside principalement dans l'économie de la description garantie par le système d'*héritage* auquel nous allons revenir tout à l'heure.

Dans l'AOO, les classes d'objets ne constituent pas un listing (ce qui a lieu dans l'EWN avec les synsets au niveau d'**Inter-Lingual-Index**) mais sont structurées de telle sorte que chacune d'elles possède sa super- et sa sous-classe. Le seul point commun entre les unique beginners et les classes de l'AOO sont les super-classes qui se trouvent au sommet de la hiérarchie dans l'AOO basées sur les classes conceptuelles de WordNet appelés « unique beginners ». Cela a permis de diviser les tâches des lexicographes et d'organiser le travail de description des objets dans l'AOO. Comme la description dans l'AOO s'effectue d'en bas de l'échelle des classes, il peut bien s'avérer un beau jour que les super-classes soient différentes de ce qui a été attendu au départ. Ce sont les descriptions des objets, la langue, qui décideront du nombre, du contenu et de la hiérarchie des classes d'objets.

Dans l'EWN, chacun des WordNets nationaux, avant de constituer une liste non-structurée des synsets (Inter-Lingual-Index), possède sa propre hiérarchie lexicale gérée par les relations de synonymie, hyperonymie, et hyponymie. Certes, la *synonymie* est la relation sémantique de base entre les mots, mais dans l'organisation des *synsets* c'est la relation de subordination, appelée dans ce contexte l'*hyponymie*, qui importe. Par exemple, le nom *nurse* a un hyponyme (subordonné) du nom caregiver, ou inversement, caregiver est un *hyperonyme* (super ordonné) du nom *nurse*. C'est cette relation sémantique qui organise les noms dans la hiérarchie lexicale wordnetienne. L'*hyperonymie* est une relation entre des sens particuliers des mots. Nous avons donc une relation d'*hyperonymie* distincte pour chaque sens du mot de WordNet. Les hyperonymes de l'infirmière (*nurse*) sont les suivants : {*nurse*} @→ {health professional, health care provider, caregiver} @→ {professional, professional person} @→ {adult, grownup} @→ {person, individual, someone, somebody, mortal, soul (a human being)} @→ {organism, being} @→ {living thing, animate thing} @→ {object, physical object} @→ {physical entity}

@→ {entity} (les parenthèses indiquent un synset, et @→ est une relation sémantique à lire « est une sorte de ». Ce dernier « entity » est le « unique beginner » pour tous les noms dans la base).

Pour chaque relation hypéronyme, il y a une relation hyponyme correspondante. La relation sémantique d'hypéronymie dans le WordNet notée dans les fichiers par @→ part du spécifique au général, est une généralisation. L'hyponymie par contre, notée ~ → peut être lue « inclut » (*ang. subsumes*) va du générique vers le spécifique est donc la relation de spécification.

De l'application des relations hyper et hyponymie émerge l'architecture lexicale de WordNet appelée aussi « tree diagram » qui a été implanté également dans l'EWN.

Malheureusement l'expression KIND OF (une sorte de) n'est pas univoque et il est nécessaire d'être vigilant pendant le diagnostic. David A. Cruse (1986 : 138) attire l'attention sur trois cas où la relation de taxinomie ne s'applique pas : une personne (*the speaker*) s'en sert faute de nom exact, exprime son doute sur les caractéristiques typiques de l'objet dont on parle, accentue avec l'intonation son étonnement ou parle avec l'ironie. Dans son étude, Andrée Borillo (1996) distingue d'autres structures qui permettent de « repérer dans les textes des structures très locales susceptibles d'être interprétées comme des mises en relation de type hypéronymique » (Borillo, 1996 : 121). Examinons des classements des noms désignant les professions pour les enfants en bas âge dans un des livres sur l'hôpital une page intitulée : « les professionnels de l'hôpital » ou l'auteur présente d'un filé les occupations comme : *médecin, interne, infirmier, chirurgien, anesthésiste, aide-soignant, brancardier, agent de service hospitalier, psychologue*. Les bases linguistiques qui décrivent et organisent les données lexicales suivant le critère linguistique comme l'AOO ou le Dictionnaire électronique du LLI (Gross, 1992, 1994a, 1994b, 1995) rangent toutes ces occupations, ces professions, dans la même classe selon les critères des opérateurs et attributs fournis par la langue. Mais dans le WN en revanche, on retrouve le brancardier (EN : *stretcher-bearer* ; PL : *noszowy*) parmi les travailleurs (*worker*) et les assistants (*assistant*) :

Noun

- ◀ S: (n) **stretcher-bearer**, litter-bearer (one who helps carry a stretcher)
 - ◀ direct hypernym / inherited hypernym / sister term
 - ◀ S: (n) attendant, attender, tender (someone who waits on or tends to or attends to the needs of another)
 - ◀ S: (n) assistant, helper, help, supporter (a person who contributes to the fulfillment of a need or furtherance of an effort or purpose) *my invaluable assistant; they hired additional help to finish the work*
 - ◀ S: (n) worker (a person who works at a specific occupation) *he is a good worker*

- ◀ S: (n) person, individual, someone, somebody, mortal, soul (a human being) *there was too much for one person to do*
- ◀ S: (n) organism, being (a living thing that has (or can develop) the ability to act or function independently)
- ◀ S: (n) living thing, animate thing (a living (or once living) entity)
 - ◀ S: (n) whole, unit (an assemblage of parts that is regarded as a single entity) *how big is that part compared to the whole?; the team is a unit*
 - ◀ S: (n) object, physical object (a tangible and visible entity; an entity that can cast a shadow) *it was full of rackets, balls and other objects*
 - ◀ S: (n) physical entity (an entity that has physical existence)
 - ◀ S: (n) entity (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

C'est vrai que pour accéder à la « profession de brancardier », il n'existe pas de diplôme spécifique mais comme nous pouvons lire sur le site du gouvernement français (in <http://www.metiers.santesolidarites.gouv.fr/IMG/pdf/12-MGA-Brancardier.pdf>), les postulants doivent cependant disposer du brevet national de secouriste (BNS). Comment classer les activités considérées par les uns comme profession et par les autres comme une activité n'ayant rien en commun avec un métier quelconque pour ne citer que quelques exemples : *nounou*, *prêtre*, *tueur à gages*, *magicien*, *vendeur dans la rue*. Avec la hiérarchie "IS A" l'AOO applique le critère linguistique selon lequel, c'est la langue qui décide de l'appartenance d'un objet

à une classe d'objets et par la suite, grâce aux opérateurs et aux attributs place une classe dans la hiérarchie régit par le système d'héritage sémantique.

3. Héritage

L'héritage est le plus apprécié dans la description des noms (pour les verbes cf. Fellbaum, 1990, 1998 ; et adjectifs cf. Gross, Miller, 1990 ; Miller, 1998). Si on nous dit par exemple que *Pierre travaille comme chirurgien*, *Pierre* (instance d'objet de *chirurgien* dans l'AOO) hérite toutes les propriétés du *chirurgien* et de ses hyperonymes allant du plus précis au plus général.

La différence entre la hiérarchie lexicale de WordNet et celle adoptée dans l'AOO est que la première est ontologique et la deuxième linguistique. Elles se recouvrent partiellement car la langue nomme la réalité et *chirurgien* 'surgeon' et le nom désignant la *profession* 'professional, professional person' du point de vue ontologique et lexicale, pourtant il ne doit pas être *adulte* du point de vue de la langue, même si, dans la réalité, c'est ce qui a lieu dans la majorité des noms de professions.

Les hyperonymes de *chirurgien* 'surgeon' dans le WordNet :

- ◀ S: (n) doctor, doc, physician, MD, Dr., medico (a licensed medical practitioner)
I felt so bad I went to see my doctor
- ◀ S: (n) medical practitioner, medical man (someone who practices medicine)
- ◀ S: (n) health professional, health care provider, caregiver (a person who helps in identifying or preventing or treating illness or disability)
- ◀ S: (n) professional, professional person (a person engaged in one of the learned professions)
- ◀ S: (n) adult, grownup (a fully developed person from maturity onward)
- ◀ S: (n) person, individual, someone, somebody, mortal, soul (a human being) *there was too much for one person to do*
- ◀ S: (n) organism, being (a living thing that has (or can develop) the ability to act or function independently)
- ◀ S: (n) living thing, animate thing (a living (or once living) entity)

- ◀ S: (n) object, physical object (a tangible and visible entity; an entity that can cast a shadow) *it was full of rackets, balls and other objects*
- ◀ S: (n) physical entity (an entity that has physical existence)
 - ◀ S: (n) entity (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

Les hyponymes de *chirurgien* 'surgeon' dans le WordNet

- S: (n) amputator (a surgeon who remove part or all of a limb)
- S: (n) cosmetic surgeon, plastic surgeon (a surgeon who beautifies the body (especially the face))
- S: (n) neurosurgeon, brain surgeon (someone who does surgery on the nervous system (especially the brain))

Si la différence dans la classification ontologique (WordNet) et linguistique (AOO) au niveau de « chirurgien » consiste en « être ou ne pas être adulte », ces deux classements divergent dans la description des activités humaines charnière, celles sur lesquelles on pourrait discuter si elles appartiennent aux professions ou pas. WordNet comme toutes les ontologies (Chandrasekaran, Josephson, Richard Benjamins, 1999) présente une sorte de hiérarchie des relations entre les nœuds qui parfois, varie considérablement des classements linguistiques. Il suffit d'analyser les occupations comme : *nounou, prêtre, tueur à gages, prostituée, magicien, vendeur dans la rue*, pour voir à quel point le classement linguistique peut varier du classement ontologique, mais aussi à quel point les classements des WordNets nationaux varient entre eux.

Nous voudrions nous attarder sur le phénomène d'héritage multiple envisagé par l'AOO, absent dans l'EWN. On parle de l'héritage multiple dans le cas où un objet hérite des propriétés sémantiques de plus d'une classe d'objets, autrement dit une classe d'objets n'a pas seulement son géniteur (une super-classe) mais aussi sa génitrice (une autre super-classe). Cela n'est pas possible quand l'objet appartient à des classes qui ne possèdent pas d'hyponymes c'est à dire des *unique begin-*

ners du WordNet, ce qui est logique, comme les premiers hommes Adam et Ève n'auraient eu de parents eux non plus.

L'héritage multiple illustre le mieux <l'amphibie> qui hérite de la classe, <voiture> et de la classe <bateau>. Le problème peut apparaître lors de la création des sous-classes d'une classe qui a hérité ses opérateurs et ses attributs de plusieurs classes car on risque de se perdre dans la construction des arrières-grands-enfants (les risques d'ambiguïté dans l'héritage multiple sera soulevé en détail dans une publication consacré à l'héritage sémantique, le problème de point de vu de la programmation à objets est décrit par Jerzy Grębosz (2003 : 526)).

L'héritage multiple est impossible si les deux objets appartiennent aux deux classes descendantes des différentes superclasses des *unique beginners*, ce qui garantit la traduction correcte. Nous allons le montrer à partir de l'exemple de « avocat ». Le mot français *avocat*, selon l'AOO, est répertorié dans deux classes différentes : <plantes> et <humains> sans en hériter pour autant des opérations et des attributs. Dans ce cas, on a affaire à deux objets différents (deux classes différentes selon le principe qu'il y a autant de classes d'objets que d'ensembles d'opérations et d'attributs différents) qui se caractérisent par un ensemble d'opérateurs et d'attributs différents : *avocat* ayant comme superclasse <humain> : (opérateurs manipulateurs :) *admettre au barreau, radier du barreau, consulter un avocat* ; (opérateurs accesseurs :) *plaide, défend, conseille*, et *avocat* ayant comme superclasse <fruits> (opérateurs manipulateurs :) *éplucher, manger* ; (opérateurs accesseurs :) *pousse, mûrit*. Ceci dit, l'architecture souple de l'AOO nous permet l'application de l'héritage multiple en agençant les classes comme par exemple pour le *médecin-chef* (pl. *ordynator*) — défini comme *médecin qui dirige un service de l'hôpital* hérite des classes : <médecins> (ou dans notre nomenclature : <personne qui travaille en soignant des êtres vivants>) et *chef*, <personne qui travaille en dirigeant>. L'approche modulaire dans la description du module *travailler comme médecin-chef* selon le principe de décomposabilité y retrouve les modules de *travailler comme médecin* et *travailler comme chef*. De même pour le *cordiste*, il suffirait de le bien placer parmi les modules, les classes etc.

Dans le WN, nous retrouvons des synsets qui possèdent plusieurs hiérarchies comme par exemple S : (n) *spouse*

◀ S: (n) spouse, partner, married person, mate, better half (a person's partner in marriage)

◀ S: (n) relative, relation (a person related by blood or marriage) *police are searching for relatives of the deceased; he has distant relations back in New Jersey*

◀ S: (n) person, individual, someone, somebody, mortal, soul (a human being) *there was too much for one person to do* etc.

- ◀ S: (n) domestic partner, significant other, spousal equivalent, spouse equivalent (a person (not necessarily a spouse) with whom you cohabit and share a long-term sexual relationship)
- ◀ S: (n) person, individual, someone, somebody, mortal, soul (a human being) *there was too much for one person to do*
etc.

Si on ne se concentrait que sur le premier synset : S: (n) spouse, partner, married person, mate, better half (a person's partner in marriage), et si on imaginait les WordNets polonais, français et anglais, on aurait bien du mal à associer les sunsets qui représentent la notion de *spouse* en anglais car la relation simplifiée entre : PL *żona* — FR *femme* — EN *wife*, et PL *małzonka* — FR *épouse* — *spouse* (en) ne se retrouve pas dans l'emploi de ces termes pour ne citer que des exemples des phrases : PL: *prezydent wraz z małzonką*, FR *le président avec son épouse*, EN *president with his wife*. Nous touchons ici le problème de distinction des sens d'un nom.

4. Désambiguisation

La façon de distinguer les sens d'un mot dans ces deux bases permettra de trancher définitivement la question débattue dans cet article, à savoir laquelle de ces deux descriptions du lexique, l'EWN ou l'AOO s'adapte mieux à la TA.

L'approche orientée objets comme son nom l'indique a choisi les objets au centre d'architecture, les objets qui sont caractérisés par leurs opérations et leurs attributs. Le sens des mots est donc vu par le biais de son entourage étant proche des travaux de John Rupert Firth (1951); Michael Alexander Kirkwood Halliday, Ruqaiya Hasan, (1976); John McHardy Sinclair, (1987) des ressemblances des familles de Ludwig Wittgenstein (1953), et des travaux de Gaston Gross et de son école où le sens de mot est défini par son emploi (*la notion d'emploi* in Gross 1998a : 104—109 ; 1996 : 195—212).

La sémantique du WordNet se base sur la notion du sens des mots que l'on rencontre d'habitude dans les dictionnaires. C'est donc dans l'organisation de ces informations que réside la particularité du WordNet qui essaie de rendre les relations sémantiques entre les sens des mots, plus explicites et plus conviviales « easier to use » (in Miller, 1998 : 23). Même si un *synset* est accompagné d'une explication connue des dictionnaires traditionnels (depuis 1989 les définitions font partie des synsets), il n'est pas une *entrée du dictionnaire* au sens propre du mot. Une entrée du dictionnaire « classique » peut être représenté par plusieurs *synsets*, car un *synset* égale à un sens du mot. Par exemple le nom : « nurse » : infirmière/nounou est représentée par deux *synsets* différents.

- S: (n) nurse (one skilled in caring for young children or the sick (usually under the supervision of a physician))
- S: (n) nanny, nursemaid, nurse (a woman who is the custodian of children)

Les synsets des wordnets différents sont par la suite mises en relation, ce que nous avons décrit brièvement dans les troisième et quatrième parties du présent article.

La garantie de la traduction correcte des unités décrites dans l'AOO est leur corrélation instantanée avec leur correspondant en langue cible. Ceci dit les mots sont désambiguïsés sur le champs en fonction des besoins qui se présentent. C'est une différence majeure entre la description AOO et EWN sur laquelle nous allons nous attarder dans nos analyses.

Tandis que dans l'AOO les équivalents en langue 2 sont inséparables de la description en L1, L'EuroWordNet laisse cette tâche aux outils informatiques. De plusieurs solutions d'intégrations des WordNets nationaux (les solutions envisagées ont été présentées par Piek Vossen, Pedro Diez-Orzas, Wim Peters, 1997:1) a été retenue celle de deux niveaux : le niveau qui dépend de la langue (le wordnet spécifique pour chaque langue (*the language-dependent layer*) et le niveau qui ne dépend pas de la langue (*the language-independent layer*) appelé aussi une base des données indépendante des langues ou *interlingual index* (ILI), un supérieur (*the top-layer*) et un domaine des ontologies (*domain ontologies*)). Nos analyses se concentraient sur les deux premiers niveaux : celui de «vrai» WordNet spécifique pour chaque langue (*Each wordnet represents a unique language-internal system of lexicalizations* : <http://www.illc.uva.nl/EuroWordNet/>) et celui de *Inter-Lingual-Index* qui permet la traduction des mots d'une langue à l'autre. Nous laissons de côté (pour ne pas aller au-delà de nos analyses) une trentaine des concepts de bases qui ont été sélectionnés pour toutes les langues européennes et les remaniements possibles de cette liste de concepts.

Le projet de EuroWordNet (selon les recommandations de la Commission Européenne) encourage à créer des bases de données qui ne sont pas les traductions de la version anglaise du WordNet (WordNet 1.5), mais qui rendent compte de la spécificité de la langue qu'elle décrivent et de sa conceptualisation. Ce principe, juste qu'il soit, rend la connexion des wordnets et ainsi la TA très complexe, exigeant des outils informatiques performants (Novell ConceptNet toolkit (Díez-Orzas *et al.*, 1995). Comme l'observe Sook Hee Park (2004 : 2611) *la secrétaire* en français (selon *Le Petit Larousse Illustré*, cf. les sens du TLFi:<http://atilf.atilf.fr/dendien/scripts/tlfiv5/visusel.exe?11;s=2991427185;r=1;nat=;sol=0;>) a cinq sens avec deux significations différentes comme « secrétaire de rédaction » (sub-editor), « secrétaire de plateau » (scriptwriter). Le troisième sens de l'anglais est absent. En Coréen /*pyo*/ (*Pyojun Grand Dictionnaire du Coréen*), qui correspondrait à la secrétaire en anglais, est homonymique et le cinquième parmi les six a quatre sens : un correspond au sens 3 du WordNet (secretary, secretarial assistant), le quatrième

ressembla à : **secretary** (a person who is head of an administrative department of government) avec la signification d'un *ministre dans les pays socialistes* qui n'est pas citée par d'autre dictionnaire. Cette multitude de sens dans le WordNet (et le EuroWordNet de même car il partage les mêmes principes) ne résout donc pas le problème de la traduction et n'en fait pas un outil parfait de désambiguïsation (Ide, Véronis, 1998 : 13 “:fine-grainedness of WordNet's sense distinctions”) sans oublier la forme féminine des noms qui parfois crée un nouveau sens (en polonais *sekretarz—sekretarka*) ou la forme masculine : *sekretarz* (secretary) qui n'a pas son homologue féminin.

La confection des dictionnaires traditionnels oblige à énumérer les sens d'un mot pour faciliter la lecture et la compréhension du mot donné. Les sens des mots des dictionnaires varient considérablement.

Reprenons l'exemple du substantif : « la secrétaire » qui paraît relativement plus facile à traduire du français vers l'anglais, vers le polonais et inversement (secrétaire — *sekretarz / sekretarka* — secretary) que vers le coréen.

Le Trésor de la Langue française informatisé (<http://atilf.atilf.fr/tlf.htm>) distingue « secrétaire » qui désigne respectivement : une personne, un manuel, un meuble, un oiseau. Dans le premier groupe nous avons les distinctions en :

A.

1. *Subst. masc.*

- a) *Vieilli*. Celui qui était attaché à un personnage de haut rang pour rédiger des lettres
- b) Membre de l'administration vaticane préposé à l'expédition des actes officiels.

2. *Secrétaire d'État*

- a) *Subst. masc., vieilli. Secrétaire d'État et des commandements.*
- b) *Subst. masc.* Titre du collaborateur du Président des États-Unis d'Amérique s'occupant des affaires étrangères.
- c) Membre du gouvernement dont le rôle et le statut sont variables.

3. ***Secrétaire d'ambassade***, *subst. masc.* Celui qui était autrefois officiellement chargé de rédiger les dépêches de l'ambassade ; de nos jours, diplomate situé, dans la hiérarchie administrative, entre l'attaché et le conseiller.

B. *Subst. masc. et fém.*

1. Personne qui rédige certaines pièces officielles

POL. Premier secrétaire (du parti). Titre du principal responsable de certains partis politiques.

2. *ADMIN.* Fonctionnaire, agent ou employé(e) qui dirige certains services ou assure certains travaux administratifs, de rédaction ou d'écritures.

3. *En partic.*

- a) ***Secrétaire (général(e))***. Titre employé dans diverses sociétés privées, dans des organismes privés ou publics ; *en partic.*, synonyme de *directeur administratif*.

b) *Secrétaire d'édition.* « Dans la presse de province, journaliste chargé de préparer une édition régionale ou locale »

C. *Subst. masc. et fém.*

1. Personne qui écrit sous la dictée ou qui rédige pour le compte de quelqu'un.

2. Employé(e) chargé(e) de divers travaux (dactylographier du courrier, rédiger du courrier, classer des documents, répondre au téléphone); *secrétaire de direction* (souvent *au fém.*). « Collaborateur immédiat d'un chef d'entreprise, d'un directeur ou chef de service, préparant et réunissant les éléments de son travail » (*Mét.* 1955).

Secrétaire médical(e). Employé(e) qui assiste un médecin, un dentiste ou travaille dans l'administration d'un établissement de soins.

3. *Secrétaire de séance.* Personne désignée ou choisie par une assemblée pour rédiger le compte rendu écrit ou le procès-verbal des travaux de cette assemblée.

4. *Vieilli.* Manuel comportant des modèles de correspondance.

II. *Subst. masc., MOBILIER.* Meuble à tiroirs où l'on range des papiers, pourvu généralement d'un abattant sur lequel on peut écrire.

III. *Subst. masc., ORNITH.* Synon. de *serpentaire*. *L'oiseau nommé secrétaire, quelques autres tels que les grues [...] sont aussi fort intrigués quand on paraît les observer*

Nous retrouvons la division des sens similaires du mot « secrétaire » proposé dans le Robert (Le Robert, 1981 : 187).

Dans les dictionnaires anglais (ne devrions nous pas dire : dans la langue anglaise) par contre apparaît encore un sens du “secretary” : (Webster's third New International dictionary, copyright 1986 by MERRIAM-WEBSTER INC.p 2052 vol III).

1: obs [*Vieilli*]

a one entrusted with the secrets or confidences of a superior

b one considered to understand the secrets of God or of the nature or to have penetrated other mysteries

2: one employed to handle correspondence and manage routine and detail work for a superior

3:

a: an officer of a business concern who may issue notices and keep records of directors [...]

b: an officer of an organization or society responsible for its records and correspondence

4:

an officer of state who superintends a government administrative department [...]

5:

a a writing desk

b secretary bookcase : a writing desk with a top section for books

6: SECRETARY BIRD

Les entrées à part constituent :

l'oiseau : **Secretary bird**, ainsi que secrétaire général : **secretary-general**, secrétaire d'État : **secretary of state**

The Oxford English Dictionary (vol. 14 : 839—840) présente (vu les 17 volumes qui le lui permettent)

- 1.a. One who is entrusted with private or secret matters, a confidant
- b. fig, of things personified
- c. Applied to those entrusted with the secrets or commands of God, or of a god
- d. secretary of nature: one acquainted with the secrets of Nature
- 2.a. one whose office it is to write for another
- b. transf. and fig
- c. One who writes (on a particular occasion) for another
- d. One skilled in letter-writing
- e. In the titles of books on the art of letter-writing
- 3.a. In the official designation of certain ministers presiding over executive departments of state.
- b. Mr. Secretary : used before the name of a secretary of state, or as a title instead of his name.
4. Short for secretary hand, type
5. A writing desk, a secretaire
6. The secretary-bird

En polonais la situation se complique avec la forme féminine du nom qui ne couvre pratiquement qu'un sens (pour le moment) du nom masculin "sekreterarz"

<http://sjp.pwn.pl/lista.php?co=sekreterka> *Słownik języka polskiego PWN*

◀ sekreterka

1. «pracownica zajmująca się pracami biurowymi związanymi z działalnością jakiejś osoby lub instytucji»
2. «urządzenie nagrywające informacje przekazywane przez osobę dzwoniącą, gdy abonent nie może odebrać telefonu»

sekreterarz

1. «pracownik zajmujący się pracami związanymi z działalnością jakiejś osoby lub instytucji»

2. «osoba kierująca pracą administracyjno-organizacyjną w instytucjach i urzędach»
3. «obieralna funkcja w różnych organizacjach»
4. «w dawnej Polsce: pisarz lub komornik królewski albo tytuł honorowy nadawany przez króla uczonym i literatom»
5. «drapieżny ptak o stalowoszarym upierzeniu żyjący w Afryce»

Słownik współczesnego języka polskiego (Warszawa: Wilga 1999, 300—301)

sekretarka

1. kobieta pracująca w sekretariacie
2. pot. urządzenie podłączone do aparatu telefonicznego automatycznie nagrywające rozmowy pod nieobecność abonenta i później odtwarzające je

sekretarz

1. pracownik prowadzący korespondencję, obsługę interesantów, spraw bieżących jakiejś instytucji lub osoby
2. funkcja w instytucji naukowej, organizacji politycznej itp. pełniona przez osobę wybieraną spośród grona pracowników, członków
3. W dawnej Polsce: pisarz lub komornik królewski; również nadawany przez króla zasłużonym uczonym, pisarzom.

Le WordNet (la version 3,0) propose 4 synsets pour “secretary”:

Noun

- ◀ S: (n) **secretary** (a person who is head of an administrative department of government) sekretarz
- ◀ S: (n) **secretary**, secretarial assistant (an assistant who handles correspondence and clerical work for a boss or an organization) sekretarz/sekretarka
- ◀ S: (n) repository, **secretary** (a person to whom a secret is entrusted) powiernik
- ◀ S: (n) **secretary**, writing table, escritoire, secretaire (a desk used for writing) sekretarzyk

Le dernier sens de “writing table” correspond selon sa taille au polonais : *sekretarzyk* ou *sekretera* selon les définitions ci-dessous :

sekretarzyk «biurko w kształcie stolika z nadstawką zawierającą szufladki»
sekretera «biurko, którego dolna część ma kształt komody, a górna, zawierająca szufladki, zasłaniana jest płytą, stanowiącą po opuszczeniu blat do pisania»

Cela nous permet de supposer, vu la spécificité et l’originalité de chacun des WordNets nationaux, qu’il y aura un problème d’équivalence dans les cas où les divisions des sens ne se retrouveraient pas. Ce problème (*matching probleme*) dont

les participants au projet parlent souvent et dont ils proposent des solutions éventuels en analysant les cas précis est bien illustré dans le *Multilingual Central Repository* où la signification : S: (n) repository, secretary (a person to whom a secret is entrusted) n'a pas de traduction espagnol ni catalan ni basque.

Dans l'AOO, le manque d'équivalent en langue cible n'apparaît même pas, car dans l'AOO un mot a autant de sens que de traductions dans la L2 donnée. Cela veut dire également que dans l'AOO qui est une approche binaire, le nombre des sens d'un mot varie en fonction des paires de langues prises en compte.

Peut-être la solution proposée par Park de créer un WorNet Coréen qui est la traduction du WordNet anglais serait à retenir ? Alessandro Artale, Anna Goy, Bernardo Magnini, Emanuele Pianta, Carlo Strapparava (1998) proposent de diminuer la multitude des sens et d'augmenter le degré de désambiguïsation en étiquetant les domaines auxquels les mots appartiennent (*Adding Subject Field Labels*). Dans l'AOO, un **champ** dans la fiche descriptive est dédié à l'indication du domaine et le nombre des sens correspond aux nombres des traductions dans la langue de départ (Banyś, 2005).

Dans l'AOO, l'objet est caractérisé par l'ensemble des opérations qu'il effectue et qui sont effectuées sur lui ainsi que par des attributs qui lui sont assignés, par contre, le WordNet place les noms dans une hiérarchie profonde et bien (sur) développé mais sans leur structure. Bonnie Jean Dorr, M. Antonia Martí et Irene Castellón (1998 : 397) proposent de résoudre ce manque en complétant le WordNet avec Lexical Conceptual Structure Representations. La sélection lexicale dans la traduction automatique est incontournable et constitue un des problèmes de l'EuroWordNet (Vossen, Díez-Orzas, Peters, 1997 : 8, a language-specific difference in lexicalisation).

- ◀ S: (n) prosecutor, public prosecutor, prosecuting officer, prosecuting attorney (a government official who conducts criminal prosecutions on behalf of the state)
 - ◀ S: (n) official, functionary (a worker who holds or is invested with an office)
 - ◀ S: (n) skilled worker, trained worker, skilled workman (a worker who has acquired special skills)
 - ◀ S: (n) worker (a person who works at a specific occupation)
 - he is a good worker*
 - ◀ S: (n) person, individual, someone, somebody, mortal, soul (a human being) *there was too much for one person to do*
 - ◀ S: (n) organism, being (a living thing that has (or can develop) the ability to act or function independently)
 - etc.

- ◀ S: (n) lawyer, attorney (a professional person authorized to practice law; conducts lawsuits or gives legal advice)
- ◀ S: (n) professional, professional person (a person engaged in one of the learned professions)
 - ◀ S: (n) adult, grownup (a fully developed person from maturity onward)
 - ◀ S: (n) person, individual, someone, somebody, mortal, soul (a human being) *there was too much for one person to do*
 - ◀ S: (n) organism, being (a living thing that has (or can develop) the ability to act or function independently)
 - etc.

Non seulement la présente hiérarchie ne garantit pas la traduction correcte car en polonais le mot anglais *prosecutor* reçoit au moins quatre traductions différentes comme : *oskarżyciel*, *prokurator*, *adwokat*, *powód* / *powódka*. La description dans l'AOO prend en compte le nombre des traductions possibles, car il y a des cas où l'indication du domaine ne garantit pas encore une traduction adéquate. C'est le cas de *prosecutor* dont toutes les traductions appartiennent au domaine juridique (English-Polish Dictionary PWN—Oxford 2005 : 937).

Mettre en relation plusieurs WordNets (Vossen, 1996) est un vrai défi. Pour améliorer la désambiguïsation des synsets des WN et ainsi garantir la connexion des synsets convenables au niveau de Inter-Lingual-Index, plusieurs solutions sont proposées par les auteurs du projet comme nous l'avons vu mais aussi par d'autres linguistes. Parmi elles, la proposition d'ajouter une nouvelle relation sémantique (Artale, Goy, Magnini, Pianta, Strapparava, 1998) appelée *pertain-to-subject* relation de l'anglais : *the lexical concept identified by synset S1 pertains to the subject field identified by synset S2*. Cela correspond aux domaines de l'AOO. Les domaines sont ainsi ajoutés aux noms, et les structures sélectives pour les verbes et les adjectifs. Aux problèmes liés à la description des noms dans l'EWN (Vossen, 1996 : 715—728) les solutions de Park (2005) étaient aussi d'ajouter les registres de langue, Dorr, Martí et Castellón (1998 : 393—397) proposent de coupler l'EWN avec les structures lexicales conceptuelles (Lexical Conceptual Structure) pour parer au “matching problem”, c'est-à-dire à la situation où un synset de la langue de départ ne trouve pas son synset équivalent en langue cible.

5. Conclusion

Dans les deux parties de notre propos traitant de l'efficacité de l'AOO et de l'EWN pour la TA, nous avons démontré qu'une approche binaire où la traduction en langue cible est inhérente à la description de la langue de départ (comme l'AOO) correspond mieux, vu les capacités des ordinateurs d'aujourd'hui, aux exigences de la TA qu'un projet de soudage des bases lexicales monolingues (comme l'EWN).

Une solution possible, ne serait-elle pas la création des WordNet-bilingues où seuls deux WordNets (d'une langue de départ et d'une langue cible) serait mis en corrélation ? Cette voie qui paraît juste et unique, semble déjà être empruntée par William J. Black et Sabri El-Kateb (2004) ou par Luisa Bentivogli MultiWordNet — le WordNet Italien et anglais (<http://multiwordnet.fbk.eu/english/home.php>)

Références

- Artale A., Goy A., Magnini B., Pianta E., Strapparava C., 1998: "Copying with WordNet Sense Proliferation". In: *Proceedings of the First International Conference on Language and Evaluation*. Granada, Spain 28—30, <http://multiwordnet.fbk.eu/paper/wordnet-granada.pdf> (accessible : mai 2012).
- Banyś W., 2002a: « Bases de données lexicales électroniques — une approche orientée objets. Partie I : Question de modularité ». *Neophilologica*, **15**, 7—29.
- Banyś W., 2002b: « Bases de données lexicales électroniques — une approche orientée objets. Partie II : Question de description ». *Neophilologica*, **15**, 206—249.
- Banyś W., 2005 : « Désambiguïsation des sens des mots et représentation lexicale du monde ». *Neophilologica*, **17**, 57—77.
- Black W.J., El-Kateb S., 2004: "A prototype English-Arabic Dictionary Based on WordNet". In: *Proceedings of 2nd Global WordNet Conference, Czech Republic, January 20—23*. Brno, 67—74.
- Borillo A., 1996 : « Exploration automatisée de textes de spécialité : repérage et identification de la relation lexicale d'hyponymie ». *LINX*, **34—35**, 113—124.
- Chandrasekaran B., Josephson J.R., Benjamins V.R., 1999: "What Are Ontologies, and Why Do We Need Them?". *IEEE Intelligent Systems*, **14**, 20—26.
- Cruse D.A., 1986: *Lexical Semantics*. Cambridge: Cambridge University Press.
- Díez-Orzas P., Cuypers I., 1995: *The Novell ConceptNet*. Internal Report, Novell Belgium NV.
- Dorr B.J., Martí A., Castellón I., 1998: "Evaluation of EuroWordNet and LCS-Based Lexical Resources for Machine Translation". In: *Proceedings of the First International Conference on Language Resources and Evaluation*. Granada, Spain, 393—397.

- Fellbaum Ch., 1990: "English Verbs as a Semantic Net". *International Journal of Lexicography*, **3/4**: 278—301.
- Fellbaum Ch., ed., 1998: *WordNet: An Electronic Lexical Database*. Cambridge, Mass., London: The MIT Press.
- Firth J.R., 1951: "Modes of Meaning". In: *Papers in Linguistics 1934—1951*. London: Oxford University Press.
- Grębosz J., 2003: *Symfonia C++*. Kraków: Oficyna Kallimach.
- Gross D., Miller K.J., 1990: "Adjectives in WordNet". *International Journal of Lexicography*, **3/4**: 265—277.
- Gross G., 1992: «Forme d'un dictionnaire électronique». In: *L'environnement traductionnel, actes du colloque de Mons, 25—27 avril 1991*. Sillery (Canada): Presses de l'Université du Québec, 255—271.
- Gross G., 1994a: «Classes d'objets et synonymie». *Annales Littéraires. Supports, opérateurs, durées*, 93—102.
- Gross G., 1994b: «Classes d'objets et descriptions des verbes». *Langages*, **115**, 15—30.
- Gross G., 1995: «Une sémantique nouvelle pour la traduction automatique — les classes d'objets». *La Tribune des industries de la langue et de l'information électronique*, **17—18—19**, 16—19.
- Gross G., 1996: «Rendre les dictionnaires plus actifs». In: *Autour de l'informatisation du 'Trésor de la langue française'. Colloque international de Nancy, 29—31 mai 1995*. Nancy, 195—212.
- Gross G., 1998: «Pour une véritable fonction "synonymie" dans un traitement de texte». *Langages*, **131**, 103—114.
- Halliday M.A.K., Hasan R., 1976: *Cohesion in English*. London: Longman.
- Ide N., Véronis J., 1998: "Word Sense Disambiguation: The State of the Art". *Computational Linguistics*, **24 (1)**, 1—40.
- Miller G.A., 1990: "Nouns in WordNet: A Lexical Inheritance System". *International Journal of Lexicography*, **3/4**: 245—264.
- Miller G.A., 1998: *Foreword*. In: Ch. Fellbaum, ed.: *WordNet: An Electronic Lexical Database*. Cambridge, Mass., London: The MIT Press.
- Miller K.J., 1998: "Modifiers In WordNet". In: Ch. Fellbaum, ed.: *WordNet: An Electronic Lexical Database*. Cambridge, Mass., London: The MIT Press, 47—68.
- Park S.H., 2004: "English-French-Canadian French-Korean multilingual electronic database of job titles." *Industrial Electronics Society 2004, IECON 2004. 30th Annual Conference of IEEE*, **3**, 2608—2613.
- Sowa J.F., 2000: *Knowledge representation: logical, philosophical, and computational foundations*. Brooks/Cole Publishing Co., Pacific Grove, CA.
- Vossen P., 1996: "Right or Wrong: Combining lexical resources in the EuroWordNet project". In: M. Gellerstam, J. Jorg, S. Malmghren, K. Noren, L. Rogstrom, C.R. Papehl: *Proceedings of Euralex-96*. Goetheborg, 715—728.
- Vossen P., Díez-Orzas P., Peters W., 1997: "Multilingual Design of EuroWordNet". In: P. Vossen, N. Calzolaris, G. Adriaens, A. Sanfilippo, Y. Wilks, eds.: *Proceedings of the ACL/EACL-97 Workshop on Automatic Information Extraction and Building of Lexical Semantic Resources for Natural Language Processing Applications, Madrid, July 1997*, 1—8.

Vossen P., Diez-Orzas P., Peters W., 1997: "The multilingual design of the EuroWordNet database". IJCAI-97: Workshop on Ontologie and Multilingual NLP, Nagoya, Japan, August 23, 1997. On-line: <http://crl.nmsu.edu/Events/IJCAI/vossen.rtf.gz> (accessible : 09.02.2000).

Wittgenstein L., 1953: *Philosophical Investigations*. Oxford: Blackwell.

Dictionnaires

Słownik języka polskiego PWN, 2012. Warszawa: PWN.

Słownik współczesnego języka polskiego, 1999. Red. B. Dunaj. Warszawa: Wilga, 300—301.

LeDiCoal'Observatoire de Linguistique Sens-Text, <http://idefix.ling.umontreal.ca/recherche/linguistique/dico/lang-pref/fr/> (accessible : mai 2012).

Trésor de la Langue française informatisé, 2010, <http://atilf.atilf.fr/tlf.htm> (accessible : mai 2012).

English-Polish Dictionary. Red. J. Linde-Usienkiewicz. Warszawa: PWN—Oxford.