

Sonia Szramek-Karcz
Université de Silésie
Katowice

L'Approche Orientée Objets ou l'EuroWordNet s'adapte mieux à la traduction automatique ? Partie I : origines, principes, organisation des données

Abstract

Lexical databases such as AOO or EWN are built in order to enable machine translation of written word. The author's intention is to test the applicability of the above listed databases in machine translation and to choose the one that proves to be more successful.

The presentation of final results consists of two parts. Part I is the analysis of factors influencing the process of machine translation with the use of AOO and EWN that is to say: steps in database design, their theoretical aspects and the categorization of lexical items. In author's belief, the above mentioned elements exert profound influence on the result of machine translation produced with the use of one of the herein described lexical databases. The second part of the presentation touches on the matters of hierarchy, semantic inheritance and word-sense disambiguation.

Keywords

Object Oriented Approach (AOO), EuroWordNet (EWN), machine translation, lexical databases, ontology.

Introduction

Après avoir présenté la description de la classe d'objets de professions dans l'Approche Orientée Objets (S. Szramek-Karcz, 2006), analysé la description des noms dans le WordNet et comparé l'Approche Orientée Objets et l'EuroWordNet (S. Szramek-Karcz, 2011), le temps est venu d'évaluer l'utilité de l'Approche Orientée Objets et de l'EuroWordNet pour la traduction automatique (désormais la TA), une traduction effectuée par une machine sans la moindre intervention d'un traducteur humain.

L'Approche Orientée Objets (désormais l'AOO) est une base de données lexicales créée par Wiesław Banyś (W. Banyś, 2002a, 2002b) à l'Université de Silésie. L'EuroWordNet est un projet géré par Piek Vossen de l'université d'Amsterdam et vise la connexion des WordNets nationaux (C. Salvador, H. Rodriguez, J. Gonzalo, 1996 ; P. Vossen, 1996 ; P. Vossen, 1999 ; P. Vossen, P. Díez-Orzas, W. Peters, 1997 ; P. Vossen, 2002 ; P. Jansen, 2004). Les deux bases de données lexicales se veulent utiles pour la TA.

Les travaux sur la TA sont menés depuis les années 1930 (A.-M. Löffler-Laurian, 1996). Après l'engouement arrive un désenchantement et l'option de la Traduction assistée par ordinateur (TAO) prend la place de la TA. Mais le rêve ne s'est pas estompé, les avantages de la TA sont innombrables car elle permet d'échanger des informations, de progresser dans les disciplines scientifiques tout en gardant l'identité culturelle de l'auteur (les priorités de la Communauté Européenne) qui pourrait rédiger dans sa langue maternelle sans risque de perdre des lecteurs de langue cible.

Or, force est de constater que la TA demeure toujours imparfaite du fait qu'elle demande une intervention humaine et l'ambiguïté des langues reste un obstacle difficile à surmonter (N. Ide, J. Véronis, 1998 ; K. Kakaes, 2012). Les bases de données lexicales fleurissent (WordNet, FrameNet, LLI XIII, Le DiCo à l'Observatoire de Linguistique Sens-Texte, l'EWN, l'AOO) et les théories se multiplient. Dès lors la traduction automatique n'est qu'un usage possible de ces bases avec la navigation textuelle, les résumés automatiques, l'analyse du contenu d'un texte, les analyses grammaticales, etc. Dans le présent propos il s'agit de voir laquelle des deux bases de données lexicales, AOO ou EWN, convient mieux aux exigences de la traduction automatique, autrement dit, **il s'agit de voir laquelle garantit le choix adéquat de machine et par cela la traduction correcte**. Nous allons approcher cette interrogation en deux parties composées respectivement des problèmes qui serviront de base à notre argumentation, dans la première partie nous aborderons les origines et les buts de la création, les principes de construction de ces bases et l'organisation des données pour passer dans la deuxième partie à la hiérarchie des données, l'héritage sémantique et la désambiguïsation des contenus des bases.

Ayant en vue la clarté de l'exposé, le lecteur ne trouvera pas ici de description détaillée, minutieuse et exhaustive de l'AOO et de l'EWN mais juste certaines esquisses qui se sont avérées nécessaires à notre argumentation. Comme l'EWN est un réseau des WordNets nationaux et base sur le WordNet américain, nous allons souvent nous référer au WordNet (<http://wordnet.princeton.edu/>) sans oublier que c'est l'EWN qui est le sujet de notre évaluation.

L'AOO telle qu'elle est appliquée actuellement dans le Département de Linguistique Appliquée et de Traduction de l'Institut des Langues Romanes et de la Traduction à l'Université de Silésie de Katowice est bien récente que le WordNet qui remonte aux années 1980 (Chr. Fellbaum, 1998). Le fabuleux projet de WordNet (désormais le WN) avec les moyens engagés ainsi que l'EWN avec tou-

tes les langues qu'il englobe, sont des projets d'envergure incomparable à l'AOO. C'est pourquoi notre évaluation concerne **les résultats obtenus et possibles grâce à l'application de l'AOO et de l'EWN à la TA.**

1. Les origines et les buts de la création

Les tâches assignées à la construction des bases de données lexicales l'AOO et le WN étaient différentes.

L'AOO, dès le début, a été conçue pour servir à la TA avec son architecture simple, flexible et ouverte aux modifications, aisément réutilisable et enrichie dans l'avenir sans que cela nuise à son intégrité. À part le lien dialectique entre les unités (W. Banyś, 1988; B. Śmigielska, A. Grigowicz, 2004) vu à la surface et rendu par les opérateurs et les attributs de l'objet en question, l'approche prend en compte aussi les informations sémantiques comme celles contenues dans la relation partie—tout, des domaines ou l'héritage sémantiques montré dans les hiérarchies des classes. Il s'agit de présenter les informations implicites et explicites dans une structure flexible au point de les rendre utiles et réutilisables, aptes aux modifications à effectuer dans un champs donné sans détruire la construction de la base de données, ni fausser la toile des relations et des dépendances des unités qui la construisent. L'AOO est une base de données lexicales polyvalente qui réunit les informations des dictionnaires existant, les enrichit, tire des informations implicites et montre les hiérarchies de dépendances sémantiques.

Dans la construction du WN la TA n'était pas envisagée et elle n'est apparue qu'avec le projet de l'EWN ce qui paraît la cause de tous les problèmes de TA du projet EWN que nous allons voir ci-dessous. Quant au WordNet (le projet l'EWN base sur le WN américain et se compose de plusieurs WordNets), nous devons nous souvenir que cette base électronique de données lexicales organisée en un ensemble de réseaux sémantiques n'est plus aujourd'hui ce qu'elle était en 1985, date officielle de sa création car à l'origine, l'idée n'était pas de construire un lexique complet mais d'identifier les principaux nœuds lexicaux et d'explorer les *patterns* (différentes sortes) de relations sémantiques qu'ils entretiennent. La théorie testée stipulait que s'il y avait un schéma correct des relations lexicales, la définition en découlerait. Il semblait alors redondant d'introduire les définitions dans le réseau des relations sémantiques du WordNet. D'un simple moteur de recherche, le WordNet est alors devenu une base lexicale autonome à laquelle les définitions ont été ajoutées (des projets liés au WN comme Mimida Project, Open Multilingual WordNet (<http://wordnet.princeton.edu/wordnet/related-projects/>) ou inspiré du WordNet comme ImageNet (<http://www.image-net.org/>) sont en cours). Depuis, le WordNet ne cesse de se développer, sa version 3.0 contenait environ 80 000 syn-

sets des noms, c'est-à-dire des ensembles de synonymes, en 2013 la version 3.1 est disponible.

L'EuroWordNet a été un projet de 36 mois financé par la Commission Européenne (P. Vossen, P. Díez-Orzas, W. Peters, 1997 ; EuroWordNet 2001, <http://www.illc.uva.nl/uroWordNet>) qui envisageait de créer des WordNets contenant des relations sémantiques de base pour plusieurs langues européennes (anglais, hollandais, espagnol et italien et en 1998 allemand, suédois, français, tchèque, estonien) dans le but de créer une base de données lexicales multilingue. Même si le financement du projet européen est terminé, aujourd'hui les WordNets nationaux sont préparés pour une cinquantaine de langues dont la liste mise à jour par Global WordNet Organization se trouve à l'adresse : http://www.globalwordnet.org/gwa/wordnet_table.htm.

L'idée de connecter les langues afin de permettre une traduction des termes d'une langue à l'autre est née avec l'EWN (cf. Inter-Lingual-Index dans la deuxième partie de nos propos) ce qui ne reste pas sans conséquence pour les résultats de la TA obtenus avec ce projet. Comme nous l'avons montré, les deux bases lexicales, à savoir l'AOO et l'EWN, n'ont pas été créées dans le même but et la TA n'est qu'une possibilité d'exploitation de cette impressionnante base de données que constitue l'EWN. L'AOO est une base de données destinée à la TA. Dans ce qui suit, nous allons voir les conséquences de ce fait dans les autres choix que les deux bases imposent.

2. Les principes de construction

De prime abord, les principes de construction des deux bases lexicales semblent relativement proches : la clarté, la cohérence, l'extensibilité, les modules etc. Nous allons regarder les détails qui font toute la différence car comme le dit l'adage : le diable est dans les détails.

En effet, l'AOO part du principe que les données lexicales devraient être décrites de manière à permettre leur réutilisation facile et adaptation dans un autre projet qui, vu le progrès prodigieux des sciences informatiques, peut voir le jour très bientôt. Le mot d'ordre est l'architecture modulaire et qui dit architecture modulaire, dit des modules autonomes et organisés dans une structure cohérente (cf. B. Meyer, 1988 ; J.A. Fodor, 1981). Une démarche modulaire respecte les concepts de décomposabilité, composabilité, compréhensibilité et continuité postulés dans l'AOO que nous avons montré et analysé à l'exemple du module *travailler comme vendeur à domicile* (S. Szramek-Karcz, 2006). Le principe de la continuité modulaire réduit l'impact des changements dans la solution des problèmes à un minimum de modules liés entre eux, et mieux à un seul module (M. Meyer, 1988 : 16) ainsi, si nous intervertissons les modules de *vanter ses produits* avec

faire les démonstrations le changement n'affectera que le module de *présenter les produits* (cf. S. Szramek-Karcz, 2006). Le critère de continuité répond à l'un des objectifs poursuivis généralement au cours de la fabrication des bases de données à savoir la recherche de l'exhaustivité. On sait bien que le caractère évolutif de la langue rend cet objectif inaccessible, il faut donc voir dans cette préoccupation un souci de mise à jour régulière.

L'architecture modulaire de l'AOO permet de satisfaire aux postulats susmentionnés comme la souplesse de la description, son ouverture aux modifications, etc. Nous retrouvons les points communs de l'approche de la description lexico-sémantique orientée objets et de la programmation à objets en informatique (B. Stroustrup, 2002 ; B. Meyer, 1988 ; B. Eckel, 2002), ainsi que de la construction des bases des données.

Les modules ordonnés temporellement, leur ordre doit rester observé, forment un script (R.C. Shank, R.P. Abelson, 1977), chacun d'eux étant un cadre (C.J. Fillmore, B.T. Atkins, 1994) se compose de sous-cadres. Nous le retrouvons dans les propos de Marvin Minsky (1974) : "If the new room is unfamiliar, no pre-assembled frame can supply fine details; more scene-analysis is needed. Even so, the complexity of the work can be reduced, given suitable subframes for constructing hypotheses about substructures in the scene". Les frames dans des bases de données lexicales destinées à la TA sont incontournables (cf. la deuxième partie de notre présentation).

Les relations sémantiques dans le WordNet de Christiane Fellbaum (1998 : 5) connectent d'une façon naturelle les mots et les concepts des domaines sémantiques communs. Dans son article "Lexicography in an Interlingual Ontology: An introduction to EuroWordNet", Peter Jansen (2004) explique que l'EWN ne contient pas des unités plus petites qu'un mot ni les informations plus larges comme cadre (frames de M. Minsky, 1974, 1981, 1986), scripts (R.C. Shank, R.P. Abelson, 1997) ou «schemata» (D.E. Rumelhart, 1980). Le mot «modulaire» apparaît dans un autre contexte. La structure modulaire de l'EWN (P. Vossen, 1996 : 1) peut être résumée comme suit : les modules de langue qui contiennent le «lexicon» conceptuel de chacune des langues, le module indépendant de la langue qui comprend Inter-Lingual-Index (ILI), domaine ontologique et l'ontologie Top-Concept. Les modules dans l'AOO sont en deux langues, c'est une approche binaire dont la description vise la langue de départ et la langue d'arrivée ce qui entraîne la traduction immédiate de chacun des modules. Dans l'EWN, la traduction n'est qu'un résultat de l'équivalence au niveau de l'ILI. Les modules sont monolingues et retrouvent leur équivalent en langue cible, grâce à un réseau de relations d'équivalence qui les connecte.

L'EWN est régi par des principes exigés dans la construction de bases de données lexicales ontologiques (P. Jansen, 2004 : 3 ; T.R. Gruber, 1995) comme : la clarté, la cohérence, l'extensibilité, minimum de panchement des codes (en anglais : *minimal encoding bias*), minimum de caractéristiques ontologiques (en

anglais : *minimal ontological commitment*). Selon le principe de la clarté, les définitions dans une base de données ontologique devraient être objectives et indépendantes du contexte social et computationnel. Dans le WN les définitions ont été ajoutées plus tard (ultérieurement) et comme dans l'AOO, elles sont introduites à titre indicatif (dans l'AOO l'objet est défini par les opérations et les attributs et dans le WordNet (ou EWN), par sa place dans la hiérarchie des concepts). La définition comparable à celle des dictionnaires classiques destinées aux lecteurs humains ne joue aucun rôle dans la TA et se trouve dans les deux bases de données AOO et EWN uniquement pour faciliter le travail des lexicographes, ainsi que pour agrandir les possibilités d'utilisation de ses bases autrement dit à d'autres fins que la TA.

Le principe de cohérence stipule que l'ontologie doit être cohérente et les définitions correspondre aux exemples de la langue naturelle. Conformément à ce principe, le WN et par conséquent l'EWN, sont munis de définitions et d'exemples. Ces bases de données (WN et EWN) ont des possibilités multiples d'exploitation et la TA n'est qu'une d'entre elles. Du point de vue de la TA, les définitions du style du dictionnaire traditionnel et les exemples sont sans importance.

Comme la langue naturelle vit, la base de données lexicales doit (comme cela est postulé par l'AOO et l'EWN) être apte aux modifications. Des mots nouveaux devraient donc être incorporés sans nuire à l'intégralité du système. Essayons d'intégrer un mot nouveau comme **cordiste** (anglais : *rope worker*, polonais : *pracownik na linach*) dans l'AOO et l'EWN pour tester leur capacité de mise à jour régulière. Dans l'AOO, le « cordiste » est un objet de la classe d'objets <professions> (S. Szramek-Karcz, 2006b) inséré dans la base sans le moindre problème, décrit selon les principes de l'héritage sémantique présenté dans la deuxième partie de nos propos. Pour introduire dans la base ontologique un « cordiste » il faudrait, arbitrairement, trancher s'il est un *travailleur* ou un *professionnel* car dans le WN, nous retrouvons les éléments de la classe d'objets <professions> parmi les hyponymes des synsets travailleurs (**worker**) et professionnels (**professional**) :

- ◀ S: (n) **worker** (a person who works at a specific occupation) "*he is a good worker*"
 - ◀ direct hypernym / inherited hypernym / sister term
 - ◀ 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))

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) 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))

La seule différence entre ces deux classements est l’hyperonyme : S: (n) adult, grownup (a fully developed person from maturity onward), ce qui nous permet de comprendre que le travailleur (*worker*) n’est pas forcément adulte comme un professionnel (*professional*).

Le fait d’être adulte pour exercer une profession (ou une occupation), ce qui paraît juste, est absent dans la classification du WordNet polonais (<http://plwordnet.pwr.wroc.pl/main/?lang=pl>). Comme il s’agit d’un trait ontologique, il n’est alors pas pris en considération dans le classement linguistique de l’AOO. Toujours dans l’esprit ontologique, le WN divise les activités en deux sortes d’occupations, à savoir celles qui exigent l’apprentissage approfondi (*learned professions*) et celles

qu'il suffit d'effectuer : *specific occupation*, des fois on ajoute *skilled professions* pour montrer qu'une telle ou telle profession exige un certain savoir faire. Pour intégrer notre **cordiste** dans le système, il faudrait décider s'il est *professional* ou juste *worker*. Sa définition de *personne qui travaille sur chantier d'accès difficiles*, nous fait choisir la deuxième catégorie où on ajouterait les hyponymes comme : *cordiste-naceliste*, *couvreur cordiste*, *couvreur-zingueur cordiste*, **maçon** *cordiste*, **ouvrier** *cordiste*, **technicien** *cordiste*, il faudrait confronter avec la classification effectuée pour : *naceliste*, *couvreur*, *couvreur-zingueur*, **maçon**, **ouvrier** et **technicien**. L'extension de cette base de données est possible, même si certains choix pourraient être discutables. Dans le EWN qui servirait à la TA, il serait souhaitable que le mot nouveau soit aussi pris en compte dans tous les WordNets qui le construisent. Dans l'AOO, le mot «cordiste», pour pouvoir être intégrée dans la base, doit recevoir son équivalent en langue cible par exemple : dans la description français-polonais : [FR] *cordiste* — [PL] *pracownik na linach* ; dans la description polonais-français : [PL] *pracownik na linach* — [FR] *cordiste* ; et pour les autres paires de langues : [FR] *cordiste* — [EN] *rope worker* ; [EN] *rope worker* — [FR] *cordiste* ; [PL] *pracownik na linach* — [EN] *rope worker* ; [EN] *rope worker* — [PL] *pracownik na linach*. Pour chaque paire et chaque direction de la traduction, il y a une base à part, souvent la relation n'est pas si simple. Par exemple, si nous partons de l'anglais vers le polonais, le nom anglais *nurse* a trois traductions distinctes (et trois classe d'objets distinctes) en polonais, celles de *pielęgniarka*, *niania*, *opiekunka* et si nous partons du mot polonais *pielęgniarka*, la traduction possible vers l'anglais n'est que *nurse*.

Le fait de choisir l'ontologie impose encore deux principes absents à l'AOO : minimum de panchement des codes et minimum de caractéristiques ontologiques. Selon ce principe, la conceptualisation devrait s'effectuer au niveau du savoir qui ne dépend pas des codages symboliques admis à un niveau de codage symbolique. Dans l'EWN, le postulat est réalisé par *common top-level ontology* où se trouvent les concepts communs pour tous les WordNets nationaux. Les problèmes pourraient arriver au moment de la traduction vers des *langues extra-européenne* (par exemple *langues africaines*), on s'imagine que dans ce cas les notions de bases peuvent varier considérablement.

Le principe de minimum de caractéristiques ontologiques semble vouloir faire tourner la description ontologique du vocabulaire vers la description linguistique où seul l'emploi d'un terme décide de sa définition et de sa place dans la hiérarchie de la base de données. Pour l'illustrer, regardons le nom : FR *prostitué*. Dans le WordNet 3.0, ses hyperonymes sont respectivement: S: (n) woman, adult female (an adult female person (as opposed to a man)), S: (n) female, female person (a person who belongs to the sex that can have babies), S: (n) person, individual, someone, somebody, mortal, soul (a human being), S: (n) organism, being (a living thing that has (or can develop) the ability to act or function independently)>. Le WordNet polonais également ne marque pas de synset qui préciserait quelconque occupa-

tion. Par contre dans le Multilingual Central Repository qui est une compilation (entre autres) de plusieurs WordNets (J. Atserias, L. Villarejo, G. Rigau, E. Agirre, J. Carroll, B. Magnini, P. Vossen, 2004) apparaissent les niveaux de *occupationalRole* et *Function*, ce qui rapproche ce nom de la classification linguistique effectuée dans l'AOO où la langue range ce mot dans la classe d'objets de <professions> par le fait d'accepter les constructions : *travailler comme prostitué, gagner sa vie comme prostitué, exercer la profession de prostitué, exercer le métier de prostitué* (cf. le même classement effectué dans le dictionnaire électronique élaboré au LLI (Laboratoire de Linguistique Informatique) (D. Le Pesant, 2000, 5^e étude : 34), ainsi que la classe d'objets de professions de P.-A. Buvet, P.-Y. Foucou (2000).

Le principe linguistique (l'AOO) et non ontologique (EWN) permet d'effectuer les classements des éléments lexicaux de façon impartiale, les convictions des chercheurs ou des lexicographes n'entrent pas en jeu. Les éléments lexicaux classés différemment dans les WordNets peuvent présenter le problème de leur juxtaposition correcte appelé « matching problem ».

3. L'organisation des données lexicales

Les résultats possibles de la TA sont également faciles à prévoir si on analyse la façon d'organiser les données lexicales. Le lexique dans des bases s'articule autour des unités de description : les objets (AOO) et les synsets (EWN).

Avant de passer à la définition des objets et des synsets, ajoutons que dans l'optique orientée objets, la direction d'analyse est inversée par rapport à la description des sens des mots appliquée dans le cadre des structures prédicats-arguments où l'on part de la fonction propositionnelle (prédicat) pour arriver à ses arguments (objets) qui saturent les positions ouvertes par ces premiers (S. Karolak, 1972 ; K. Bogacki, S. Karolak, 1991 ; W. Banyś 1981, 1983, 1984) ; c'est-à-dire **dans l'AOO nous partons d'un objet** et cherchons les prédicats qui peuvent lui être assignés. Ces prédicats sont répartis en attributs (adjectifs et les constructions *N(Prép)N*) et en opérations (verbes) qui pour des raisons d'organisation descriptive des objets, se trouvent divisés en opérateurs **constructeurs, manipulateurs et accesseurs**. Les constructeurs comme leurs noms l'indique, construisent la classe d'objets en question ou construisent la situation où la classe d'objets n'apparaît pas, les accesseurs fournissent les informations sur le comportement et la structure de la classe, les manipulateurs par contre effectuent toutes sortes d'opérations sur la classe d'objets ou que la classe d'objets peut effectuer. La description dans les deux sens est importante, mais il fallait bien choisir une des options possibles et ce sont les objets qui ont été choisis comme le pivot autour duquel tournent (entre

autres) les opérateurs et les attributs. Les différences entre l'opérateur approprié / hérité vs l'opérateur approprié / général sont soulignées : les *opérateurs appropriés* s'appliquent à une ou plusieurs classes en question comme *travailler, jardiner* etc. et les *opérateurs généraux* s'appliquent à un grand nombre de classes comme *regarder, observer, admirer* etc. Tous les opérateurs des super-classes sont hérités par des sous-classes. Nous allons revenir à la question d'héritage dans la deuxième partie de notre présentation.

Dans l'optique AOO, l'objet est défini par ses opérations et attributs, son statut est fonctionnel et ses caractéristiques ontologiques ne sont pas prises en compte (comme c'est le cas de la classification dans l'EuroWordNet). Dans l'AOO, l'objet est caractérisé par l'ensemble des opérations qu'il effectue ou qui sont effectuées sur lui ainsi que par des attributs qui lui sont assignés. Ces objets dans l'AOO sont regroupés en classes d'objets, ce qui veut dire que la classe d'objets est un ensemble d'objets qui partagent les mêmes opérations et (ou) les mêmes attributs. La notion se rapproche, sans être identique, des classes d'objets de G. Gross (G. Gross 1992, 1994a, 1994b, 1995a, 1995b ; D. Le Pesant, M. Mathieu-Colas, 1998 : 7—33). Une classe d'objets est définie par les opérations qu'elle admet et une même opération peut bien constituer un élément de faisceau définitionnel de classes d'objets différentes. Il est facile de s'imaginer que l'opérateur «travailler» va appartenir aux classes différentes suivant les exemples de son emploi ci-dessous :

machine travaille (COUFFIGNAL, *Mach. penser*, 1964, p. 61)

vin travaille (NADAUD, *Chansons*, 1870, p. 284)

argent travaille (BLOY, *Lieux communs*, 1902, p. 31)

goutte travaille (SANDEAU, *M^{lle} de La Seiglière*, 1848, p. 231)

Dans l'AOO on distingue autant de classes d'objets que d'ensembles d'opérations et d'attributs. Par exemple, dans le cas de <gendarme>, nous avons affaire à trois classes différentes :

gendarme

- personne chargée du maintien de l'ordre et de la sûreté publique <professions>
- épointement rocheux situé sur une arête, pouvant constituer un obstacle à la progression des alpinistes <obstacles>
- saucisse séchée et fumée, vendue par paire, de couleur brune dorée <aliments>

Les objets constituent une classe d'objets, mais la réalisation concrète d'un objet <chirurgien> comme par exemple Monsieur Duval est son instance, une instance d'objet. Nous retrouvons les instances des objets, parmi les hyponymes de tout en bas de la hiérarchie du EWN.

Les informations dans l'AOO sont présentées sous forme d'une fiche-entrée qui contient respectivement les informations sur la morphologie, la syntaxe et la

sémantique. La partie sémantique de la fiche descriptive contient les informations, toujours bilingues, sur la classe d'objet à laquelle l'objet décrit par la fiche appartient, sa définition, ses synonymes, ses super-classes, domaine, attributs, partie—tout, opérateurs : constructeur, manipulateur, accesseur, extensions et cadre (*frame*).

Les synonymes dans la fiche descriptive de l'AOO ont une fonction d'orientation et ne définissent pas l'unité lexicale (c'est le devoir des opérateurs et des attributs), dans la plupart des cas se sont des registres différents d'un mot comme : *médecin*, synonyme *toubib*. Par synonyme dans l'AOO on comprend les mots interchangeables dans tous les contextes, dans l'EWN les synonymes sont des mots interchangeables dans presque tous les contextes, ce qui est une différence considérable pour la TA. Le WN a servi à détecter les relations entre les mots et non à retrouver les équivalents dans les autres langues. Le fait de laisser les synsets dans le projet de l'EWN ne facilite pas du tout la tâche de la TA, la description des noms dans l'EWN (comme celle de WN) s'articule autour des synsets. Les synsets (S:) (le nom vient de *synonym set*) sont des ensembles de synonymes et constituent le noyau de la construction de l'EuroWordNet.

Précisons et soulignons encore une fois que dans l'EWN on traite des synonymes des mots qui peuvent s'interchanger dans certains (mais pas tous les) contextes. (La notion de synset se rapproche, sans être identique, de *run* introduit par J.K. Spark (1986), d'après G.A. Miller (1998 : 24)). Dans le projet de l'EWN les types des relations internes à la langue ont été largement agrandis, des équivalents complexes ont été introduits comme par exemple : *eq_near_synonyme*, *has_eq_hyperonym* ou *has-Eq_hyponym*, car l'EWN doit faire face aux problèmes de la connexion de plusieurs WNs. Cette connexion s'effectue par intermédiaire de l'Inter-Lingual-Index. Inter-Lingual-Index est une liste des synsets du WN anglais version 1.5 qui est agrandie si le besoin se présente (P. Vossen, L. Bloksma, W. Peters, C. Kunze, A. Wagner, K. Pala, K. Vider, F. Bertagna, 1999). Ainsi chacun des synsets du WordNet monolingue retrouve au moins un équivalent parmi les synsets du Inter-Lingual-Index. Nous y voyons le plus grand obstacle à la TA : Inter-Lingual-Index — une liste des concepts anglais comme qu'un point de repère pour les autres WordNets nationaux. Les *near_synonymes* et autres relations de gradation ajoutées compliquent la description et éloignent l'EWN de l'application à la TA. Le risque de l'erreur dans la classification comme synonymes proches ou plus lointains sur l'axe de la synonymie est élevé, les classifications peuvent varier d'une personne à l'autre. Le critère linguistique de l'AOO paraît plus fiable car c'est la langue qui décide du classement des unités de description.

Mais avant de retrouver sa place dans le listing des synsets de l'Inter-Lingual-Index, les données des WordNets, ainsi que les objets dans l'AOO, sont structurées dans une hiérarchies avec, ce qui est encore plus intéressant pour nos propos et le problème débattu, la possibilité d'héritage sémantique y compris l'héritage multiple. Dans la deuxième partie nous allons analyser l'efficacité de l'AOO et de

l'EWN pour la TA en examinant leur hiérarchie, l'héritage et la désambiguïsation des unités lexicales de description.

Références

- Atserias J., Villarejo L., Rigau G., Agirre E., Carroll J., Magnini B., Vossen P., 2004: "The MEANING multilingual central repository". In: *Proceedings of the Second International Global WordNet Conference (GWC'04)*. Brno.
- Banyś W., 1981: « Description indéfinies : arguments ou prédicats en position d'argument ? » *Linguistica Silesiana*, 4.
- Banyś W., 1983: *Ambiguïté référentielle des phrases à descriptions indéfinies en français*. Katowice, Wydawnictwo UŚ.
- Banyś W., 1984: « Sémantique, structure, syntaxe et lexique ». *Cahier de Lexicologie*, 45.
- Banyś W., 1988: « Sur le dictum thématique : articulations secondaires du rhème ». In: Banyś W., Karolak S., eds.: *Structure thème-rhème dans les langues romanes et slaves*. Wrocław, Ossolineum.
- Banyś W., 1998: « Dictionnaires électroniques et conception 'modifié-modifieur' » (mimeograph).
- Banyś W., 2002a: « Bases de données lexicales électroniques — une approche orientée objet. Partie I : Questions de modularité ». *Neophilologica*, 15, 7—28.
- Banyś W., 2002b: « Bases de données lexicales électroniques — une approche orientée objet. Partie II : Questions de description ». *Neophilologica*, 15, 206—248.
- Bogacki K., Karolak S., 1991: « Fondement d'une grammaire à base sémantique ». *Lingua e Stile*, 26, 3.
- Buvet P.-A., Foucou P.-Y., 2000: « Extraction automatique de noms de profession sur le web ». *Linguisticae Investigationes*, 23 (2), 219—228 (http://www-lli.univ-paris13.fr/membres/biblio/956_textel5.rtf).
- Eckel B., 2002: „Wprowadzenie do obiektów”. In: *Thinking in C++*. Warszawa, HELION, 25—66.
- Fellbaum Chr., ed., 1998: *WordNet: An Electronic Lexical Database*. Cambridge, Mass., London: The MIT Press.
- Fillmore C.J., Atkins B.T., 1994: „Starting where the dictionaries stop: The challenge for computational lexicography”. In: B.T.S. Atkins, A. Zampolli, eds.: *Computational Approaches to the Lexicon*. Clarendon Press.
- Fodor J.A., 1981: *Representations*. Cambridge MA: MIT Presse.
- 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., 1995a: « A propos de la notion d'humain ». *Linguisticae Investigationes Suplementa*, 17: *Lexiques-grammaires comparés en français : actes du Colloque international de Montréal (3—5 juin 1992)*. John Benjamins Publishing Company, 71—80.

- Gross G., 1995b : « Une sémantique nouvelle pour la traduction automatique — les classes d'objets ». *La Tribune des industries de la langue et de l'information électronique*, 17—19, 16—19.
- Gruber T.R., 1995 : “Toward principles for the design of ontologies used for knowledge sharing”. Presented at the Padua workshop on Formal Ontology, March 1993, later published in *International Journal of Human-Computer Studies*, 43, 4—5, Novembre, 907—928.
- Ide N., Véronis J., 1998 : “Word Sense Disambiguation: The State of the Art”. *Computational Linguistics*, 24(1), 1—40.
- Jansen P., 2004 : “Lexicography in an Interlingual Ontology: An Introduction to EuroWordNet”. *Canadian Undergraduate Journal of Cognitive Science*, 1.
- Kakaes K., 2012 : “Why Computers Still Can't Translate Languages Automatically”, future tense. On-line: http://www.slate.com/articles/technology/future_tense/2012/05/darpa_s_transtac_bolt_and_other_machine_translation_programs_search_for_meaning_.html (accessible: 10 février 2013).
- Karolak S., 1972: *Zagadnienia składni ogólnej*. Warszawa, PWN.
- Le Pesant D., 2000 : *Six études de sémantique lexicale sur les noms locatifs*. [Manuscrit présenté pour l'Habilitation à Diriger des Recherches, LLI, Université Paris 13].
- Le Pesant D., Mathieu-Colas M., 1998 : « Introduction aux classes d'objets ». *Langages*, 131, 6—32.
- Loffler-Laurian A.-M., 1996 : *La traduction automatique*. Villeneuve d'Ascq, Presses Universitaires du Septentrion.
- Meyer B., 1988: *ObjectOriented Software Construction*, Upper Saddle River (N.J.). Prentice Hall International Series in computer Science.
- Miller G.A., 1998: *Foreword to WordNet: An Electronic Lexical Database*. Cambridge, Mass., London: The MIT Press.
- Minsky M., 1974: “A Framework for Representing Knowledge”. *MIT-AI Laboratory Memo*, 306, June.
- Minsky M., 1981: « A Framework for Representing Knowledge ». In : J. Haugeland, ed. : *Mind Design*. Cambridge MA, MIT Press, 245—262.
- Minsky M., 1986: *The Society of Mind*. New York, Simon & Schuster (trad. fr. *La société de l'esprit*, 1988, Paris, InterÉditions).
- Rumelhart D.E., 1980: “Schemata: The Building Blocks of Cognition”. In: R.J. Spiro, B. Bruce, W.F. Brewer, eds.: *Theoretical Issues in Reading and Comprehension*. Hillsdale, NJ: Erlbaum.
- Salvador C., Rodriguez H., Gonzalo J., 1996: *Definition of the links and subsets for nouns of the EuroWordNet project*. EuroWordNet Project LE4003, Deliverable D005. Amsterdam, University of Amsterdam, 1—130.
- Shank R.C., Abelson R.P., 1977: *Scripts, Plans, Goals and Understanding. An Inquiry into Human Knowledge Structures*. Hillsdale, New Jersey, Lawrence Erlbaum Associates.
- Stroustrup B., 2002: *The C++ Programming Language*. Addison-Wesley Pub. Co.
- Szramek-Karcz S., 2006 : « Description des professions dans l'approche orientée objets ». In : K. Bogacki, A. Miatluk, eds. : *Proceedings of the International Conference : Semantic Relations in Language and Culture, Białystok 24—26 october 2005*. Białystok: Wydawnictwo Uniwersytetu w Białymstoku, 300—308.

- Szramek-Karcz S., 2011 : « La description des noms dans l'EuroWordNet et l'Approche Orientée Objets ». *Romanica Cracoviensia*, **11**, 415—421.
- Śmigielska B., Grigowicz A., 2004 : « Description lexicographique fondée sur la modification conceptuelle : conception 'modifié-modifieur' ». *Neophilologica*, **16**, 42—51.
- Vossen P., 1996 : "Right or Wrong: Combining lexical resources in the EuroWordNet project". In: M. Gellerstam, J. Jaborg, S. Malmghren, K. Noren, L. Rogstrom, C.R. Papmehl: *Proceedings of Euralex-96*. Goetheborg, 715—728.
- Vossen P., ed., 1999: *EuroWordNet: Multilingual database with lexical semantic networks*. Kluwer Academic Publishers.
- Vossen P., 2002 : « Piek 'WordNet, EuroWordNet and Global WordNet' ». *Revue française de linguistique appliquée* **1**(7), 27—38 (<http://www.cairn.info/revue-francaise-de-linguistique-appliquee-2002-1-page-27.htm>, accessible : 10 février 2013).
- 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, 1—8.
- Vossen P., Bloksma L., Peters W., Kunze C., Wagner A., Pala K., Vider K., Bertagna F., 1999: *Extending the Inter-Lingual-Index with new concepts EuroWordNet (LE2-4003) Deliverable 2D010*. August 2, University of Amsterdam.