

Aleksandra Żłobińska-Nowak

*Université de Silésie
Katowice*

Validité de la modélisation objet dans la langue et dans la traduction

Abstract

The aim of this paper is to discuss one of the methods applicable in description of the representation of the world in language, that is object modeling.

The author emphasizes the notion of the *Hierarchical Semantic Network* proposed by A.M. Collin and M.R. Quillian in the 1960's which constitutes an example of the way of representing knowledge in the memorization process. These networks will be described in parallel with a class of structures of the data easily applicable to computer processing.

This analysis is a starting point for elaborating on the notion of the class of objects. Its specificities, content and relations with others groups of the same type, the property of the semantic inheritance will show to what extent the notion is useful word sense disambiguation according to W. Banyś and also in the learning process.

Keywords

Modeling object language, hierarchical semantic network, class of objects, word sense disambiguation, learning process.

Dans la linguistique contemporaine les travaux sur le traitement automatique des langues constituent aujourd'hui le véritable défi pour les linguistes. Cette branche demande de leur part beaucoup de patience, de savoir-faire et tout ceci à la base d'une méthodologie précise qui permettrait d'adapter les données provenant de la langue aux capacités et à l'opérationnalité d'une machine.

Cette présentation s'appuiera sur une des méthodes de la représentation du monde dans la langue, à savoir, la modélisation objet. Nous allons analyser la conception des réseaux sémantiques formant une structure hiérarchique pour l'étude de la mémorisation proposée par A.M. Collin et M.R. Quillian dans les années soixante. Sera souligné, dans ce contexte-là, le rôle des réseaux constituant certains modes

de représentation de connaissances étant en même temps, une classe de structures de données représentables facilement dans un ordinateur.

Cette analyse constituera le point de départ à l'explication de la notion de classe d'objets, intimement appliquée actuellement dans le traitement automatique des langues.

Les spécificités des classes, leur contenu, les relations qu'elles entretiennent avec d'autres ensembles de même nature, la propriété de l'héritage sémantique permettront de montrer leur utilité non seulement dans la désambiguisation des sens des mots suivant les principes de l'approche orientée objets à la W. Banyś, mais aussi dans le processus de l'apprentissage.

On observe aujourd'hui un grand intérêt pour les langages à objets qui conduit à la certitude que les concepts d'objets prennent probablement une place de premier plan dans tous les domaines de l'informatique y compris la linguistique informatique. Tout ceci, en gros, grâce à l'identification aisée, quasi immédiate des objets informatiques avec leur contrepartie physique.

Cette vision de choses possède quelques origines que nous allons tenter de présenter dans ce court travail. D'un côté elle a été empruntée à la programmation informatique élaborée par Alan Kay dans les années soixante-dix, l'auteur de la programmation orientée objet (POO) ou programmation par objet, qui consistait en la définition et l'interaction de briques logicielles appelées objets. Dans cette acception l'objet représentait un concept, une idée ou toute entité du monde physique, il possédait une structure interne et un comportement, et il savait communiquer avec ses pairs. L'objectif était donc de représenter ces objets et leurs relations [http://fr.wikipedia.org/wiki/Programmation_orient%C3%A9e_objet].

Les systèmes d'exploitation, de bases de données, de méthodologies et de représentations des connaissances par objets sont actuellement largement disponibles pour les utilisateurs.

L'intérêt croissant pour une telle approche découle de la facilité à décrire les entités du monde réel et de les gérer en regroupant toutes les données et les procédures qui les manipulent dans une même unité sémantique qui est l'objet.

L'approche objet est née de l'idée d'une programmation structurée vu que les programmes utilisés devenaient de plus en plus complexes et volumineux. Par conséquent leur maintien et exploitation étaient difficiles, il était donc question d'établir un regroupement en fonction des caractéristiques communes.

La modélisation objet se concentre autour d'objets, d'entités ayant une existence matérielle (p. ex. *arbre, chat, livre*, etc.) ou bien virtuelle (p. ex. *Agence Nationale pour emploi, contrat à durée déterminée*, etc.). Un objet de par sa nature peut être caractérisé par :

- les attributs (propriétés, structure de l'objet), une sorte de données qui permettent de le décrire et fournissent des informations portant sur son état,

- son identité grâce à laquelle on peut le distinguer ou le comparer aux autres objets, qui découle naturellement de ses attributs et décide de sa place dans une classe d'objets,
- les opérations (appelées aussi méthodes, fonctions membres ou actions qui rendent compte du comportement de l'objet) qu'il est susceptible de réaliser et qui lui permettent d'agir sur les autres objets.

Une autre origine, un peu plus ancienne, qu'on peut indiquer dans cette modélisation à objet, sont les réseaux sémantiques dont l'invention est attribuée à deux psychologues A.M. Collin et M.R. Quillian. Ils ont proposé dans les années soixante une vision de la mémorisation et de la structuration de la mémoire appuyée sur des structures hiérarchiques suivant la règle principale selon laquelle *un X est un Y* (*X* est une sorte de *Y*). Un réseau dans cette acception est un graphe fini, étiqueté, avec les sommets appelés *nœuds* représentant des concepts et les arcs appelés *liens* qui renvoient aux relations binaires entre ses concepts. Cette approche insiste plus sur la hiérarchie arborescente des objets et sur les relations entre eux, ce qui est une source, à son tour, de la formation de leurs classes. Ce procédé permet de comprendre comment fonctionne la mémoire humaine et comment le locuteur peut procéder à certaines inférences ainsi que faire appel à des connaissances enregistrées dans sa mémoire.

Ce modèle souligne qu'il existe une sorte de pyramide dans le cerveau humain s'appuyant sur des notions détaillées (p. ex. une race de chiens comme *chihuahua*) sur lesquelles reposent les catégories plus généralisées (p. ex. *animaux*). Au milieu de cette pyramide se trouvent les représentations des catégories étant des notions de statut intermédiaire (p. ex. *chiens*, *oiseaux*, *poissons*). En gros, ce qui devient expliqué par F. Rastier se demandant comment les réseaux sémantiques pouvaient être utilisés pour représenter le contenu lexical (F. Rastier, 1987 : 84) :

- l'étiquette du nœud hiérarchiquement supérieur est un sémème,
- toutes les étiquettes des nœuds descendants représentent des composants de ce sémème.

Il est à noter que chaque sémème acquiert ici le statut d'interaction structurée p. ex. *animal* — *chien* — *Cachou* et n'étant pas seulement une collection de traits comme dans le cas des analyses différentielles.

Ainsi, chaque élément qui appartient à la catégorie supérieure possède-t-il les traits caractéristiques de tous les animaux, et, qui plus est, les traits qui ne sont propres qu'à lui.

On arrive en résultat à la constatation qu'il existe une économie de mémoire pour la représentation de nombreuses relations et pour la facilité dans la recherche d'informations nécessaires à certains raisonnements et inférences.

Il est intéressant de noter que la question de la connaissance apparaît pratiquement à chaque époque marquant la pensée linguistique (J.-P. Desclés, 1987), à signaler entre autres « l'arbre de Porphyre » de l'Antiquité ou les travaux de Leibnitz souvent cités dans le cadre de l'Intelligence Artificielle.

D. Delas (1978), dans le même esprit, remarque que certains attributs sont attachés à certains signes et permettent à ces signes-là de se combiner avec d'autres dont on pense qu'ils partagent les mêmes attributs et qui sont même susceptibles de les remplacer, le cas échéant.

Selon D. Delas qui analysait la place de la métaphore au sein de la grammaire générative, la forme la plus simple, binaire au stade primitif du développement de l'humanité se présentait sous la forme d'une grille conceptuelle basée sur des oppositions à caractère binaire du type : *concret/abstrait, animé/inanimé, statique/dynamique, totalité/partie, dense/diffus* etc.

Passons maintenant à l'objet lui-même et sa description du point de vue du type de modélisation où il constitue le centre.

Nous pouvons appeler objet un élément identifiable du monde soit concret soit abstrait. Il faut souligner que les notions d'objet et de classe sont interdépendantes. Chaque objet existe toujours comme instance d'une classe alors qu'une classe regroupe des objets portant les mêmes propriétés (attributs et opérations).

Ainsi *un bateau* peut être caractérisé par :

- nombre de membres d'équipage,
 - vitesse de croisière,
 - bateau-torpilleur,
 - bateau-transport,
 - bateau sous-marin etc.
- étant ses propriétés (attributs), ou
- arrimer,
 - arriver,
 - sortir du port etc.

étant ses opérations qui font preuve de son comportement.

Dans ce contexte nous devons toucher la question de classes d'objets qui réunissent les ensembles des objets partageant les propriétés semblables.

Les classes d'objets sont des classes à portée sémantique définies à l'aide des critères distributionnels. Elles ont servi étroitement à la confection des dictionnaires électroniques surtout dans les travaux du Laboratoire de linguistique informatique (LLI, actuellement Lexiques, Dictionnaires, Informatique — LDI) de l'Université Paris 13 sous la direction de G. Gross. Ces ensembles des mots homogènes sémantiquement et réunis en fonctions de l'organisation syntaxique des phrases dans lesquelles se trouvent leurs éléments constituent le point de départ non négligeable dans le traitement automatique des langues par le fait d'offrir des débouchés importants dans le domaine de la polysémie ou du figement.

Elles se sont avérées un remède là où les systèmes purement formels avaient montré leurs insuffisances dans le cas des structures linguistiques réduites à une représentation formelle, indépendantes du lexique et du sens, sans contenu.

La notion de classe d'objets se trouve au carrefour du lexique et de la grammaire. Chacune des classes construites est définie à partir des prédictats sélectionnant

de façon appropriée les unités qu'elle contient et tout ceci prenant en compte, bien entendu, leurs relations qui interviennent sur l'axe syntagmatique.

La nécessité d'avoir une vision à caractère plus global dans l'analyse des unités linguistiques a été déjà postulée sur le plan logique, p. ex. par G. Frege (1884 dans : D. Le Pesant, M. Mathieu-Colas, 1998 : 6—7) et son principe de contextualité selon lequel la signification des mots ne peut être dévoilée que dans le contexte d'une proposition d'où ils tirent leurs significations.

E. Benveniste souligne, dans la même lignée de pensée, que «le sens d'une unité linguistique se définit comme sa capacité d'intégrer une unité de niveau supérieur» (E. Benveniste, 1966 : 127 in : D. Le Pesant, M. Mathieu-Colas, 1998 : 7).

Les classes d'objets semblent découler naturellement de la répartition de la phrase en prédicat/arguments grâce à laquelle chaque phrase possède un noyau de nature prédicative complété par un ou des arguments nominaux. Cette perspective est plus opératoire qu'un simple découpage en sujet/prédicats qui ne rend pas possible l'explication de toutes les positions constitutives de la phrase. Grâce à la structure prédicats/arguments le sens de chaque prédicat devient plus facile à découvrir par l'intermédiaire de son entourage. On l'analyse à travers le prisme du type d'objets qu'il implique.

Selon cette approche, la caractérisation des termes correspondant aux arguments impliqués par les prédicats devrait se faire avec plus de rigueur qu'une simple indication postulée dans d'autres travaux linguistiques en *concret/abstrait, animé/inanimé* etc.

Ceci est également proposé par les dictionnaires de langue comme le *Nouveau Petit Robert* où la définition du verbe *manger — avaler pour se nourrir* précise qu'il s'agit non seulement d'un concret mais aussi de son sous-type, UN ALIMENT (solide ou consistant). Cette précision rend possible la création d'une classe d'objets qui comporte les éléments de nature semblable, qualifiés d'aliments et le nom d'aliment lui-même est appelé dans ce cas-là *substantif classifieur*.

Ainsi, procède-t-on à un certain *typage* des arguments (D. Le Pesant, M. Mathieu-Colas, 1998 : 12) ce qui revient à dire d'ailleurs que chaque emploi du prédicat peut être caractérisé en prenant en considération la catégorie sémantique de ses compléments.

Nous pouvons donc remarquer en résumant que les classes d'objets sont des ensembles de mots apparentés du point de vue sémantique en soulignant que lors de leur regroupement sont prises en compte leurs propriétés syntagmatiques.

Avant de passer à l'utilité de ces ensembles-là dans le cadre de l'approche orientée objets, voyons les principes généraux qui conduisent les travaux sur la désambiguïsation des sens des mots dans l'Institut des Langues Romanes et de Traduction à l'Université de Silésie.

L'approche orientée objets conçue par W. Banyś (2002, 2005) s'appuie sur la proposition de G. Gross et son équipe quant à l'analyse du sens des unités lexicales à la base de la notion de classe d'objets.

Cependant y sont ajoutées les spécifications des liens d'héritage sémantique en forme de super- et sous-classes. Ce mécanisme d'héritage sémantique que nous avons essayé d'expliquer plus haut facilite le travail linguistique. Les attributs et les opérations devraient s'ajouter automatiquement à chaque nouvelle sous-classe. Voyons l'analyse du mot *coteau* :

<i>stok</i>	<i>coteau</i>
[Super-klasa 5 :] <i>przestrzeń</i>	[Super-classe 5 :] <i>espace</i>
[Super-klasa 4 :] <i>część przestrzeni</i>	[Super-classe 4 :] <i>portion de l'espace</i>
[Super-klasa 3 :] <i>obszar</i>	[Super-classe 3 :] <i>étendue terrestre</i>
[Super-klasa 2 :] <i>powierzchnia pochyła</i>	[Super-classe 2 :] <i>surface inclinée</i>
[Super-klasa 1 :] <i>pochyłość</i>	[Super-classe 1 :] <i>pente</i>

Ainsi, les attributs et les opérations relevés pour une super-classe pourraient être partagés par ses sous-classes. Même si certains n'apparaissaient pas dans l'en-tourage des éléments appartenant à ces sous-classes-là, étant réservés uniquement à la super-classe, ils ne seraient pas hérités par le simple principe du manque de leur présence dans le contexte. Le cas contraire cependant est d'une utilité importante grâce à la possibilité d'élimination d'éventuels oubli ou négligences involontaires de la part du linguiste construisant le dictionnaire suivant les principes de l'approche orientée objets.

À titre d'exemple voyons encore quels attributs ou opérations peuvent être indiqués pour le mot *coteau* :

[Atrybuty :]	[Attributs :]
<i>lagodny stok</i>	<i>coteau en pente douce</i>
<i>latwy stok</i>	<i>coteau facile</i>
<i>ośnieżony stok</i>	<i>coteau enneigé</i>
<i>stok nawietrzny</i>	<i>coteau au vent</i>
<i>stok podwietrzny</i>	<i>coteau sous le vent</i>
<i>stok południowy</i>	<i>coteau sud</i>
<i>stoki klifu</i>	<i>coteaux d'une falaise</i>
<i>stromy stok</i>	<i>coteau abrupt</i>
etc.	
[Operacje :]	[Opérations :]
<i>jeździć na nartach na stoku</i>	<i>faire du ski sur un coteau</i>
<i>wspinac się na stok</i>	<i>grimper sur un coteau</i>
etc.	

La description possède un caractère ouvert, il est donc possible d'apporter des changements, de fournir de nouvelles informations dans l'analyse d'un objet. Dans

cette description on focalise l'attention sur le comportement linguistique de l'objet en exposant, de façon aussi détaillée que possible, son entourage lexical.

Toutes ces informations concernant l'objet analysé créent ensemble une structure qui porte le nom de *schéma* (frame) et qui, dans l'approche orientée objets à la W. Banyś, reçoit un format d'une fiche descriptive prenant en compte les données indispensables à une présentation exhaustive d'un lexème. L'origine de la notion de *schéma* est attribuée à M. Minsky (1986) et a été reprise par les chercheurs en domaine de l'Intelligence Artificielle constituant un des formalismes appliqués pour la représentation des connaissances. Ces schémas-là ont servi également dans l'explication des séquences d'actions stéréotypées connues sous le nom de *scénarios* (scripts). Chaque scénario met en jeu les éléments (objets et acteurs qui y participent), des conditions initiales qui doivent être accomplies pour qu'un événement ait lieu et les résultats possibles du scénario. Une telle analyse permet d'accéder à l'interprétation du texte et en extraire les prédictions (ou les résultats) (P. Bouillon, 1998 : 153—154) qui n'est pas sans influence sur la désambiguïsation des sens des mots là où les principes descriptifs appliqués ne paraissent pas suffisants et, par conséquent, le problème de l'ambiguité ne devient pas résolu.

L'analyse des lexèmes polysémiques dans le traitement automatique peut s'avérer efficace grâce à l'adoption d'une des méthodologies qui soulignent l'utilité et l'importance de l'objet, entre autres, l'approche orientée objets suivant laquelle nous effectuons nos recherches.

Il est notoire que le caractère polysémique des mots entraîne des difficultés dans la traduction classique avec le facteur humain et, qui plus est, dans la traduction automatique.

Dans la désambiguïsation des verbes, par exemple, l'objet n'est pas cette fois-ci décrit suivant son comportement linguistique qui équivaut à expliciter ses attributs et ses opérations mais il sert de point d'appui dans la reconnaissance d'un sens précis du verbe soumis à l'analyse.

Différents emplois verbaux (ou autres prédictifs), comme nous venons de le voir, dépendent donc du caractère des objets qui les entourent, la différence est visible dans la traduction, prenons comme exemple le verbe français *sortir* et ses équivalents polonais dans les deux emplois choisis :

X — [ANM] — **sortir** (— *de/sur/dans/à/en* — Y — [CONC <lieu>]) — **wyjść/wychodzić z/na/na/do/na**

Jean sort de son appartement.

Après les cours, les enfants sortent du collège.

Ce matin j'avais du mal à sortir de mon lit.

X — [CONC <illumination ; dégagement de lumière ; phénomène lumineux> ; <quantité d'air ou de gaz>] — **sortir** — *de/par* — Y — [CONC] — **wydobyć/wydobywać się z/przez**

La lueur d'un feu sort de la cheminée.

Le train ralentit, des étincelles sortent des roues alors que la vitesse diminue.

Le gaz sort de la bouteille.

Comme on peut le remarquer, les emplois relevés activent différents types de classes d'objets qui sont complétées d'éléments de nature semblable. Il est à observer en même temps que les spécificités des classes constituées font qu'elles peuvent parfois s'appliquer uniquement à l'analyse d'un seul prédicat même si leur caractère paraît identique, telle la classe des locatifs qui, dans le cas extrême, en fonction d'être impliquée soit par le verbe *sortir* soit par le verbe *monter* entraînera d'autres constituants :

Anne-Marie monte sur la colline.

**Anne-Marie sort de la colline.*

Nous pouvons remarquer également que les prédicats événementiels corrélés aux établissements introduits dans la classe des locatifs peuvent être applicables seulement pour les emplois du verbe *sortir* tandis que dans le cas de *monter* les éléments de la classe caractérisée par le même substantif classifieur seront naturellement exclus :

Jean sort du bal.

**Jean monte au bal.*

Voyons encore à titre d'exemple quelques objets des deux classes mentionnées dans les emplois ci-dessus :

illumination ; dégagement de lumière ; phénomène lumineux	oświetlenie ; wydzielanie światła ; zjawisko świetlne
clarté (f)	jasność (f)
étincelle (f)	iskra (f)
feu (m)	ogień (m)

quantité d'air ou de gaz	mieszanina powietrza lub gazów
air (m)	powietrze (n)
azote (m)	azot (m)
gaz (m)	gaz (m)
nuage (m)	obłok (m)

Dans la construction des classes d'objets notre tâche consiste à la préparation minutieuse de leur contenu et la présentation aussi complète que possible des élé-

ments qui répondent à leurs critères. Il existe bien entendu le risque de leur incomplétude, cependant, comme nous l'avons souligné, le caractère ouvert des analyses dans le cadre de l'approche orientée objets rend possible l'enrichissement progressif de ce travail.

Dans quel domaine peuvent être utiles les classes d'objets à part leur importance notable démontrée dans la désambiguïsation ?

Les classes constituent en gros des groupes de mots ordonnés suivant quelques principes. Elles présentent alors un type de dictionnaire thématique, un thésaurus qui relie les lexèmes par la relation d'association non seulement dans les limites d'une classe mais aussi, grâce à l'organisation hiérarchique, permet d'établir des liens entre plusieurs classes d'objets et par conséquent structurer une base de données ou de connaissances.

Pour quelqu'un qui souhaite connaître ou approfondir une langue étrangère, les classes peuvent s'avérer un outil intéressant et efficace dans le processus d'apprentissage, de plus, leur composition, formée souvent d'un nombre considérable d'éléments constitue une sorte de dictionnaire détaillé souvent plus riche que d'autres types de supports lexicaux disponibles sur le marché.

À côté de ses informations purement dictionnaires basées sur les approches à objets il existe d'autres apports que la traduction automatique effectuée par une machine peut fournir dans l'enseignement, y compris la traduction dont nous parlons dans la désambiguïsation suivant l'approche orientée objets. Ce type de traduction automatique peut être utilisée en vue de sa comparaison avec des traductions humaines. Ainsi peut-on attirer l'attention des élèves sur les ambiguïtés des langues naturelles d'ordre lexical ou syntaxique, sur la question de justesse de traduction, sur la différence entre le traitement humain et le traitement automatique, ou bien différencier par l'intermédiaire de ces deux types de traductions l'importance des situations de communication qui permettent de déterminer le style des traductions (A.-M. Loffler-Laurian, 1996 : 127—131).

Dans ce travail nous avons tenu à faire le point sur une des principales méthodes utilisées actuellement dans le traitement automatique des langues. Nous espérons que les origines de la modélisation objet, présentation de son développement, de l'importance des classes d'objets dans l'analyse des phénomènes linguistiques liés à la polysémie ainsi que des problèmes auxquelles elles peuvent se heurter et de ses apports possibles dans l'enseignement des langues étrangères, éveilleront l'intérêt du lecteur désireux d'approfondir ce sujet ou de poursuivre ces recherches.

Références

- Ayache M., Flory A., 1996 : *Approche Orientée Objet*. Paris, Economica.
- Banyś W., 2005 : « Désambiguïsation des sens des mots et représentation lexicale du monde ». *Neophilologica*, 17, 57—76.
- Banyś W., 2002 : « Bases de données lexicales électroniques — une approche orientée objets : Partie I et II ». *Neophilologica*, 15, 7—29 et 206—249.
- Benveniste E., 1966 : *Problèmes de linguistique générale*. Paris, Gallimard.
- Bouillon P., 1998 : *Traitements automatiques des langues naturelles*. Paris-Bruxelles, Éditions Duculot.
- Collin A.M., Quillian M.R., 1969 : “Retrieval time for semantic memory”. *Journal of Verbal Learning and Verbal Behavior*, 8, 240—247.
- Collin A.M., Quillian M.R., 1970 : “Does category size effect categorization time?” *Journal of Verbal Learning and Verbal Behavior*, 9, 432—438.
- Desclés J.-P., 1987 : « Réseaux sémantiques : la nature logique et linguistique des relations ». *Langages*, 87 [Paris, Larousse], 55—78.
- Delas D., 1978 : « La grammaire générative rencontre la figure. Lectures ». *Langages*, 51 [Paris, Larousse], 65—104.
- Fuchs C., 1993 : *Linguistique et traitements automatiques des langues*. Paris, Hachette Supérieur.
- Frege G., 1884 : *Les fondements de l'arithmétique*. [trad. fr. : 1970]. Paris, Le Seuil.
- Gleason J.B., Ratner N.B., 2005: *Psycholinguistyka*. Gdańsk, Gdańskie Wydawnictwo Psychologiczne.
- Gross G., 1994b : « Classes d'objets et description des verbes ». *Langages*, 115 [Paris, Larousse], 15—31.
- Le Pesant D., Mathieu-Colas M., 1998 : « Introduction aux classes d'objets ». *Langages*, 131 [Paris, Larousse], 6—33.
- Loffler-Laurian A.-M., 1996 : *La traduction automatique*. Villeneuve-d'Ascq, Presses Universitaires de Septentrion.
- Minsky M., 1986 : *The Society of Mind*. New York, Simon & Schuster (trad. fr. : *La Société de l'esprit*. Paris, InterEditions, 1988).
- Prandi M., 1998 : « Contraintes conceptuelles sur la distribution : réflexions sur la notion de classe d'objets ». *Langages*, 131 [Paris, Larousse], 34—44.
- Rastier F., 1987 : « Représentation du contenu lexical et formalismes de l'Intelligence Artificielle ». *Langages*, 87 [Paris, Larousse], 79—102.