

KRZYSZTOF BOGACKI

QUELQUES RÉFLEXIONS SUR LES PRIMITIVES,
LES PRÉDICATS, LE SENS DES TERMES
DANS LES ONTOLOGIES ET DANS LE LEXIQUE STANDARD

SOME THOUGHTS ON PRIMITIVES, PREDICATES
AND THE MEANING OF TERMS
IN ONTOLOGIES AND IN GENERAL VOCABULARY

Abstract

The article deals with two terms: “primitives” and “predicates” found in technical reports on ontology, on natural languages and in programming languages. The way they are used seems to confirm the Latin phrase: *Si duo dicunt idem non est idem*. However, one must attenuate this supposition which is due to the fact that the descriptions of ontologies, of the programming code and of the natural language texts are made in formal language. The obtained descriptions have two aspects: ‘formal’ and ‘lexical’ or ‘semantic’. Both predicates and primitives appear in linguistics and computer science on the formal level, and not on the ‘lexical’ one. A glance at the Charter of Fundamental Rights of the European Union as the base-text used to develop a micro-ontology leads to the conclusion that the authors introduce the concept of *dignité humaine* ‘human dignity’ by an arbitrary stipulative definition rather than by reporting one. The meaning assigned to this term does not coincide with any of the registered meanings of this expression in the standard language

Key words: ontology, predicates, primitives, terms in ontologies, natural language vocabulary.

Un des avantages du web facilement accessible par l’intermédiaire des moteurs de recherche est qu’il offre la possibilité d’examiner les contextes dans lesquels apparaissent les termes. Cette propriété nous a permis de nous pencher sur le sens donné à des termes tels que *prédicats* et *primitifs* (ou

primitives). Ils se retrouvent non seulement dans le discours des linguistes examinant la syntaxe et la sémantique des langues naturelles mais aussi dans le code des informaticiens et, last but not least, sous la plume des philosophes. Nous espérons pouvoir montrer que les représentants de ces trois disciplines se servent des mêmes termes mais le sens qu'ils leur confèrent varie selon la discipline représentée. Est-il toutefois légitime d'affirmer de façon catégorique : *Si duo dicunt idem non est idem* ? Ou faut-il plutôt adopter une position plus nuancée ?

Notre article se décline en plusieurs volets. Tout d'abord, nous rappellerons les notions principales sur lesquelles va porter notre réflexion en commençant par celle d'ontologie.

Après avoir jeté un rapide coup d'œil sur les techniques d'enrichissement de l'ontologie en TAL, nous évoquerons la notion de prédicat et celle de primitives. Pour terminer, nous examinerons le terme de *dignité* tel qu'il apparaît dans la micro-ontologie formée à partir de la « Charte des droits fondamentaux de l'Union européenne » en version française et le sens que le même lexème a dans la langue.

UN PREMIER COUP D'ŒIL SUR L'ONTOLOGIE ET SUR LES ONTOLOGIES

A l'origine, l'ontologie était située entièrement sur le terrain de la philosophie. S'intéressant à la nature et à l'organisation de la réalité, constituant une branche de la métaphysique et s'interrogeant sur l'être, elle était la « science de l'être en tant qu'être » par excellence. Elle s'occupait non seulement de la justification de l'être, mais aussi de sa description et ouvrait un vaste champ de recherches pour des générations de philosophes. En informatique, l'ontologie apparaît aux alentours de 1980 sous l'impulsion des recherches en logique, en intelligence artificielle et en sciences cognitives. En effet, on a vite découvert l'intérêt de l'ontologie – toujours au sens de questionnement sur les êtres représentés dans les textes manipulés – pour la recherche et l'extraction d'information. Grâce aux travaux entrepris par les informaticiens, une distinction s'est établie entre ontologie philosophique (au singulier) et ontologies (des informaticiens et, par extension, de représentants d'autres disciplines, au pluriel). A la question de savoir quel est le sens de ce terme pour les informaticiens, on répond le plus souvent par la formule de T. Gruber (2009 :1959) : « In the context of computer and infor-

mation sciences, an ontology defines a set of representational primitives with which to model a domain of knowledge or discourse. » Une autre formule, celle de Mondary (2011 : 10) stipule qu'une ontologie « est une vue abstraite et simplifiée du monde que l'on souhaite représenter dans un but donné. » Pour les informaticiens, le terme d'ontologies désigne aujourd'hui « a special kind of information object or computational artifact. » [un certain type d'information ou d'un artéfact informatique'.] (Staab, Studer 2009 : 2). C'est un artéfact symbolique, car pour pouvoir être utilisé par un ordinateur, il est construit dans un langage formel apte à manipuler des entités conceptuelles organisées en arbre ou en treillis. Les rapports reliant les entités conceptuelles sont explicites et les relations mises en jeu sont le plus souvent celles de subsomption et de méréonymie. Cette spécification explicite d'une conceptualisation des données fournies par une hiérarchie d'entités devrait rendre possible le raisonnement valide (tirer des conclusions). Si les informaticiens exigent que cette représentation soit formulée dans un langage utilisable par une machine, c'est parce qu'ils placent au centre de leurs préoccupations l'aspect « ingénierie », contrairement aux philosophes. Vu la polyvalence de la réalité que les informaticiens essaient de modéliser, ils ne prennent en compte que ces propriétés de la réalité qui sont importantes voire indispensables pour leurs applications. D'où le caractère nécessairement fragmentaire et incomplet de leurs propositions.

Les ontologies non-philosophiques prennent des formes variées : listes des termes, glossaires offrant à côté des termes leurs définitions, thésaurus qui en fait sont des hiérarchies de termes, taxonomies – dans lesquelles les primitives sont structurées hiérarchiquement afin de permettre l'héritage des propriétés.

LES LANGAGES POUR UNE ONTOLOGIE ?

S'il y a différentes visions des rapports entre l'ontologie et le réel représenté¹, les chercheurs sont d'accord pour faire la différence entre ontologie qui conceptualise le réel et lexique qui est une émanation de la langue. Nul doute d'autre part qu'on a besoin pour en parler d'un langage qui, en plus, doit avoir certaines caractéristiques pour être exploité par la machine.

¹ Cf. P. Saint-Dizier et P. Viegas Evelyn (eds.) (1995) pour une discussion du problème de la primauté de la langue et de ses conséquences pour la forme que revêtirait l'ontologie.

Différents types de langages ont été conçus pour le stockage et la manipulation des ontologies. Parmi les plus connus on notera tout d'abord le plus ancien Dublin Core (DC), Knowledge Interchange Format (KIF), DARPA Agent Markup Language et le plus utilisé de tous : OWL (Web Ontology Language depuis 2001)² et ses dialectes (OWL Lite, OWL DL et OWL Full). D'autres langages, associés à des systèmes plus ou moins connus, et imposant un formalisme spécifique sont aussi accessibles. Pour une présentation plus détaillée nous renvoyons le lecteur à G. Antoniou et al. (2005) et à S. Staab et R. Studer (2009).

TYPES D'ONTOLOGIES

Les ontologies existantes sont très diversifiées et peuvent être classées de différents points de vue. Les classements qu'on en propose tiennent compte de l'opposition entre ontologie globale (ou de haut niveau qui représente les concepts de sens commun³) et celles de domaine qui sont limitées à un secteur particulier du réel. Celles-ci peuvent porter sur des réalités variées : moyens de transports, maladies, musique, film, astronomie, perception de paysages, éléments visuels, etc.

Il est possible de proposer un autre critère de classement – celui du public visé. En effet, une ontologie décrivant les moyens de transport : voitures, avions, trains, etc. destinées à un large public à la veille des vacances tiendra compte de moins de détails que celle dont se servent les techniciens chargés de l'entretien des mêmes véhicules. Elle mettra aussi en valeur d'autres propriétés d'objets décrits. L'étendue de l'ontologie et la granularité de la description conceptuelle varieront en fonction de ce critère. Les ontologies d'application sont construites en fonction des besoins spécifiques qui découlent d'une tâche donnée.

L'examen des ontologies accessibles sur le web permet de constater que la description des entités qu'on y trouve n'est pas identique d'une ontologie à l'autre. Ainsi P. Buitelaar (2010 : 219) note 5 sens suivants du mot *article* :

1. An individual thing or element of a class [...]
2. A particular section or item of a series in a written document [...]

² Cf. Lacot (2005).

³ La plus étendue est l'ontologie de la Cycorp (cf. <https://www.cyc.com>) qui, en accès libre, offrait en juillet 2015 environ 500000 notions et plus de 70.000 prédicats reliées par environ 7 millions d'assertions.

3. A non-fictional literary composition that forms an independent part of a publication [...]
4. The part of speech used to indicate nouns and to specify their application
5. A particular part or subject; a specific matter or point

Or l'ontologie COMMA⁴ retient *article* avec le sens 2 :

```
<rdfs:Class rdf:ID='Article'>
<rdfs:label xml:lang='en'>article</rdfs:label>
<rdfs:subClassOf rdf:resource='#Document'/>
<rdfs:subClassOf rdf:resource='#ExtractedDocument'/>
<rdfs:comment xml:lang='en'>Document corresponding to a piece of
writing on a particular\index {particular} subject and which purpose is to
fully realize a particular objective in a relatively concise form
e.g.:demonstrate something.</rdfs:comment>
</rdfs:Class>
```

alors que GOLD⁵ le décrit comme représentant le sens 4 (*article* est une sous-classe de la classe : *determiner*).

```
<owl:Class rdf:ID='Article'>
<rdfs:label xml:lang='en'>article</rdfs:label>
<rdfs:subClassOf rdf:resource='#Determiner'/>
<rdfs:comment>
```

```
An article is a member of a small class of determiners that identify a noun's
definite or indefinite reference, and new or given status... </rdfs:comment>
</owl:Class>
```

ONTOLOGIE FORMÉE À PARTIR D'UN TEXTE – EXTRACTION DE RELATIONS SÉMANTIQUES

Dans le processus de construction d'ontologies, on utilise parfois des corpus de textes partant du principe que la trace des concepts se rapportant au domaine décrit peut être repérée par extraction terminologique. Ainsi par exemple pour l'enrichissement semi-automatique d'ontologies on a recours à l'extraction de relations sémantiques entre noms. Celles qui sont recherchées le plus souvent sont la synonymie, la méronymie et l'hypéronymie. Diffé-

⁴ Cf. <http://www.si.fr.atosorigin.com/sophia/comma/Htm/HomePage.htm>

⁵ Cf. <http://linguistics-ontology.org/version>

rentes méthodes ont été proposées et plusieurs logiciels élaborés. Dans sa thèse, P. Séguéla (1999) passe en revue les logiciels et les méthodes utilisées qu'il divise en deux groupes : acquisition de marqueurs par analyse des collocations et acquisition de marqueurs par projection de modèles du domaine. Elles utilisent le repérage des marqueurs linguistiques spécifiques pour chaque type de relation et cette méthode est parfois combinée avec le recours à la statistique, ce qui renforce considérablement son rendement.

Selon Morin (1999 :62) les étapes principales de l'opération sont les suivantes : on choisit la relation sémantique pour laquelle on fournit une amorce constituée de termes. On passe ensuite à l'examen d'un corpus (lemmatisé) dont on extrait toutes les occurrences du couple de termes avec les contextes qui les renferme. Ceux-ci révèlent les schémas lexico-syntaxiques candidats qui doivent être validés et vont servir à l'extraction de nouveaux couples candidats. On travaille ainsi dans une boucle qui sert à enrichir la liste de structures lexico-syntaxiques. Dans son corpus, Morin (1999 : 69) relève différents types de schémas lexico-syntaxiques : 3 phrases correspondant au schéma « SN être composé SN », 3 pour « SN être constitué SN », 5 pour « production de SN par SN », 6 pour « étude de l'effet de SN sur SN », 8 pour « SN comme LISTE », 10 pour « SN tel que LISTE » et 35 phrases ne correspondant à aucun des schémas précédents.⁶

Il existe dans les textes certaines formes régulières qui constituent des indices d'une relation mais leur repérage n'est pas toujours chose aisée, tellement les marqueurs peuvent être variés. Prenons comme exemple la relation de méronymie. Le marquage se fera tout d'abord par l'expression *faire partie de* et, dans un deuxième temps, par ses synonymes, p. ex. *grouper, former, posséder, assembler, unir, joindre, associer, rassembler, fusionner, combiner, réunir*, etc. Or, pour cette notion, que nous avons explorée à partir du verbe *réunir*, <http://www.crisco.unicaen.fr/des/> mentionne 90 synonymes organisés en 180 cliques. La liste fournie par les dictionnaires courants de synonymes devrait être complétée par des données tirées du vocabulaire technique. Ainsi la phrase : *Le polyol possède une chaîne carbonée de cinq atomes* allongerait notre liste par le verbe *posséder*. Ce qui plus est, la méronymie est réalisée grâce à des structures que l'on retrouve pour l'expression d'autres relations. C'est le cas de X ETRE DET Y que l'on retrouve avec les synonymes : *Le glycérol est un polyol*. Finalement, la méronymie peut aussi être exprimée par des structures non-verbales :

⁶ Cf. aussi la thèse de N. Hernandez (2005) et sa bibliographie.

Parmi les membres de l'équipe, il y avait un Nord Africain et trois Asiatiques. Le glycérol, plus communément appelé glycérine, est une molécule qui se retrouve dans des produits pharmaceutiques ou cosmétiques pour ses propriétés hydratantes, ou alimentaires comme solvant ou émulsifiant.

C'est dire que le dispositif informatique risque d'être incomplet et laissera échapper bon nombre d'éléments présents dans le corpus.

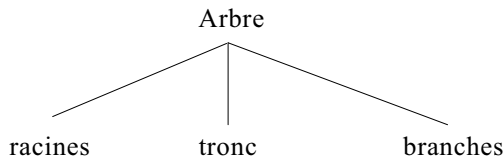
GRAPHE ONTOLOGIQUE

L'ontologie peut être représentée sous forme d'une arborescence ou d'un autre graphe. Compte tenu des besoins pour un même fragment de la réalité on peut avoir plusieurs ontologies de granularité différente. Pour la description d'un arbre on pourrait imaginer une formule tenant compte du fait que l'arbre a des racines, des branches et un tronc.

Dans le langage OWL, ce fragment aurait la forme suivante :

```
...
<owl:Class rdf:ID="arbre" />
<rdfs:subClassOf rdf:PartOf="racines"/>
<rdfs:subClassOf rdf:PartOf="tronc"/>
<rdfs:subClassOf rdf:PartOf="branche"/>
...
</owl:Class>
```

En forme de graphe simplifié (et donc considérablement appauvri !), l'ontologie décrivant un arbre aurait la forme suivante :

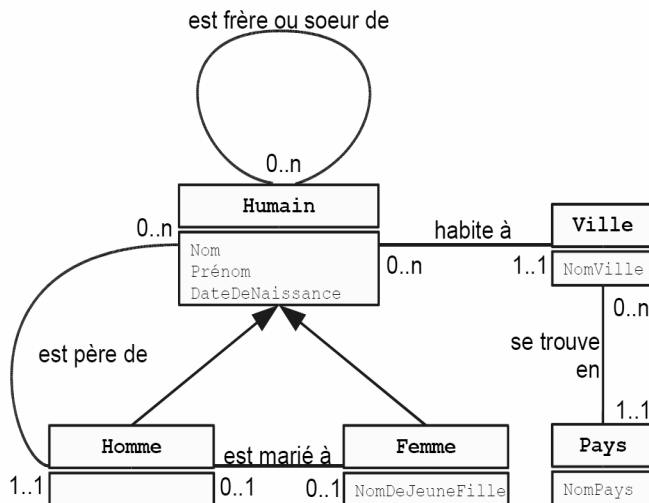


Ces informations sont réductibles à une série de définitions qui peuvent être transformées en représentations sémantiques formées de prédicats avec leurs arguments. En effet, le diagramme correspond aux trois assertions ci-dessous :

- L'arbre a des racines*
- L'arbre a un tronc*
- L'arbre a des branches*

sous-tendues par une représentation sémantique telle que PARTIE_DE (x, y)

Les graphes peuvent avoir des formes plus compliquées et rendre compte de relations plus sophistiquées. D'après X. Lacot (2005) nous reproduisons ci-dessous un treillis correspondant à une micro-ontologie. Elle décrit les relations possibles au sein d'une population composée d'humains, hommes et femmes. Les relations et les objets dont on tient compte sont : frère/sœur, habiter à, ville, pays, être marié(e), nom de jeune fille, nom, date de naissance.



LES PRIMITIVES DANS DIFFÉRENTS CONTEXTES : TROIS PERSPECTIVES

Deux termes sont incontournables dans le contexte informatique et linguistique. Ce sont les primitives et les prédicats. Ils n'apparaissent pas par contre dans les discussions des philosophes sur l'être, l'essence, les accidents, etc. bref, dans le cadre de l'ontologie philosophique. Plusieurs points en relation avec les primitives sont à rappeler ici.

La notion de primitive est plus facile à cerner en **linguistique** où de nombreux travaux, datant de Leibniz et fréquents surtout à la fin du XX^e siècle ont permis de fixer un certain nombre de points. L'idée générale de primitives n'est pas sans rappeler la situation des recherches en physique. L'exemple classique est celui de la représentation, dépassée à l'époque actuelle, qu'on se faisait de la matière comme étant composée d'atomes considérés comme entités indivisibles. En linguistique, la notion de primitives, souvent liée de façon explicite à la technique de la décomposition du sens,

a été confondue avec celle d'universaux. D'un côté, les primitives constituent le point d'aboutissement d'une analyse décompositionnelle descendante des lexèmes livrant au bout du compte une liste d'entités en nombre réduit, de l'autre elles forment un système « de représentation du sens à partir d'ensembles limités de primitives. »

Comment identifier les primitives en linguistique ? Prenons l'exemple emprunté à la sémantique. Le point de départ est donné par l'observation de la complexité sémantique des unités lexicales telles que *verdir*, *blanchir*, *bleuir*, *rougir* et d'autres verbes dérivés d'adjectifs dont le sens peut être décomposé en RENDRE + Adj. L'analyse peut être continuée en appliquant le même mécanisme aux composants dégagés. Ainsi RENDRE serait décomposé en FAIRE+DEVENIR (ou en CAUSER + DEVENIR), etc. jusqu'au moment où aucune décomposition ne serait possible. Cette démarche devrait impliquer la possibilité de faire le chemin inverse, c'est-à-dire on devrait pouvoir reconstituer un concept complexe à partir de ses éléments constitutifs.

Est-il possible de structurer nos connaissances, dans une ontologie, autour d'un petit nombre d'éléments abstraits indivisibles ? L'avis qui prévaut est qu'il est difficile voire impossible d'isoler et de décrire en ontologie des primitives conceptuelles capables d'exprimer par combinatoire le sens de n'importe quel concept non seulement au sein de l'ontologie fondationnelle (de haut niveau) mais même au sein d'un domaine donné⁷. Ce qui plus est, on observe aussi les entités conceptuelles-éléments d'une ontologie qui ne sont pas nécessairement sémantiquement simples (donc indécomposables a priori). Dans une ontologie, on considère comme primitive un concept qui, certes, n'apparaît pas comme décomposable dans cette ontologie-là sans se soucier de la possibilité théorique de l'utiliser dans une autre ontologie de granularité plus élevée où il apparaîtrait comme analysable. Par exemple *droit d'asile* et *droit de propriété* peuvent faire partie, tous les deux, de deux ontologies : O_1 et O_2 . Dans O_1 ils se trouveraient en bas d'une hiérarchie de concepts constituant le nœud final, point d'aboutissement d'une décomposition possible. Dans O_2 , au contraire, le système admettrait une différenciation des droits : il y a aurait donc un droit d'asile, un droit de propriété, un droit à l'éducation, les citoyens ont le droit de vote, le droit de pétition, le droit d'accès aux documents etc. (cf. *Charte*). Le statut de primitive reviendrait dans O_2 à *droit* et non à *droit d'asile*. En d'autres termes, les

⁷ Cf. Declerck et al. 2012

« primitives » des informaticiens sont des primitives d'un modèle : indécomposables dans telle ontologie, décomposables (donc n'ayant pas le statut de primitives) dans telle autre.

PRIMITIVES ONTOLOGIQUES ET PRIMITIVES LINGUISTIQUES

Il semble que le statut épistémologique des primitives ontologiques des informaticiens ne soit pas clair. D'autre part, sur le plan linguistique, on ne saurait leur accorder le statut d'entités sémantiquement simples. Ainsi en examinant la possibilité d'élaborer automatiquement une taxonomie des actions, Weinland et al (2006) identifient 22 actions auxquelles ils attribuent le statut de « primitives ». On y trouve, entre autres : *jump*, *lower firm leg*, *turn in new position*, *rebound after jump*, *crouch before jump*, *lift leg firm*, *lift arm ahead*, *lower arm from ahead*, *lower arm sideways*, *lift arm ahead (single)*, *lift both arms ahead (single)*, *rotate both arms lifted*, *lift arm sideways*, *lift arm sideways (single)*, *do nothing (single)*, *lift both arms ahead*, *turn both arms sideways*, *lower both arms sideways*, *lift leg bending*, etc. Or sur le plan purement linguistique le caractère sémantiquement complexe de tous ces concepts est évident.

S'il existe un courant en linguistique qui se caractérise par une recherche des primitifs sémantiques, cette préoccupation est étrangère aux recherches dans la confection des ontologies domaniales ou de celles de haut niveau et en informatique qui ne se soucient guère de proposer des notions sémantiquement indécomposables. A notre connaissance, dans le domaine des ontologies, contrairement aux recherches en linguistique qui définissent, tant bien que mal, la nature des éléments primitifs, il n'existe aucune tentative de caractérisation d'unités élémentaires postulés non décomposables qui par des combinaisons successives fournirait des entités de rang supérieur.

Bachimont (2000) montre cependant que dans le discours sur les ontologies il existe un niveau où il serait justifié de parler des primitives. C'est le niveau formel incluant les relations, les fonctions, les rôles thématiques, les symboles de variables, les prédicats (chacun caractérisé par une certaine arité et un type particulier d'arguments), les connecteurs, les quantificateurs. Ce niveau existe à côté de la couche « lexicale » (ou « sémantique ») manipulée par cet appareil formel. C'est grâce à elle qu'une ontologie médicale a un contenu différent de celui d'une ontologie décrivant la structure d'une organisation multinationale même si les deux ontologies sont décrites à

l'aide d'un même langage formel. C'est à ce niveau qu'on parlera de « primitives de représentation » ayant chacune une caractéristique « lexicale ». Elles seraient différentes d'un domaine à l'autre et, contrairement aux primitives sémantiques que le linguiste découvre en langue par une série de décompositions lexicales, elles n'existent pas au niveau de l'ontologie philosophique.

DES PRÉDICATS À TOUT BOUT DE CHAMP

En **linguistique** « le terme de prédicat recouvre un ensemble d'acceptions qui constitue presque un continuum s'étendant entre deux pôles : une acception essentiellement grammaticale et une acception essentiellement logique. » (<http://www.semantique-gdr.net/dico/index.php/Pr%C3%A9dicat>). Dans chaque cas de figure, en linguistique comme en logique, à côté de cette caractéristique « formelle », les prédicats ont une caractéristique lexicale : certains expriment le remplacement, d'autres la suppression, la transformation, la copie, la possibilité, la possession, la connaissance, la perception, etc.). Les prédicats sont dégagés de différentes manières et présentent une ressemblance plus ou moins grande avec les structures syntaxiques de surface (le prédicat y est rapproché du schéma syntaxique du verbe). Morphologiquement, ils peuvent revêtir différentes formes : verbale avant tout mais aussi substantivale, adjectivale ou prépositionnelle.⁸ Sémantiquement ils peuvent être simples ou complexes. Dans ce dernier cas, ils peuvent être décomposés en éléments sémantiquement plus simples et leur analyse, poussée jusqu'au bout, devrait se solder par l'établissement de prédicats « atomiques » indécomposables.⁹ Leur liste formerait un inventaire d'outils « primitifs » de base. L'inventaire d'outils comprend aussi les constantes logiques (négation, différents types de connecteurs : connecteurs *de* coordination, d'égalité, d'alternative etc.). Le mécanisme analytique autorise aussi la prise en compte des éléments déictiques de type temporel (instants : t_0 , t_{-1} , t_{+1}).

Parfois les recherches sur les prédicats se combinent avec celles sur les primitives sémantiques – ou en d'autres termes – on attribue aux supposés primitives la valeur de prédicat ayant une arité (unaire ou binaire).¹⁰ De

⁸ Cf. G. Gross (2012).

⁹ Cette conception, qui remonte aux travaux de Leibniz (1975), a été développée récemment par A. Wierzbicka (1969, 1971, 1975, 1980, 1996, 2004), Bogacki & Karolak (1991) et d'autres.

¹⁰ Soit dit entre parenthèses que si l'espoir d'arriver à un inventaire complet et définitif de primitives sémantiques est vain, leur prise en compte ouvre tout de même des perspectives de

telles préoccupations sont absentes de la pratique informatique où l'on appelle parfois prédicats des structures sémantiquement fort complexes.

Un prédicat **en informatique** est défini de différentes façons. C'est une méthode qui représente un ensemble de critères et décide, pour un objet donné, si ces critères sont remplis ou non. Si l'on adopte cette position, on dira que dans les langages informatiques un prédicat est assimilable à une fonction qui renvoie une valeur booléenne : « VRAI » ou « FAUX ».

Dans l'exemple ci-dessous, le prédicat **transform** calcule la puissance 3 des nombres contenus dans la variable

```

.....
std::vector< int > cubes( 10 ); // instantiate vector cubes
// calculate cube of each element in v; place results in cubes
std::transform( v2.begin(), v2.end(), cubes.begin(), calculateCube );
cout << "\n\nThe cube of every integer in Vector v is:\n";
std::copy( cubes.begin(), cubes.end(), output );
cout << endl;
return 0;
} 11

```

Une autre façon de concevoir un prédicat consiste à dire que c'est « une formule qui peut être prouvée ou réfutée, ou qui peut faire partie des hypothèses sous lesquelles on fait une preuve. »¹²

Comme en linguistique, le prédicat des informaticiens, qu'il soit simple ou complexe, implique un ou plusieurs arguments. Il se caractérise donc par une arité et par le type d'argument(s) convoqué(s).

On parle aussi des prédicats dans différents systèmes **ontologiques** (différentes ontologies) ce qui n'étonne pas : on a affaire à un artéfact qui relève de l'informatique. Ainsi Guarino et al. (2009 : 2) disent : « The ontology engineer analyzes relevant entities [...] and organizes them into concepts and relations, being represented, respectively, by unary and binary predicates. » Effectivement le mécanisme que l'on retrouve au niveau de l'ontologie où certaines entités (concepts, propriétés, relations) sont vérifiées par d'autres rappelle le mécanisme prédicatif qui met en jeu les prédicats qui pour être vrais exigent certains arguments. (cf. http://semantics.univ-paris1.fr/pdf/Programme_OP_2008.pdf).

recherches intéressantes p. ex. via l'idée de parallélisme isomorphe entre le comportement syntaxique des lexèmes et les propriétés des prédicats formant leur représentation sémantique.

¹¹ http://www.java2s.com/Tutorial/Cpp/0480__STL-Algorithms-Modifying-sequence-operations/stdtransformwithpredicate.htm

¹² Cf. « *Manuel de référence du langage B* », Version 1.8.5 du 10 janvier 2002 (p. 29).

CHARTER ET ONTOLOGIE

Etant donné que langue et ontologie sont inséparables on comprend que l'élaboration d'ontologies se fait souvent en exploitant un corpus de textes ce qui présuppose l'adhésion à la thèse de la bi-directionnalité de la relation entre ontologies et labels textuels correspondant aux concepts trouvés dans les ressources lexicales (Hirst 2004). Il nous paraît donc légitime d'examiner les termes saillants repérés dans la *Charte*. Nous allons les comparer avec le sens qu'ils ont dans la langue. La procédure commence toutefois par la définition des besoins auxquels elle doit répondre et se termine par une opérationnalisation qui revient à représenter l'ontologie dans un langage de représentation des connaissances (p. ex. graphes conceptuels ou logiques de description). Des inférences devraient être possibles pour classer, généraliser des concepts ou découvrir les faits qui n'ont pas été mentionnés ou décrits expressément). La micro-ontologie déduite de la *Charte* fait partie d'ontologies de domaine. Elle devrait être mise en relation avec une ontologie de haut niveau qui offre un cadre d'accueil pour les catégories formant l'ossature générale de la *Charte*.

STRUCTURE DE LA CHARTER

La *Charte* que nous examinons dans sa version française, précédée d'un préambule, est subdivisée en 7 parties. Les six premières forment l'essentiel du contenu, la 7^e décrit le champ d'application, la portée et l'interprétation des droits et des principes énumérés dans les parties précédentes et précise le niveau de protection en ajoutant un principe supplémentaire qui est l'interdiction de l'abus de droit. La conjonction des concepts introduits dans les six premières parties sont à entendre comme une liste énumérative de notions, chacune ayant le statut de notion analysable donc, en fin de compte, non-primitive. Ce sont : *dignité humaine, libertés, égalité, solidarité, citoyenneté, justice*. Les développements proposés pour chacun des nœuds ci-dessus constituent le profil de chacun d'eux comparables à ce qu'on trouve dans la micro-ontologie décrivant l'arbre qui a un tronc, des racines et des feuilles.

Une analyse ontologique naïve de la *Charte* devrait conduire à la conclusion que les concepts essentiels-nœuds constituent un arbre ontologique réutilisable dans une ontologie de niveau plus élevé. Ces entités devraient correspondre aux exigences des spécialistes dans le domaine traité : juristes, sociologues, philosophes, hommes politiques etc.

Deux observations s'imposent :

1. Certains des concepts-nœuds sont étiquetés avec des labels qui apparaissent à différents endroits de l'arbre ontologique. Ainsi, il est question de droit dans la définition

- ◆ de la dignité (droit à la vie, droit à l'intégrité de la personne),
- ◆ des libertés (droit à la liberté et à la sûreté, droit de se marier et droit de fonder une famille, droit à l'éducation, droit de propriété, droit d'asile),
- ◆ de l'égalité (droits de l'enfant, droits des personnes âgées),
- ◆ de la solidarité (droit à l'information et à la consultation des travailleurs au sein de l'entreprise, droit de négociation et d'actions collectives, droit d'accès aux services de placement),
- ◆ de la citoyenneté (droit de vote et d'éligibilité aux élections au parlement européen, droit de vote et d'éligibilité aux élections municipales, droit à une bonne administration, droit d'accès aux documents, droit de pétition),
- ◆ de la justice (droit à un recours effectif et à accéder à un tribunal impartial, droit à ne pas être jugé ou puni pénalement deux fois pour une même infraction)

Nous avons là, semble-t-il, des entités distinctes qui concourent à définir l'entité de l'ordre supérieur. Or, on est loin de la liste complète et définitive d'entités formant une ontologie d'un fragment de la réalité. En effet, les ontologies du droit proposent de longues listes d'entités dont elles se servent.¹³

2. Il existe un écart entre l'organisation de l'ontologie et celle, lexicographique, des labels.

Considérons le concept de dignité. Les dictionnaires¹⁴ notent 4 sens de *dignité* :

1. Sentiment de la valeur intrinsèque d'une personne ou d'une chose, et qui commande le respect d'autrui.

2. Attitude de respect de soi-même, fierté. Synon. *Respectabilité*. Comportement traduisant ce respect.

3. Prérogative ou prestige inaliénables dont jouit une personne en raison de son comportement, ou qui sont attachés à une chose, et qui leur valent considération et respect ou y donnent droit.

4. Prérogative (charge, fonction ou titre) acquise par une personne (un groupe de personnes),

¹³ Cf. Breuker, J. et al. (2007) et Hoekstra, R. et al. 2007) ainsi que le projet européen Estrella (www.estrellaprojecf.org).

¹⁴ Cf. Le Petit Robert et le TLFi.

Dans la *Charte* la zone d'applicabilité de ce concept est délimitée par une conjonction de 4 traits : **droit à la vie, droit à l'intégrité de la personne, interdiction de la torture et des peines ou traitements inhumains ou dégradants, interdiction de l'esclavage et du travail forcé**. Nous sommes donc dans le domaine qui implique le prédicat déontique. Sémantiquement, il relève de la possibilité affirmée (pour *droit*) ou soumise à la négation (pour *interdiction*) : POSSE (R(x)) ou NON POSSE (R(x)). Les sens de *dignité* dans la langue se situent au contraire dans un domaine qui peut être décrit par le prédicat de sentiment RESPECT : 'sentiment qui porte à accorder à qqn une considération admirative, en raison de la valeur qu'on lui reconnaît, et à se conduire envers lui avec réserve et retenue, sentiment qui incite à traiter quelqu'un avec égards, considération, en raison de ...'. Dans la représentation sémantique, on retrouve le prédicat axiologique (BON/MAUVAIS) combiné avec l'action et la causalité (protection) et la croyance combinée avec l'activité et une relation indéterminée (respect).

Tout porte à croire donc qu'en proposant une vision nouvelle de la dignité, radicalement différente de celle attestée en langue et repérable dans les dictionnaires, la *Charte* a recours à une procédure de stipulation. Elle revient à assigner un sens nouveau à un label déjà existant en langue.

CONCLUSION

A première vue, quoi de plus éloigné que le discours des informaticiens ou des linguistes sur les entités de base utilisées dans leurs analyses ? En effet, les uns et les autres semblent utiliser les mêmes termes – *primitives* et *prédicats* – de façon différente. Il s'avère cependant qu'on peut trouver des points communs.

Faisant une distinction, à la suite de Bachimont (2009), entre leur aspect « formel » et « lexical » ou « sémantique », on sera obligé d'admettre qu'au niveau « formel » du langage utilisé en **informatique** et dans certaines analyses linguistiques, les notions de primitives et de prédicats se ressemblent. On y trouve les mêmes caractéristiques s'exprimant en termes d'arité, d'arguments, de quantificateurs, d'opérateurs logiques etc. Au contraire, au niveau « lexical » la situation est toute différente. En **linguistique** on trouve deux démarches. D'un côté, le courant représenté par A. Wierzbicka et ses disciples a fait des efforts considérables pour dresser une liste exhaustive de primitives indécomposables qui se combinant entre elles forment les repré-

sentations sémantiques des lexèmes. Les représentants d'autres écoles, moins soucieux d'utiliser dans leurs recherches des éléments ultimes, arrêtent la granularité de leurs analyses à un niveau moins poussé¹⁵. Pour les informaticiens élaborant leur code ou construisant leurs ontologies, la complexité des entités de base n'a jamais constitué le problème majeur. En effet, leurs ontologies contiennent des entités dont le degré de complexité est fonction du but visé.

Si l'on adopte la *Charte des droits fondamentaux de l'Union Européenne* comme point de départ textuel pour l'établissement d'une micro-ontologie pour le concept de dignité, on s'aperçoit qu'il ne coïncide avec aucun sens de ce mot existant en langue. Dans le système ontologique de la *Charte* il est introduit par une décision qui ressemble à une définition stipulatoire loin de la définition descriptive qui est celle des lexicographes.

BIBLIOGRAPHIE

- Antoniou Grigoris, Franconi Enrico, and van Harmelen Frank, 2005, Introduction to Semantic Web Ontology Languages [in :] N. Eisinger and J. Maluszynski (Eds.) REWERSE 2005, Springer-Verlag .
- Bachimont Bruno, 2000, Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en ingénierie des connaissances [in :] *Ingénierie des connaissances. Evolutions récentes et nouveau défis*, Charlet, Jean et al., Eds, Eyrolles, 305-323. On-line: https://www.utc.fr/~bachimon/dokuwiki/_media/fr/ontologie-icbook.pdf. [Consulté le 19 Juillet 2015].
- Bogacki Krzysztof & Karolak Stanisław, 1991, Fondements d'une grammaire à base sémantique. [in :] *Lingua e Stile*, 26(3), 309-345.
- Breuker Joost, Hoekstra Rinke, Boer Alexander, van den Berg Kasper, Rubino Rosella, Sartor Giovanni, Palmirani Monika, Wyner Adam, and Bench-Capon Trevor, 2007, OWL ontology of basic legal concepts (LKIF-Core). Deliverable 1.4, <http://www.estrellaproject.org/doc/D1.4-OWL-Ontology-of-Basic-Legal-Concepts.pdf> [Consulté le 19 Juillet 2015].
- Buitelaar Paul, 2010, Ontology-based Semantic Lexicons: Mapping between Terms and Object Descriptions, [in :] Hu-Ren Huang & al. (eds.) *Ontology and the Lexicon. A Natural Language Processing Perspective*, , 212-223. Cambridge : Cambridge University Press.
- Charte des droits fondamentaux de l'Union Européenne, 2000, Journal Officiel des Communautés européennes. http://www.europarl.europa.eu/charter/pdf/text_fr.pdf. [Consulté le 19 Juillet 2015].
- Declerck Gunnar, Audrey Baneyx, Xavier Aimé, Jean Charlet, 2012, A quoi servent les ontologies fondationnelles ? [in :] 23^{èmes} *Journées francophones d'Ingénierie des Connaissances* (IC 2012), June 2012, Paris, France. 67-82.

¹⁵ Cf. G. Gross (2012) qui parle de substantifs prédicatifs (*décision, gifle, réflexion, choix* etc.) et se place loin de primitives de Wierzbicka.

- Gross Gaston, 2012, *Manuel d'analyse linguistique. Approche sémantico-syntaxique du lexique*. Villeneuve d'Ascq : Presses Universitaires du Septentrion, coll. « Sens et structures ».
- Gruber Thomas, 2009, Ontology [in :] *Encyclopedia of Database Systems*, Ling Liu and M. Tamer Özsu (Eds.), Springer-Verlag. <http://tomgruber.org/writing/ontology-definition-2007.htm>. [Consulté le 19 Juillet 2015].
- Guarino Nicola, Oberle Daniel, Staab Rudi, 2009, What is an Ontology ? [in :] Staab S. & Studer R. (eds.) *Handbook on Ontologies*. 1-17. Springer Dordrecht Heidelberg.
- Hernandez Nathalie, 2005, *Ontologies de domaine pour la modélisation du contexte en recherche d'Information*. Thèse de doctorat, Université de Toulouse.
- Hirst Graeme, 2004, Ontology and the lexicon [in :] Staab S. & Studer R. (eds.). *Handbook on ontologies*, 209-230. Springer, Berlin Heidelberg.
- Hoekstra Rinke, Breuker Joost, Di Bello Marcello, and Boer Alexander, 2007, The LKIF Core ontology of basic legal concepts [in :] Casanovas, P. et al. (ed) *Proceedings of the Workshop on Legal Ontologies and Artificial Intelligence Techniques (LOAIT 2007)*. Estrella. <http://www.leibnizcenter.org/docs/hoekstra/Hoekstra-LOAIT07.pdf> [Consulté le 16 Août 2015].
- Lacot Xavier, 2005, Introduction à OWL, un langage XML d'ontologies Web. http://lacot.org/public/introduction_a_owl.pdf. [Consulté le 19 juillet 2015].
- Manuel de référence du langage B Version 1.8.5, 2002, Accessible on-line : http://www.irit.fr/~Martin.Strecker/Teaching/Common/B_manuel_de_reference.pdf. Consulté le 29 septembre 2015.
- Mondary Thibault. 2011, *Construction d'ontologies à partir de textes. L'apport de l'analyse de concepts formels*. Thèse de doctorat de l'Université Paris-Nord - Paris XIII.
- Morin Emmanuel, 1999, *Extraction de liens sémantiques entre termes à partir de corpus de textes techniques*. Thèse en Informatique, Université de Nantes, Décembre 99.
- Saint-Dizier Patrick, Viegas Evelyn (eds.), 1995, Computational Lexical Semantics. [Online]. *Studies in Natural Language Processing*. Cambridge: Cambridge University Press. Available from: Cambridge Books Online <<http://dx.doi.org/10.1017/CBO9780511527227>> [Consulté le 02 août 2015].
- Séguéla Patrick, 1999, *Construction de modèles de connaissances par analyse linguistique de relations lexicales dans les documents techniques*. Mars 2000. Thèse de doctorat. Toulouse III.
- Staab Steffen & Studer Rudi (eds.), 2009, *Handbook on Ontologies, International Handbooks on Information Systems*. Springer-Verlag Berlin .Heidelberg.
- Le Trésor de la Langue Française Informatisé*, Nancy, CNRS Editions, <http://atilf.atilf.fr/tlf.htm>
- Weinland Daniel, Ronfard Remi, Boyer Edmond (2006) Automatic Discovery of Action Taxonomies from Multiple Views (in :) Andrew Fitzgibbon and Camillo J. Taylor and Yann LeCun. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '06)*, Jun 2006, New York, United States. IEEE Computer Society.1639–1645.
- Wierzbicka Anna, 1969, *Dociekania semantyczne*, Wrocław: Ossolineum.
- Wierzbicka Anna, 1971, *Kocha, lubi, szanuje. Medytacje semantyczne*. Warszawa: Wiedza Po-wszechna.
- Wierzbicka Anna, 1975, Rozważania o częściach ciała [in :] E. Janus (Ed.) *Słownik i semantyka. Definicje semantyczne*. Wrocław: Ossolineum, pp. 91-103.
- Wierzbicka Anna, 1980, *Lingua Mentalis. The Semantics of Natural Language*. Sydney: Academic Press.
- Wierzbicka Anna, 1996, *Semantics: Primes and Universals*. Oxford: Oxford University Press.
- Wierzbicka Anna, 2004, Polish and Universal Grammar [in :] *Studies in Polish Linguistics* 1, 9-27.

SITOGRAFIE

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.100.9654&rep=rep1&type=pdf>
<http://www.crisco.unicaen.fr/des/>
<https://www.cyc.com>
http://www.java2s.com/Tutorial/Cpp/0480__STL-Algorithms-Modifying-sequence-operations/stdtransformwithpredicate.htm
<http://linguistics-ontology.org/version>
<http://www.estrellaproject.org>
<http://www.ksl.stanford.edu/projects/DAML/UNSPSC.daml>
<http://www.semantique-gdr.net/dico/index.php/Pr%C3%A9dicat>
<http://www.si.fr.atosorigin.com/sophia/comma/Htm/HomePage.htm>

KILKA REFLEKSJI NA TEMAT PRYMITYWÓW, PREDYKATÓW,
ZNACZENIA TERMINÓW W ONTOLOGIACH I W SŁOWNICTWIE OGÓLNYM

Streszczenie

Artykuł traktuje o « prymitywach » i « predykatach », które pojawiają się w ontologiach, w opisie języków naturalnych i w językach programowania. Sposób ich użycia zdaje się potwierdzać maksymę: *Si duo dicunt idem non est idem*. Jednakże używanie języków formalnych w ontologiach informatycznych oraz w tekstach językoznawczych każe zniuansować ten pogląd. Zarówno predykaty, jak i prymitywy pojawiają się w językoznawstwie i w informatyce w warstwie formalnej, nie zaś „leksykalnej” opisu. Pod tym ostatnim względem bliżsi ideału stworzenia listy prymitywów są językoznawcy, dążący niekiedy do zidentyfikowania pojęć nierozkładalnych, które powinny pozwolić na zrekonstruowanie reprezentacji semantycznej leksemów.

Rzut oka na *Kartę praw podstawowych Unii Europejskiej* jako na podstawę tekstową służącą do opracowania mikro-ontologii prowadzi do wniosku, że jej autorzy wprowadzają pojęcie *dignité humaine* ‘godność człowieka’ definicją projektującą. Znaczenie przypisywane temu pojęciu nie pokrywa się z żadnym z zarejestrowanych w języku.

Słowa kluczowe: predykaty, prymitywy, ontologia, terminy w ontologiach, słownictwo języka naturalnego.