

---

## **REDENE - Recherche documentaire assistée par ontologies de domaine adaptatives**

**Xavier Aimé<sup>\*\*\*</sup> — Frédéric Fürst<sup>\*\*</sup> — Pascale Kuntz<sup>\*</sup> — Francky Trichet<sup>\*</sup>**

*\* LINA - Laboratoire d'Informatique de Nantes Atlantique (UMR CNRS 6241)  
Équipe COD - Connaissance & Décisions  
2 rue de la Houssinière BP 92208 - F-44322 Nantes Cedex 03  
{pascale.kuntz,francky.trichet}@univ-nantes.fr*

*\*\* LARIA - Laboratoire de Recherche en Informatique d'Amiens (CNRS-FRE 2733)  
UPJV, 33 rue Saint Leu - 80039 Amiens Cedex 01  
frederic.furst@u-picardie.fr*

*\*\*\* Société TENNAXIA - 19 rue de Réaumur - 75003 Paris  
xaime@tennaxia.com*

---

*RÉSUMÉ. La subjectivité des connaissances devient une dimension incontournable qui se doit d'être intégrée et prise en compte dans le processus d'Ingénierie des Ontologies (IO). Le projet REDENE est fondé (i) sur une formalisation des résultats obtenus en psychologie cognitive sur le fonctionnement de la mémoire humaine - en tenant compte des hypothèses établies dans le domaine des neurosciences - et (ii) sur l'intégration et l'exploitation d'une telle formalisation au sein des processus de recherche d'information basés sur l'usage d'ontologies.*

*ABSTRACT. Subjective knowledge becomes an unavoidable subject, which must be integrated and developed in Ontological Engineering. REDENE is a project which is based on (i) the formalization of some results defined in Cognitive Psychology on human memory - by considering assumptions established in the field of the Neurosciences - and (ii) the integration and the exploitation of such a formalisation within the processes dedicated information retrieval based on the use of ontologies.*

*MOTS-CLÉS : Ontologie Pragmatifiée Vernaculaire du Domaine, Prototypicalité Conceptuelle et Lexicale, Catégorisation, Gradient, Personnalisation, Adaptation, Pragmatique.*

*KEYWORDS: Conceptual and Lexical prototypicality, Computational ontology, Typicality, Information Retrieval, Gradient, Personalisation, Pragmatics, Subjective knowledge.*

---

## 1. Introduction

D'un point de vue linguistique, la pragmatique s'intéresse aux éléments du langage dont la signification ne peut être comprise qu'en fonction d'un contexte d'interprétation donné. Dans le cadre des ontologies de domaine (qui sont des spécifications formelles de conceptualisations partagées (Gruber, 1993)), il s'agit d'enrichir la sémantique formelle intrinsèque à une ontologie de domaine (OD) à l'aide d'éléments caractéristiques d'un contexte de création ou d'usage comme la culture, le mode d'apprentissage, ou encore l'état émotionnel. Pour ce faire, nous prenons comme point de départ les ontologies telles qu'elles sont définies aujourd'hui au moyen du langage OWL. Nous y ajoutons deux gradients, dont le but est de modéliser une différence intuitive de degré de vérité dans le processus de catégorisation : (1) une pondération des liens « *is-a* » dans la hiérarchie des concepts, les **gradients de prototypicalité conceptuelle**, (2) une pondération des synonymes du label utilisé pour dénoter un concept donné, le **gradient de prototypicalité lexicale**. Ces gradients permettent de prendre en considération différents points de vue pour une même ontologie, et par là-même d'enrichir la sémantique formelle initiale intrinsèque à l'ontologie d'une pragmatique inhérente au contexte d'usage considéré (fonction de la culture, des modes d'apprentissage ou encore du contexte émotionnel). Le projet REDENE<sup>1</sup> se situe dans un cadre interdisciplinaire, et vise à améliorer la pertinence d'un moteur de recherche en s'appuyant sur des ontologies de domaine adaptatives qui tiennent compte à la fois de la psychologie, des neurosciences et de l'ingénierie des connaissances. Après avoir exposé une typologie des ontologies dans la section 2, le projet REDENE et les gradients de prototypicalité conceptuelle et de prototypicalité lexicale, fondés sur les processus cognitifs de catégorisation, sont explicités dans la sections 3.

## 2. Proposition d'une typologie des ontologies

Il ne s'agit pas ici de revisiter les longs débats philosophiques sur les « *connaissances absolues* » et les « *connaissances relatives* », mais de préciser les spécificités des connaissances sur lesquelles se base notre travail. Pour un domaine donné, une connaissance peut se définir (1) soit par rapport à un *référentiel propre à un individu*, où les connaissances seront qualifiées de *conviction*, (2) soit par rapport à un *référentiel propre à un endogroupe*, où les connaissances seront qualifiées de *connaissances relatives*. Chaque endogroupe possède ainsi un référentiel propre, plus ou moins différent de celui d'un autre endogroupe sur un même domaine. En réalité, toute objectivité individuelle est une subjectivité au niveau de l'endogroupe, et toute objectivité endogroupale est une subjectivité universelle. Par transitivité, toute objectivité individuelle est une subjectivité universelle.

---

1. Ce travail de recherche est réalisé dans le cadre d'un projet mené en collaboration et financé par la société Tennaxia (<http://www.tennaxia.com>), société de service et de conseils en veille juridique et réglementaire dans le domaine Hygiène, Sécurité, Environnement et Développement Durable (HSE-DD).

Notons que cette nouvelle proposition de typologie d'ontologies ne se fonde pas sur la nature même des connaissances considérées, mais sur le niveau de subjectivité/objectivité de ces connaissances. Elle est donc complémentaire des typologies d'ontologies déjà proposées en Ingénierie des Ontologies (IO), comme celle introduit dans (Gomez-Perez *et al.*, 2003).

Le niveau le plus élevé de cette typologie est le Savoir absolu. Comparable au concept universel des treillis, il est considéré comme l'union de toutes les ontologies de domaine. Une *ontologie de domaine* (OD), dans sa version computationnelle, est vue en tant que spécification formelle et explicite d'une conceptualisation partagée (Gruber, 1993). Elle correspond à une modélisation manipulable et intelligible des connaissances propres à un domaine particulier, par des agents tant logiciels qu'humains. Nous considérons les OD comme étant une conceptualisation universelle absolue et asymptotique (dans le sens d'une limite dont, en réalité, nous nous approchons de plus en plus sans jamais l'atteindre) pour un domaine particulier, et partagée (c'est à dire comprise et admise comme étant vraie) par la totalité des individus de l'univers - au sens de population dans un système donné fini. Il s'agit d'un sous-ensemble du Savoir absolu pour une thématique précise. Par rapport à notre typologie, nous qualifions une OD de Connaissance absolue pour un domaine, de partie du Savoir absolu.

Dans la pratique, définir une ontologie pour un domaine particulier consiste à établir une synthèse consensuelle de connaissances d'individus appartenant à un endogroupe propre à ce domaine et à un instant donné. Cette modélisation se fonde sur les connaissances relatives exprimées par leurs détenteurs - que ce soit au travers d'interviews ou de documents (textes, images, sons, etc.). Nous qualifions ces ontologies de domaine d'**ontologies vernaculaires du domaine** (OVD), le qualificatif *vernaculaire* provenant du latin *vernaculus* signifiant indigène. Dès lors que nous établissons des connaissances pour un même domaine suivant des axes propres à un endogroupe, c'est à dire dans le référentiel de cet endogroupe, nous passons au stade de connaissances relatives.

Rosch qualifie d'*écologique* ce processus d'élaboration de connaissance, dans le sens où il est certes fonction de l'endogroupe, mais avec une influence certaine du contexte dans lequel il évolue (Rosch, 1975). Ainsi, pour une OVD, un endogroupe donné et à un temps  $t$ , un minimum de trois facteurs définissent un contexte particulier : (1) la culture de l'endogroupe, (2) ses modes d'apprentissage, et (3) son contexte émotionnel (ou état mental). De ce fait, les ontologies peuvent être vues d'une manière pragmatique, suivant différents points de vue possibles, où chacun des points de vue, bien que ne remettant pas en cause la sémantique (formelle) inhérente à l'ontologie considérée, permet d'adapter le degré de vérité des relations de subsomption définies entre les concepts ; nous qualifions ce type d'ontologies d'**ontologie pragmatisée vernaculaire du domaine** (OPVD). Une OPVD est une OVD placée dans un contexte particulier.

La passage du stade d'OVD à OPVD se fait par le biais d'un processus de pragmatisme d'une OVD, qui consiste dans un premier temps à définir un contexte. Ce contexte est décrit par des valeurs affectées à des paramètres pondérant les facteurs de culture, de modes d'apprentissage et de contexte émotionnel (ou état mental) de l'endogroupe concerné. Ces trois valeurs sont utilisées par la suite au sein de calculs des gradients de prototypicalité (cf. section suivante). Ces derniers permettent de décorer les hiérarchies et par la même de représenter (au sens tracer) la pragmatique inhérente au contexte d'usage de l'OPVD considérée.

### 3. REDENE

Le projet REDENE (*Recherche d'information et Décisions dans le cadre d'une approche Ecologique, Neurale et Emotionnelle*) a pour vocation de s'inspirer - et de formaliser - des résultats obtenus en psychologie cognitive sur le fonctionnement (*i.e.* construction et utilisation) de la mémoire humaine, et d'appliquer - entre autre - cette formalisation à la recherche d'information, sur la base de la typologie des ontologies définie précédemment. REDENE est, à la base, un processus d'ingénierie ontologique avec comme objectif la recherche d'information à partir d'ontologies. Sans remettre en cause les différents travaux effectués à ce jour en IO, l'objectif de ce projet est véritablement de faire converger sur le thème de la Recherche d'Information par Ontologie les trois facettes de la cognition que sont la psychologie, les neurosciences et l'ingénierie des connaissances. Nous qualifions notre approche de (1) *écologique*, car inspirée de l'anthropologie cognitive comme activité cognitive des hommes différentes suivant la société à laquelle ils appartiennent, et leur culture, et (2) *émotionnelle*, car nous prenons - entre autre - en compte cette dimension tant dans la phase d'acquisition que d'utilisation des connaissances au travers du processus de pragmatisme des ontologies.

#### 3.1. Gradients de prototypicalité

Le mécanisme de catégorisation est un phénomène cognitif présent dès notre plus jeune âge, qui s'accompagne d'un classement implicite de représentativité des éléments composant chaque catégorie. Nous modélisons ces processus par l'intermédiaire de gradients de prototypicalité (pondération des liens *is-a* dans la hiérarchie catégorielle). Mais avant de présenter ces différents gradients, illustrons leur principe à travers un exemple simple et intuitif.

Soit le concept *Chien* avec deux relations d'attribut - *pouvoir aboyer* et *pouvoir ronger des os* - et considérons deux de ses sous-concepts : *Chien de défense* (un *Chien de défense est-un Chien*) et *Chien de berger* (un *Chien de berger est-un Chien*). Par héritage, chacun de ces sous-concepts hérite des relations d'attribut du concept père, relations auxquelles il peut rajouter ses propres éléments. Un *Chien de défense*

peut aboyer et ronger des os mais il sait également *protéger un individu, protéger des biens* et peut *être considéré comme une arme* (trois attributs supplémentaires par rapport au concept père). Un *Chien de berger* possède les qualités propres à tout chien, mais il sait en plus *rassembler des moutons* (une seule relation d'attribut supplémentaire par rapport au concept père). Ainsi, d'un point de vue structurel, nous avons le sous-concept *Chien de défense* qui va spécialiser le concept *Chien* au point de s'en distinguer très nettement (de la sorte, un *Chien de défense* n'est pas n'importe quel *Chien*). Par contre, le sous-concept *Chien de berger*, en ne se distinguant que très peu du concept *Chien*, va en être beaucoup plus représentatif. Cette appréciation, basée sur la structure (sur l'intention), va constituer la composante *objective* de notre gradient, dans le sens où les intentions - ainsi définies - forment un consensus définitionnel et stable au niveau de l'endogroupe. Mais, pourquoi, dans notre société occidentale et qui plus est française, citons-nous plus souvent le labrador que le rotweiller comme étant représentatif de la catégorie *Chien* ? Tout simplement par habitude, par la fréquence de rencontre dans notre univers, par la familiarité d'un élément, etc. Cette composante *subjective*, fondée uniquement sur nos percepts et sur leurs interprétations, sera calculée à partir de la fréquence d'apparition du concept (sous toutes ses formes) dans les supports relatifs au domaine - en sachant que chaque nouveau support, porteur d'informations, sera à même de modifier ce jugement.

Maintenant, posons-nous la question de savoir comment notre évaluation de la pertinence de nos connaissances face à une situation donnée peut varier en fonction de notre humeur. De multiples travaux, dont (Walker *et al.*, 1997), ont été réalisés en psychologie sur le sujet. La conclusion de ces travaux peut être résumée ainsi : lorsque nous sommes dans un état mental négatif, déprime par exemple, nous avons tendance à nous recentrer sur ce qui nous est cher ou sur ce qui nous paraît le plus fort émotionnellement parlant. Dans notre cas, nous aurons tendance à resserrer notre univers à tout ce qui nous est familier ; notre chien personnel (ou celui d'un proche) - qui au départ est forcément le plus caractéristique de la catégorie - devient le et quasi unique chien. A l'inverse, dans un état mental positif (amour, joie), nous sommes beaucoup plus ouverts dans notre jugement et nous acceptons bien plus facilement les éléments jugés jusque-là peu caractéristiques. Ainsi, jugement et état émotionnel se trouvent fortement liés.

Nos travaux reposent sur l'idée fondamentale que tous les concepts ne sont pas constitués de membres de représentativité équivalente par rapport à la catégorie qui les subsume, mais qu'ils comportent des membres qui sont de meilleurs représentants que d'autres (Kleiber, 2004). Ce phénomène est également applicable pour l'ensemble des termes désignant un concept (ensemble composé d'un terme saillant (Sowa, 1984), considéré comme le plus fort lexicalement pour désigner le concept, et d'une liste de synonymes). C'est sous cette hypothèse, validée par des travaux de psychologie cognitive tels que (Cordier, 1985), que nous proposons deux mesures - dont une présentation détaillée de la formalisation est proposée dans (Aimé *et al.*, 2008) : (1) le **gradient de prototypicalité conceptuelle**, qui correspond à une pondération des liens « *is-a* » dans la hiérarchie des concepts, et qui permet

de mesurer les différentes représentativités des sous-concepts ; (2) le **gradient de prototypicalité lexicale**, qui, pour un concept donné, correspond à une pondération des synonymes du terme saillant.

Nous qualifions ces mesures de « *gradients* » dans le sens où elles modélisent une différence intuitive de degré de vérité dans un processus de catégorisation. Elles s'appliquent à une OPVD, *i.e.* à un ensemble de concepts (voire de relations) pour un domaine donné, communément utilisé et admis par tous les individus d'un endogroupe propre à ce domaine, et placé dans un contexte particulier. Ce contexte va être fonction de trois composantes : la culture, l'apprentissage et l'état émotionnel. Si, pour ce dernier facteur, nous pouvons établir différentes valeurs d'après les travaux de Russel (Posner *et al.*, 2005), l'appréciation des deux premiers reste - à ce jour - encore à définir.

### 3.1.1. *Gradients de prototypicalité conceptuelle local (GPCL)*

Le gradient de prototypicalité conceptuelle local (GPCL) possède une composante intentionnelle qui repose sur la comparaison des propriétés conceptuelles du concept père et du concept fils, et une composante extensionnelle qui repose sur la comparaison des fréquences d'apparition des termes désignant les concepts père et fils dans l'univers de l'endogroupe considéré.

### 3.1.2. *Gradient de prototypicalité lexicale (GPL)*

Ce gradient a pour but de valuer le fait que tous les synonymes d'un terme considéré comme le terme saillant, *i.e.* le *label préconisé* pour un concept, n'ont pas forcément la même représentativité au sein de l'endogroupe. La question est en effet « *pourquoi nommons-nous plus le concept x avec le vocable y plutôt que z ?* » Pour définir ces variations de représentativité lexicale, nous reprenons le gradient calculé précédemment, à la différence près que nous n'allons pas utiliser la composante objective liée aux propriétés (du fait que nous sommes sur la même catégorie). Nous allons nous inspirer de la formule calculant le contenu en information d'un concept, en prenant pour évaluer ce gradient le rapport entre la fréquence d'utilisation de ce terme et la somme des fréquences de tous les termes du concept dans l'ensemble des documents propres à l'endogroupe.

## 3.2. *Evaluation d'ontologies et Recherche d'information*

Les gradients de prototypicalité peuvent contribuer au processus d'**évaluation d'ontologie** en tant qu'indicateurs de qualité de catégorisation, et par là même de qualité de la modélisation inhérente à une ontologie de domaine formalisée (au moyen d'OWL par exemple). D'un côté va se poser le problème des concepts jugés très peu typiques (une valeur de GPCL faible peut indiquer la nécessité de créer des concepts intermédiaires par exemple). Inversement, le fait qu'un concept soit jugé le plus typique d'une catégorie est-il conforme au jugement des experts du domaine.

Dans le cas où les ontologies ont été créées essentiellement à partir de textes, la composante subjective de la connaissance extraite est très forte, et il peut être intéressant d'équilibrer les points de vue en faisant varier les différents paramètres contextuels.

Les gradients de prototypicalité peuvent également contribuer au processus de **recherche d'information**, lequel va s'articuler autour (1) d'une OPVD qui va nous guider sur le focus de recherche en permettant une extension des requêtes conformément aux connaissances relatives de l'ontologie de domaine, (2) de la prototypicalité conceptuelle qui va nous guider sur l'ordre de pertinence des résultats et (3) de la prototypicalité lexicale qui va nous aider non seulement dans la reformulation des requêtes mais également sur l'ordre de pertinence des résultats. Les gradients de prototypicalité peuvent ainsi permettre de classer les résultats d'une requête (sous-entendue d'une requête dite *étendue*) suivant un ordre de pertinence, en plaçant au premier plan les éléments les plus représentatifs d'une catégorie ou d'un terme donné. Nous pouvons également adjoindre à cette notion de classement (point de vue *qualitatif*) le fait de pouvoir proportionner le nombre de réponses (point de vue *quantitatif*) en fonction de la valeur des gradients. La recherche d'information devient ainsi adaptable, voire personnalisable, puisqu'il est possible de faire varier les contextes de point de vue (*i.e.* privilégier l'intention à l'extension ou inversement, travailler avec des états mentaux différents, etc.).

#### 4. Conclusion

La finalité de nos travaux, focalisés sur la notion d'ontologie pragmatifiée vernaculaire du domaine, est - en nous fondant sur des travaux de psychologie cognitive, pour certains validés par la neurophysiologie - de prendre en compte la subjectivité de la connaissance via sa spécificité à un endogroupe et un domaine, son aspect écologique, et l'importance de son contexte émotionnel. Cet objectif nous conduit (1) à étudier la dimension pragmatique d'une ontologie, et (2) à dresser une typologie des ontologies en partant du Savoir absolu pour en arriver aux Connaissances fortement relatives. En nous inspirant notamment des travaux d'E. Rosch sur la prototypicalité, nous avons développé deux mesures identifiant deux gradients de prototypicalité, l'un conceptuel qui mesure la représentativité relative d'un concept par rapport à ses concepts frères, l'autre lexical, qui mesure la représentativité relative d'un terme par rapport à ses synonymes. Bien évidemment, ces gradients ne modifient en rien la sémantique formelle inhérente à l'ontologie considérée ; les liens de subsomption restent valides. Ces gradients ne sont que le reflet de la prise en compte de la pragmatique, à savoir une sur-couche pragmatique sur la sémantique. Ils sont l'une des bases sur lequel repose le projet REDENE qui, dans un premier temps, a pour vocation la recherche d'information contextualisée à partir d'OPVD. Ces travaux se poursuivent actuellement sur l'affinement des gradients, en se référant à de nombreux travaux de psychologie cognitive et sociale. Une extension de ces gradients aux hiérarchies de relation est également en projet. Enfin, une application de ces résultats est actuellement en cours de déve-

loppement dans le cadre expérimental d'une recherche d'information au sein de textes réglementaires dédiés au domaine "Hygiène, Sécurité et Environnement (HSE)".

## 5. Bibliographie

- Aimé X., Fürst F., Kuntz P., Trichet F., « Gradients de prototypicalité conceptuelle et lexicale », *Revue des Nouvelles Technologies de l'Information - Extraction et Gestion des Connaissances (EGC'2008)*, vol. 1 (11), Cépaduès, p. 127-132, 2008. ISBN 978.2.85428.819.3.
- Cordier F., « Formal and locative categories. Are there typical instance ? », *Psychologica Belgica*, vol. XXV, n° 2, p. 115-125, 1985.
- Gomez-Perez A., Fernandez-Lopez M., Corcho O., *Ontological Engineering*, Springer, Advanced Information and Knowledge Processing, 2003.
- Gruber T., « Toward principles for the design of ontologies used for knowledge sharing », in , N. Guarino, , R. Poli (eds), *Formal Ontology in Conceptual Analysis and Knowledge Representation*, Kluwer Academic Publishers, Deventer, The Netherlands, 1993.
- Kleiber G., *La sémantique du prototype*, Presses Universitaire de France - coll. Linguistique Nouvelle, mars, 2004. ISBN 2-1304-2837-1, 2e édition.
- Posner J., Russel J., Peterson B., « The circumplex model of affect : An integrative approach to affective neuroscience, cognitive development, and psychopathology », *Development and Psychopathology*, vol. 17, p. 715-734, 2005.
- Rosch E., « Cognitive Representations of Semantic Categories », *Journal of Experimental Psychology*, n° 104, p. 192-233, 1975.
- Sowa J. F., *Conceptual structures : Information Processing in Mind and Machine*, Addison-Wisley Publishing Company, 1984. ISBN 0-201-14472-7.
- Walker W., Vogl R., Thompson C., « Autobiographical memory : Unpleasantness fades faster than pleasantness », *Applied Cognitive Psychology*, n° 11, p. 399-414, 1997.