Correction d'ontologies construites à partir de la structure de documents

Mouna Kamel, Nathalie Aussenac-Gilles, Marion Laignelet

Correction d'ontologies construites à partir de la structure de documents

Mouna Kamel, Nathalie Aussenac-Gilles, Marion Laignelet

IRIT, Université de Toulouse,
{kamel, aussenac, laignele}@irit.fr

Résumé : Les logiciels de construction d'ontologies à partir de textes réalisent une interprétation fixée a priori du contenu des textes, qu'un expert du domaine ou une ontologie doit vérifier. Or une étude précise des limites des techniques d'analyse des textes permet de guider la correction de l'ontologie apprise en définissant des règles d'aide à la correction. Ces règles attirent l'attention de l'ontographe sur des parties d'ontologie contenant des « anomalies » et tiennent compte du texte d'origine et de l'analyse réalisée pour proposer des corrections. Dans cet article, nous illustrons la notion de règle de correction dans le cas où les connaissances apprises viennent de l'exploitation de structures énumératives parallèles présentes en corpus.

Mots-clés : Ingénierie des connaissances, construction automatique d'ontologies, correction d'ontologies.

1 Introduction

La construction d'ontologies à partir de textes suppose d'appliquer des logiciels de traitement automatique des langues (TAL) pour repérer des indices de connaissances en vue de les représenter et de les organiser au sein d'une ontologie. Des plateformes comme Terminae (Aussenac-Gilles *et al.*, 2008) ou Text2Onto (Cimiano & Völker, 2005) existent et font appel à différents types de logiciels (analyseurs, extracteurs de termes ou de relations, logiciels de clustering de termes ou de recherche de segments répétés, etc.). Les capacités des logiciels de TAL et la nature même des textes (ils contiennent rarement toutes les informations requises pour définir une ontologie et sont sujets à plusieurs interprétations) obligent à prévoir une intervention humaine dans ce processus. Une ou plusieurs phases de sélection et d'interprétation des données tirées des textes ont pour but de décider si elles seront intégrées ou non à l'ontologie et sous quelle forme. Cette interprétation est fixée *a priori* dans le cas de la construction automatique d'ontologie (*ontology learning*) (Buitelaar *et al.*, 2005) (Cimiano 2007), (Maedche, 2002). Pour rattraper les erreurs possibles, l'ontologie apprise est alors considérée comme un brouillon que l'ontographe va ajuster et corriger (Brewster *et al.*, 2002).

Dans cet article, nous présentons des règles de correction d'une ontologie construite automatiquement à partir de la structure textuelle de documents selon la

méthode décrite dans (Kamel & Aussenac-Gilles, 2009). Nous illustrons notre propos sur l'étude des structures énumératives parallèles dans des documents de spécification de bases de données géographiques (références pour la localisation d'information relative aux problèmes d'aménagement, d'environnement, d'urbanisme) collectés dans le cadre du projet GEONTO[1]. Les structures énumératives sont repérées de manière automatique à partir d'éléments de mise en forme (alignements, indentation,…), de symboles (tirets, séparateurs, ponctuation), *etc*. rendus explicites par des balises XML. Ces structures permettent d'identifier des étiquettes de concepts et des relations sémantiques. Des traitements automatiques ont été définis pour les exploiter, faisant des hypothèses fortes sur la sémantique des relations et la manière de repérer les noms des concepts. Or la variabilité des énumérations en corpus conduit à des ajouts de connaissances parfois incorrects dans l'ontologie. L'étape de correction consiste à identifier ces cas et à proposer à l'ontographe les corrections appropriées.

L'article situe d'abord notre travail dans le panorama actuel de la recherche sur la construction d'ontologies à partir de textes, pour souligner l'originalité d'une démarche visant à corriger une ontologie « apprise » (§2). Nous détaillons ensuite notre analyse des structures énumératives parallèles présentes dans le corpus d'étude (§3). Nous présentons ensuite l'approche mise en œuvre sur ce corpus, en particulier les interprétations faites *a priori* des énumérations, leurs limites et les règles définies pour en corriger les erreurs (§4). Enfin, les résultats obtenus sur ce corpus (présentés en §5) permettent de dégager des perspectives (§6).

2 Correction d'ontologies

Nous appelons « correction d'ontologie » le fait de reprendre systématiquement les concepts et relations d'une ontologie de manière à vérifier des critères définis a priori et de la faire évoluer pour qu'elle respecte ces critères. Ces critères peuvent s'appuyer, comme dans la méthode OntoClean, sur des principes sémantiques, dits méta-propriétés des concepts, pour suggérer de modifier la hiérarchie des concepts et les relations dans l'ontologie (Guarino & Welty, 2004).

D'autres travaux font appel à la syntaxe du langage de représentation et à la nature des concepts. C'est ce qu'expriment les patrons de conception d'ontologie (*ontology design patterns*) (Gangemi, 2005) : ils caractérisent une « bonne manière » de représenter certains types de connaissances. Cette notion a été élargie et ont été proposés[2] des patrons s'appuyant sur les propriétés des formalismes (OWL en particulier) ; des patrons exploitant la logique pour assurer une cohérence formelle ; des patrons « de contenu » suggérant de bonnes pratiques de structuration pour des concepts et des relations particuliers, par exemple pour représenter les objets prenant part à un événement ; des patrons liés à la forme, au nommage ou à l'annotation des éléments de l'ontologie. Ces patrons de présentation doivent améliorer la lisibilité et

[1] Projet ANR-07-MDCO-005-01 GEONTO http://geonto.lri.fr/

[2] *cf.* le site http://ontologydesignpatterns.org

l'utilisabilité de l'ontologie en garantissant des choix homogènes dans les identifiants des concepts et l'utilisation systématique des labels, ces patrons étant les seuls en lien direct avec les termes. Dans le cadre de la construction d'ontologies à partir de textes, des travaux récents articulent des patrons de conception d'ontologie à des patrons lexico-syntaxique caractérisant des relations sémantiques (Aguado de Cea *et al.,* 2008). Le patron de conception indique alors comment formaliser le fragment d'ontologie qui peut être défini à partir de phrases reconnues à l'aide de ce patron.

A l'inverse, des patrons peuvent caractériser des configurations à éviter, des structures de l'ontologie considérées comme erronées parce qu'elles correspondent à des incohérences logiques ou à des erreurs non détectables formellement (Corcho *et al.,* 2009). Lorsqu'ils sont reconnus, ces patrons conduisent à suggérer de corriger des parties de l'ontologie. Ils jouent le rôle de guides de bonne pratique.

Avec un objectif analogue, nous avons choisi de définir des règles de correction d'ontologie qui caractérisent (sous forme de patrons) des configurations de l'ontologie « à corriger » et dont la conclusion aiguille vers les parties de l'ontologie à modifier. Ces patrons sont « non-formels » dans la mesure où ils repèrent des parties d'ontologie qui posent problème à l'interprétation humaine et non au système formel. Ces règles de correction s'utilisent sur une ontologie apprise automatiquement à partir de textes, et s'appuient sur des connaissances linguistiques d'une part, et sur les limites connues des techniques d'extraction utilisées d'autre part.

3 Structures énumératives et ontologies

Les méthodes classiques de construction d'ontologies à partir de textes privilégient l'analyse du texte brut, le but étant de cibler les termes qui désignent les concepts, et les relations sémantiques qui lient ces concepts. Une des méthodes pour repérer ces relations consiste à utiliser des marqueurs linguistiques qui caractérisent l'expression dans la langue au sein d'une même phrase. Or l'analyse globale d'un texte montre l'existence de relations entre unités textuelles n'appartenant pas à la même phrase, et susceptibles de s'étendre au paragraphe. Ces liens se manifestent par les relations du discours (Asher *et al.,* 2001), par l'utilisation de phénomènes linguistiques tels que les anaphores et les ellipses (Cornish, 2006), ou encore à travers la structure matérielle du texte (Luc & Virbel, 2001).

À terme, notre volonté est d'étendre la recherche de relations sémantiques aux relations supportées par la structure explicite des textes, qu'elle soit matérialisée par des balises XML, par la mise en forme ou par des marques de discours. La structure du texte joue un rôle important dans son interprétation car elle donne "forme et sens au contenu" (Jacques, 2005). Nous avons choisi de nous focaliser dans cet article sur les structures énumératives.

Une étude approfondie sur les structures énumératives a permis de spécifier les propriétés de ces objets textuels (Luc, 2001) et d'en donner des définitions (Virbel, 1999). Une ***énumération*** est un ensemble d'items qui peuvent entretenir des relations sémantiques entre eux. Un ***item*** est une entité coénumérée, discernable par une

marque typographique, dispositionnelle, lexico-syntaxique ou une combinaison de ces marques. Une ***amorce*** est une phrase introduisant une énumération, qui est repérable par une marque lexico-syntaxique, dispositionnelle ou syntaxique.

Luc (2001) propose une typologie des énumérations au sein de laquelle il définit les *énumérations parallèles* comme étant paradigmatiques (*i.e.* tous les items sont fonctionnellement équivalents), visuellement homogènes (*i.e.* tous les items sont visuellement équivalents) et isolées (*i.e.* aucun item n'est en relation avec une unité textuelle extérieure à l'énumération et l'énumération ne contient aucune autre énumération). Dès qu'une de ces conditions n'est pas vérifiée, la structure est dite non parallèle. La relation sémantique portée par la structure énumérative parallèle se situe entre l'amorce et chacun des items. Le plus souvent hiérarchique, cette relation est fréquemment de hypéronymique ou méronymique. Les énumérations parallèles regroupent des items fonctionnellement équivalents, ce qui peut être traduit, en termes d'ontologie, par des concepts de même niveau, liés par la même relation au concept présent dans l'amorce, comme illustré en figure 1. Cette analyse forme la base d'un processus d'interprétation automatique, qui permet de rendre compte de chaque structure énumérative par un fragment d'ontologie.

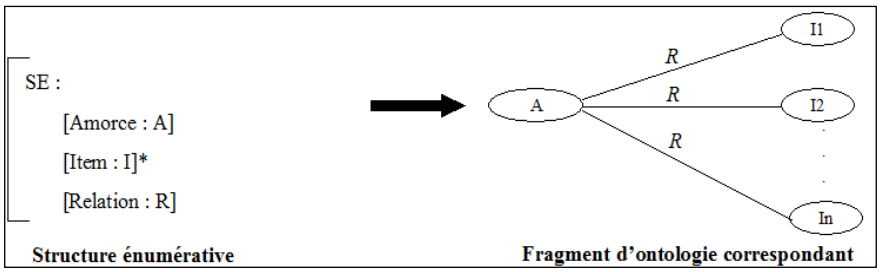

Fig. 1 – Représentation d'une structure énumérative (à gauche de la figure) et du fragment d'ontologie associé (partie droite)

4 Construction et correction d'ontologie

Nous proposons de construire une ontologie en exploitant, outre la langue naturelle présente dans les textes, leur structure et les informations présentes dans les segments de type énumération. L'automatisation de ce processus, décrite ci-après, conduisant à des fragments d'ontologie parfois non valides, les règles de correction ont pour but de repérer les cas posant problème et d'en suggérer des modifications.

4.1 Construction automatique d'ontologie à partir de la structure textuelle

Nous illustrons notre approche dans le cas où les textes analysés sont des documents de spécification de base de données en cartographie, tous conformes au

même schéma XML. Ces spécifications décrivent de nombreux concepts du domaine et leurs propriétés, en vue de guider le bon stockage dans des bases de données d'entités cartographiques. Ce travail se base sur la méthode proposée dans (Kamel & Aussenac-Gilles, 2009a) pour construire une ontologie en tirant profit de la structure de documents Cette approche exploite ainsi la structure explicitée par les balises XML, leur sémantique et leur hiérarchie, la structure accessible par certains éléments de mise en forme, et le texte en langage naturel. Elle est comparable à l'exploitation de la structure XML proposée dans (Role & Rousse, 2006). Un extrait du document analysé au format XML est présenté dans la figure 2.

```
<className="tronçon de chemin">
 <className>Tronçon de chemin</className>
        <description type="definition">Voie de communication terrestre non ferrée …</description>
        <description type="extensionalDefinition"> nature</description>
        <description type="selectionPrinciple">Voir les différentes valeurs de <Nature>… </description>
        <attributes>
                <attributeName>Nature</attributeName>
                <valueType>Énuméré</valueType>
                <description type="definition">Permet de distinguer plusieurs types de voies de comm.
                … </description>
                <enumeratedValues>
                        <valueName>Sentier</valueName>
                        <description type="definition">Chemin étroit ne permettant pas le passage
                        de véhicules.</description>
                        <description type="extensionalDefinition"> Allée piétonne (étroite) | Piste de
                        cross | Ruelle étroite | Sentier </description>
                        <description type="selectionPrinciple">Seuls les principaux sentiers sont
                        inclus.</description>
                </enumeratedValues>
        </attributes>
</className>
…
```

Fig. 2 – Extrait des spécifications de BDTopo au format XML

Ce corpus comporte des balises (<*packageName*>, <*className*>, <*attributeName*>, <*valueName*>) qui encadrent des termes et non des paragraphes, et qui sont souvent des étiquettes de concepts. L'imbrication des balises reflète des relations sémantiques entre les concepts associés aux termes marqués par ces balises. Ce sont le plus souvent des relations hiérarchiques : par exemple, la relation EST-UN existe entre un concept marqué par <*packageName*> et celui marqué par le <*className*> qui l'englobe. On retrouve le même type de relation entre <*attributeName*> et <*valueName*>. L'imbrication traduit dans certains cas le fait qu'un terme renvoie à une propriété d'un concept ou à une relation non hiérarchique (relation entre <*className*> et <*attributeName*>). Une analyse préalable des

documents de spécification, de leur schéma et des consignes de rédaction qui les concernent a permis d'inventorier toutes les balises permettant de définir des concepts et leurs étiquettes, ainsi que la manière d'interpréter l'imbrication entre balises.

Dans cette approche, la notion de *patron structurel* (Kamel et Aussenac, 2009b) rend compte de la sémantique portée par un élément de structure. Un patron structurel repère les termes concernés en corpus et produit un fragment d'ontologie selon cette sémantique. Il spécifie les balises introduisant les termes donnant lieu à la définition de concepts, la localisation de ces balises dans le document ainsi que la relation sémantique entre ces termes. Ces informations sont issues d'une analyse manuelle en corpus. Enfin, le patron génère le fragment d'ontologie correspondant. L'application successive de patrons structurels offre la possibilité d'exploiter toute la structure du document et d'avoir une approche incrémentale pour parvenir à une ontologie. L'ontologie obtenue contient 1196 concepts.

4.2 Les structures énumératives du corpus

Les structures énumératives présentes dans ces documents possèdent des propriétés lexico-syntaxiques, typographiques et dispositionnelles différentes des structures énumératives classiques tout en conservant leurs caractéristiques sémantiques. Les composants de ces structures (amorce, items) sont introduits par des balises spécifiques. Les amorces ne sont pas des phrases mais des termes. Les items correspondent soit à des syntagmes nominaux ou adjectivaux, soit à des énumérations, chaque item de l'énumération étant un syntagme nominal ou adjectival. Nous avons donc élargi la notion de parallélisme aux structures énumératives possédant des items renfermant une énumération ; nous les appellerons par la suite structures énumératives parallèles. Par ailleurs, les amorces doivent être parfois calculées dans la mesure où elles sont exprimées à travers différents éléments de structure. L'intervention d'un expert est alors nécessaire pour préciser les balises, spécifier la nature de la relation entre l'amorce et les items, et parfois pour caractériser l'amorce.

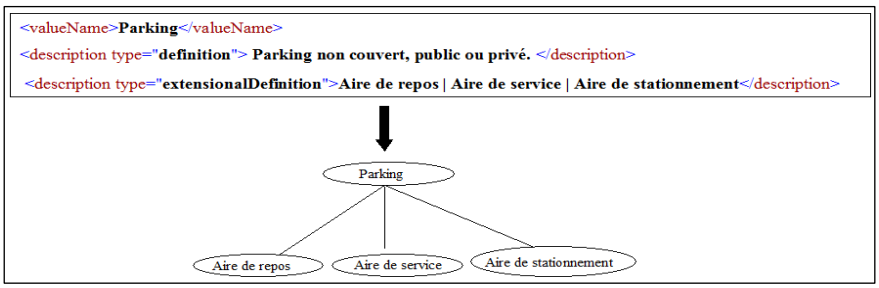

Fig. 3 –Traduction d'une structure énumérative extraite du corpus GEONTO

La figure 3 montre un exemple de structure énumérative extraite du corpus GEONTO. L'expert identifie d'une part les balises <*valueName*> et <*description type="extensionalDefinition"*> comme introduisant respectivement l'amorce et les items, et d'autre part, la relation *est-un* entre les termes marqués par ces balises. Le

traitement automatique des structures énumératives peut produire des fragments d'ontologie incorrects d'un point de vue conceptuel. Cela est dû à l'usage possible de variations de la langue au sein des structures énumératives.

Table 1. Exemples d'énumérations pour lesquelles la règle de traitement des énumérations produit des fragments d'ontologie à vérifier ou corriger selon les corrections proposées.

Items de forme adjectivale	Items numériques ou quantitatifs
<attributeName>**Régime des eaux**</attributeName> <enumeratedValues> <valueName="**Intermittent**" /> <valueName="**Permanent**" /> </enumeratedValues>	<attributeName> **Voltage** </attributeName> <enumeratedValues> < valueName ="**63 kV**" /> < valueName ="**90 kV**"/> < valueName ="**150 kV**"/> </enumeratedValues>
Correction proposée : Associer aux adjectifs le nom du concept de l'amorce	Correction proposée : Considérer le concept de l'amorce comme une propriété valuée.
Items énumératifs	Items booléens
<attributeName> **Transport par câble** </attributeName> <enumeratedValues> < valueName ="**câble transporteur**"/> < valueName ="**Télécabine, téléphérique, téléski**"/> </enumeratedValues>	<attributeName> **Multi-canton** </attributeName> <enumeratedValues> < valueName ="**oui**" /> < valueName ="**non**" /> </enumeratedValues>
Correction proposée : re-découper la liste présente en autant de concepts fils du concept de l'amorce.	Corrections proposées : soit faire du concept de l'amorce un attribut, soit lui attribuer 2 concepts fils.
Item phrastique	Inclusion lexicale
<attributeName> **Tronçon de voie ferrée électrifié** </attributeName> <enumeratedValues> < valueName ="**Ligne ferroviaire n'utilisant pas l'énergie électrique pour la propulsion des locomotives** " /> < valueName ="**Ligne ferroviaire utilisant l'énergie électrique pour la propulsion des locomotives**" /> </enumeratedValues>	<value> <valueName> **Autre Classement** </ valueName > < description type= "extensionalDefinition"> **Voies goudronnées (voies communales, chemins ruraux ou voies privées) | Rues | Rues piétonnes**> </description> </value>
Correction prévue : traiter le texte à l'aide de patrons.	Correction proposée : créer le concept Rues piétonnes comme fils de Rues.

4.3 Erreurs liées à l'interprétation des structures énumératives

L'exemple d'énumération de la figure 3 est "idéal" dans la mesure où chaque item est un syntagme nominal, et peut être assimilé à un nom de concept. Il existe cependant des énumérations non conformes à ce schéma (Table 1) parce qu'elles produisent des concepts dont les labels ne font pas sens. Ainsi, on trouve des items de forme adjectivale, numérique ou quantitative, énumérative, booléenne, phrastique, des inclusions syntagmatiques entre deux items, etc. qui, tels quels, ne sont pas de bons candidats à devenir des labels de concepts de l'ontologie. Nous avons fait un inventaire des cas où l'interprétation a priori de l'énumération engendre des erreurs dans la modélisation.

4.4 Règles de correction

Pour chacun de ces cas, nous proposons des règles de correction. La règle comporte deux parties : un patron qui caractérise une configuration de l'ontologie non valide d'un point de vue de l'interprétation des étiquettes de concepts, et une conclusion qui suggère à l'ontrographe des modifications de l'ontologie (ajout/suppression de concept, de relation, etc.). Plus précisément, nous nous intéressons aux patrons repérant les cas liés à l'interprétation des structures énumératives précédentes. Ces patrons s'intéressent à la forme lexico-syntaxique des labels des concepts appris. Pour cela, nous utilisons un étiqueteur morpho-syntaxique, Tree-Tagger, et des ressources externes (un lexique d'unités de mesures que nous avons créé). Nous décrivons ci-dessous trois patrons de correction relatifs aux anomalies les plus fréquemment observées dans l'ontologie.

4.4.1 Item adjectival

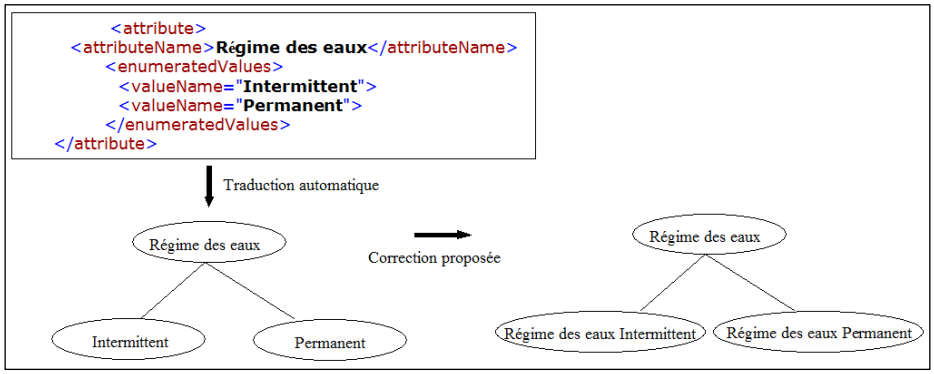

Fig. 4 –Exemple de correction des concepts dont les labels sont des adjectifs

Si le label d'un concept C est de type adjectival (Fig.4), alors un nouveau concept C' est créé dont le label est le résultat de la concaténation du label du concept père de

C et du label de C. C' est rajouté comme sous-concept du concept père de C, la relation entre C et son concept père est effacée, et le concept C est supprimé.

4.4.2 Item numérique ou quantitatif

Nous traitons pour le moment les items qui représentent des valeurs numériques, éventuellement associées à des unités de mesure (Fig.5). Nous avons créé un lexique qui répertorie les unités de mesure, et qui nous sert à annoter les items.

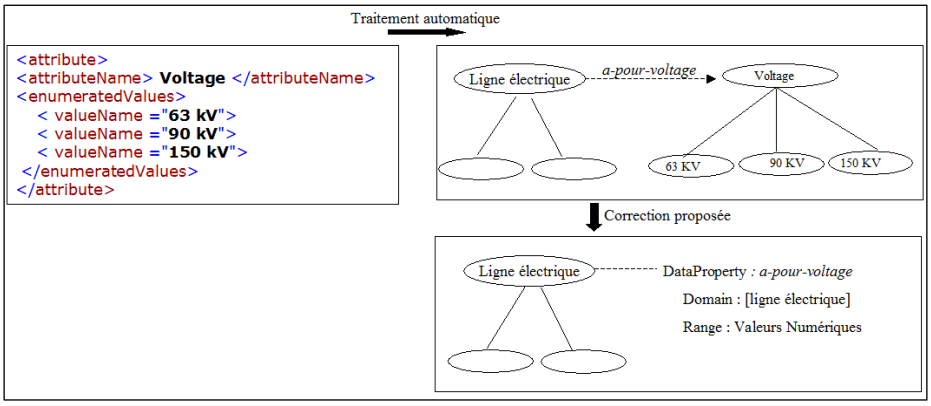

Fig. 5 –Correction de concepts ayant des valeurs quantitatives

La correction proposée consiste alors à supprimer la relation entre le concept C (relatif à l'item quantitatif ou numérique) et son concept père, puis à supprimer le concept C. Si le concept père n'a alors plus de fils, et s'il est le co-domaine d'une relation de type *ObjectProperty*, cette relation est changée en une relation de type *DataProperty*, ayant pour domaine le domaine de valeurs associé au concept père, et le concept père est supprimé.

4.4.3 Item énumératif

Si le label d'un concept C correspond à une énumération (Fig. 6), alors une classe C' est créée pour chaque élément de l'énumération. La relation *est-un* entre C' et le concept père de C est établie. Une fois l'énumération traitée, le lien entre C et sa classe père est rompu, et la classe C détruite.

Afin de ne pas perdre la sémantique véhiculée par le fait que les éléments de l'item énumératif aient été placés au sein d'un même item, nous créons un concept *A_Définir* qui sera labellisé par l'expert. Ce concept aura comme concepts fils, les concepts correspondant aux éléments de la liste contenue dans l'item.

De la même façon, nous avons défini des patrons de correction pour les items booléens et les énumérations où il y a inclusion lexicale.

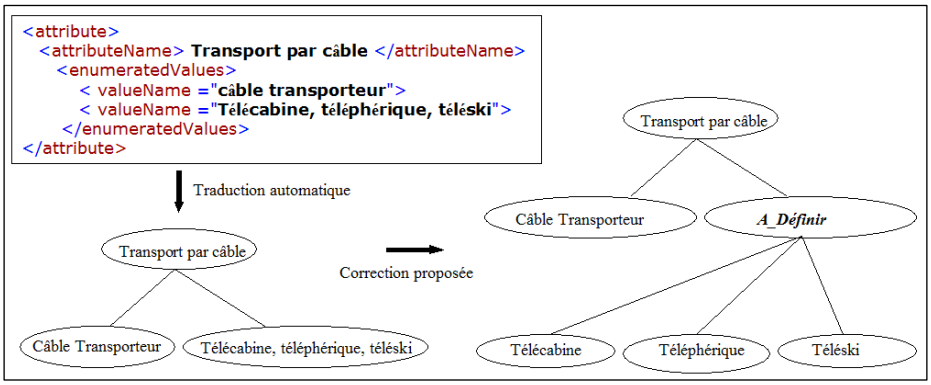

Fig. 6 –Proposition pour corriger des concepts correspondants à une liste

5 Résultats

Ces règles ont été expérimentées sur l'ontologie obtenue automatiquement à partir des spécifications de la base de données BDTopo. Le tableau de la Fig. 8 récapitule, pour chacun des cas que nous venons de répertorier, le nombre d'occurrences de ce cas dans le document et le nombre de cas reconnus par les règles de correction.

Items à corriger	**Nombre d'occurrences**	**Nombre d'occurrences reconnues par les patrons**
Adjectival	18	11
Numérique/Quantitatif	18	18
Phrastique	2	2
Enumératif	239	224
Booléen	4	4

Fig. 7 –Résultats des corrections

Plusieurs raisons expliquent ces résultats. Les items adjectivaux qui ne sont pas proposés à la correction sont des termes ambigus dans la mesure où ils désignent à la fois un nom et un adjectif, comme par exemple le terme *religieux*. Comme TreeTagger attribue l'étiquette NOM à ces termes, les labels ne s'unifient pas au patron. Les items énumératifs sont composés soit d'items séparés par un caractère spécial ("|" ou "/"), soit d'items séparés par la virgule et/ou la disjonction "ou". Dans cette évaluation, ce dernier type d'item énumératif ainsi que les items phrastiques ne sont pas traités.

Nous avons constaté par ailleurs que l'ordre d'application des patrons de correction est important. Des items peuvent combiner plusieurs phénomènes linguistiques, comme une liste d'adjectifs.

```
<attribute>   <attributeName> Bâtiment </attributeName>
  <enumeratedValues> < valueName ="Industriel, agricole ou commercial"> </valueName>
                     < valueName ="Religieux"> </valueName>
  </enumeratedValues> </attribute>
```

Fig. 8 – Enumération nécessitant la combinaison de plusieurs règles de correction

Dans le cas de la figure 8, il convient d'exécuter en premier lieu la règle concernant l'item énumératif puis la règle traitant les items adjectivaux.

6 Perspectives

Nous avons montré que dans le cas favorable où des structures énumératives parallèles peuvent être repérées automatiquement dans des textes au format XML, grâce aux balises caractéristiques de leurs amorces et items, l'exploitation automatique de ces structures mène dans certains cas à produire des fragments d'ontologies incorrects. Nous avons défini des règles de correction pour certains de ces cas, visant à proposer à l'expert ontographe des corrections. Les évolutions prévues à ce travail sont les suivantes. Il s'agit tout d'abord d'améliorer la fiabilité des patrons, en introduisant par exemple des heuristiques dans le traitement des items adjectivaux. La définition de patrons de correction pour les items phrastiques est également envisagée. Cela consisterait à extraire de l'item la clause principale, lorsque celui-ci n'est composé que d'une seule phrase. Des stratégies sont alors à définir pour traiter les items constitués de plusieurs phrases. Enfin la validation de cette approche sur un corpus de structures énumératives repérables par leur structure matérielle permettra de généraliser notre méthode et de l'appliquer par la suite à tout type de document. Enfin, une dernière validation sera de tester ces règles sur des ontologies qui n'auraient pas été construites à partir de la structure et dans laquelle on souhaite vérifier systématiquement les labels des concepts, de manière à avoir des formes linguistiques cohérentes.

Références[3]

AGUADO DE CEA G., GÓMEZ-PÉREZ A., MONTIEL-PONSODA E., SUÁREZ-FIGUEROA M. (2008), Natural Language-Based Approach for Helping in the Reuse of Ontology Design Patterns. In A Gangemi, J. Euzenat (Eds.): *Knowledge Engineering: Practice and Patterns, 16th International Conference, EKAW 200*8, Proceedings. LNCS 5268 Springer Verlag, 32-47.

ASHER N., BUSQUET J., VIEU L. (2001), La SDRT: une approche de la cohérence du discours dans la tradition de la sémantique dynamique. *Verbum* 23, 73-101.

AUSSENAC-GILLES N., DESPRES S., SZULMAN S. (2008), The TERMINAE Method and Platform for Ontology Engineering from texts. *Bridging the Gap between Text and Knowledge* -

[3] Toutes les URL mentionnées ont été testées en avril 2010

Selected Contributions to Ontology Learning and Population from Text. P. Buitelaar, P. Cimiano (Eds.), IOS Press, 199-223.

BREWSTER C., CIRAVEGNA F., WILKS Y. (2002). User-centered ontology learning for knowledge management. In B. Andersson, M. Bergholtz, and P. Johannesson (eds.), *proceedings* of *NLDB*, LNCS 2553, Springer Verlag, 203–207.

BUITELAAR P., CIMIANO P., MAGNINI B. (2005). *Ontology Learning From Text: Methods, Evaluation and Applications.* IOS Press.

CIMIANO P., VÖLKER J. (2005). Text2Onto - A Framework for Ontology Learning and Data-driven Change Discovery. In A. Montoyo, R. Munoz, E. Metais, *Proceedings of the 10th International Conference on Applications of Natural Language to Information Systems (NLDB*), LNCS 3513, Springer Verlag, 227-238.

CIMIANO P. (2007). *Ontology Learning and Population from Text. Algorithms, evaluation and applications*. Springer, Berlin.

CORCHO O., ROUSSEY C., VILCHES BLAZQUEZ L.M. (2009). Catalogue of Anti-Patterns for formal Ontology debugging, Dans *Atelier Construction d'ontologies : vers un guide des bonnes pratiques, Plate-forme AFIA 2009*, Hammamet (Tunisie). En ligne : http://liris.cnrs.fr/publis/?id=3990

CORNISH F. (2006). Relations de cohérence en discours : critères de reconnaissance, caractérisation et articulation cohésion-cohérence. *Corela*, D. Legallois (Ed.), numéro spécial, Organisation des textes et cohérence des discours. En ligne : http://edel.univ-poitiers.fr/corela/document.php?id=1280

GANGEMI A. (2005). Ontology Design Patterns for Semantic Web Content, *The Semantic Web – ISWC* 2005, Vol 3729, 262-276, Berlin:Springer Verlag.

GUARINO N, C. WELTY. 2004. An Overview of OntoClean. In S. Staab & R. Studer (eds.), *The Handbook on Ontologies*, 151-172. Berlin:Springer-Verlag.

JACQUES M-P. (2005). Structure matérielle et contenu sémantique du texte écrit. *Corela*, Volume 3, Numéro 2. En ligne : http://corela.edel.univ-poitiers.fr/document.php?id=769

KAMEL M, AUSSENAC-GILLES N. (2009a). Construction automatique d'ontologies à partir de spécifications de bases de données. *Journées Francophones d'Ingénierie des Connaissances (IC 2009), Hammamet (Tunisie),* F. Gandon (Ed.), Univ. Hassan II, 85-96. http://ic2009.inria.fr/docs/papers/KamelAssenacGilles_IC2009_10.pdf

KAMEL M, AUSSENAC-GILLES N. (2009b). Utiliser la structure du document dans le processus de construction d'ontologie. *Conférence internationale TIA (Terminologie et Intelligence Artificielle),* Toulouse, nov 2009. http://www.irit.fr/TIA09/thekey/articles/kamel-aussenac.pdf

LUC C. (2001). Une typologie des énumérations basée sur les structures rhétoriques et architecturales du texte. *Actes de TALN 2001*, *Université de Tours*, juillet 2001, 263-272.

LUC C., VIRBEL J., (2001), Le modèle d'architecture textuelle : fondements et expérimentation. *Verbum*, Vol. XXIII, N. 1, 103-123.

MAEDCHE A., (2002), *Ontology learning for the Semantic Web*, volume 665. Kluwer Academic Publisher.

ROLE F., ROUSSE G. (2006). Construction incrémentale d'une ontologie par analyse du texte et de la structure des documents. *Document numérique*, vol. 9, 77-91.

VIRBEL J. (1989). The contribution of Linguistic Knowledge to the Interpretation of Text Structures. *In* André J., Quint V., Furuta R. (eds,) *Structured documents*, 161-181, Cambridge University Press.