

NEED : NEW E-LEARNING ENVIRONMENT AND DEMONSTRATORS

Maria Teresa Pazienza

DISP, Université de Rome Tor Vergata
Via del Politecnico 1, 00133 Roma, Italie
pazienza@info.uniroma2.it

Armando Stellato

DISP, Université de Rome Tor Vergata
Via del Politecnico 1, 00133 Roma, Italie
stellato@info.uniroma2.it

Fabio Massimo Zanzotto

DISP, Université de Rome Tor Vergata
Via del Politecnico 1, 00133 Roma, Italie
zanzotto@info.uniroma2.it

Résumé : Les communautés européennes et méditerranéennes ont la nécessité de garder leurs langues et leurs cultures nationales afin de favoriser le processus d'intégration. Parler des langues étrangères, être en mesure d'interagir au sein de cultures différentes facilite l'accès à tout type de documentation écrite. NEED, système en développement à l'université de Rome Tor Vergata, aura pour objectif d'explorer des solutions innovatrices pour l'étude des langues en intégrant des outils de TALN tels que des analyseurs morpho-syntaxiques, des dictionnaires, glossaires, et ontologies.

Abstract : European countries as well as those in the Mediterranean area search for maintaining both language and national culture in order to continue to have an independent role in the international community. Nevertheless, it is important to speak more than one language to better interact in a foreigner country: this has a great relevance for people moving for working out of their original location. We have modelled a framework for introducing NLP technologies in support of developing an e-learning systems for natural language.

Mot-clés : e-learning, TALN, analyseurs morphosyntaxiques, dictionnaires, ontologies.

Keywords : e-learning, NLP, morphosyntactic analysis, dictionaries, ontologies.

INTRODUCTION

Les communautés européennes et méditerranéennes ont la nécessité de garder leurs langues et leurs cultures nationales afin de favoriser le processus d'intégration.

Parler des langues étrangères, mais surtout être en mesure d'interagir au sein de cultures différentes facilite l'accès à tout type de documentation écrite (administrative, légale, technologique, social, économique, sociale, politique etc.) mais de plus confère une meilleure capacité de compréhension des textes.

En effet, on peut considérer le multilinguisme, ainsi que la réelle capacité à communiquer à différents niveaux de compréhension et à partir de différentes sources de connaissance comme un atout majeur, surtout lorsque l'information est fournie dans une langue non maîtrisée.

Par conséquent, il serait donc opportun d'améliorer les services d'instruction et de formation qui incluent l'enseignement d'une langue étrangère pour les élèves d'âge scolaire, mais aussi, de manière plus générale, pour tous les citoyens impliqués dans un processus d'étude tout au long de leur vie.

Au jour d'aujourd'hui, ces services d'instruction désormais bien rodés, doivent accéder à l'étape successive qui est d'être intégrés dans plusieurs applications informatiques. Cependant, il reste encore à comprendre avec précision comment les outils de technologie avancée utilisés pour le traitement de texte contribuent effectivement à la connaissance des langues.

Nous aimerions considérer l'apprentissage électronique (l'e-learning) comme une approche de l'apprentissage « autogéré » et étayé par de nouveaux instruments. Ceci permettrait à un grand nombre de personnes se trouvant dans des lieux différents de suivre une formation efficace même dans des domaines cognitifs en fréquente mutation. On s'attend à ce qu'un projet bien conçu augmente la capacité d'apprentissage et diminue aussi bien le temps de réaction aux changements nécessaires dans la préparation d'un cours que les coûts de gestion pour chaque étudiant.

Il semble que la nouveauté plus importante dans le domaine de l'apprentissage électronique soit l'utilisation d'outils attractifs et technologiquement avancés. Le Traitement Automatique des Langues Naturelles (TALN) joue un rôle très important dans un processus d'apprentissage globalisé.

Actuellement il est difficile, autant pour les enseignants que pour les étudiants, d'utiliser les technologies de TALN existantes et amplement disponibles. De plus, l'apprentissage électronique souffre de l'absence d'un cadre logique d'évaluation comprenant des mesures objectives dans des contextes d'environnement technologique. L'apprentissage électronique est fondé sur l'hypothèse que les technologies de l'information et de la communication (TIC) sont utilisées afin d'améliorer une formation dans un cadre de travail totalement anonyme et en condition d'éloignement géographique des participants (enseignants et étudiants)

Il y a encore des aspects de l'apprentissage électronique qui restent inexplorés tel que, par exemple le problème majeur lié à l'éloignement géographique des participants qui implique des modèles asynchrones d'apprentissage puisqu'il s'en suit une séparation temporaire (les délais durant lesquels chaque étudiant apprend et les délais d'envoi du matériel d'apprentissage étant bien entendu différents)

Puisque l'apprentissage n'est pas conduit par un enseignant (autogestion du temps de la part de l'étudiant), il est important de fonder l'organisation du matériel de cours sur des concepts indépendants qui pourront être utilisés de manière autonome et exhaustive ainsi les réponses de l'étudiant se feront par niveau conceptuel.

L'organisation par niveau de concepts permet à l'utilisateur de jouir d'une grande flexibilité d'utilisation ainsi que de la possibilité de commander les sujets de cours en fonction de son propre intérêt et de ses aspirations.

Le second problème est celui lié au développement des systèmes d'apprentissage électronique dans le domaine des compétences linguistiques étant donné que le langage est le véhicule lui-même de l'information pour tout matériel didactique.

Les technologies linguistiques (TALN) jouent un rôle très important dans le cadre de la formation dite « globale ». il est demandé à chaque étudiant d'avoir un niveau élémentaire d'informatisation ainsi qu'une réelle capacité à interagir à différents niveaux de compréhension grâce à différentes sources de connaissance

(donc d'avoir la capacité de réagir à tout type de phénomène conceptuel). En outre, il existe un thème spécifique lié à l'apprentissage électronique des langues. En effet, il est important de mettre en évidence quelles sont les relations qui existent entre les descriptions conceptuelles et leurs différentes représentations linguistiques superficielles dans les textes. Cela peut sembler étroitement lié à un langage spécifique, cependant il est impératif de développer des structures de représentations conceptuelles suffisamment générales afin qu'elle puissent être employées pour la compréhension des contenus des textes à un niveau multilingue.

NEED, système en phase de développement à l'université de Rome Tor Vergata, aura pour objectif d'explorer des solutions innovatrices pour l'étude des langues en intégrant des outils de TALN tels que des analyseurs morpo – syntaxiques, des dictionnaires, glossaires, ontologies.

Par le biais d'une GUI simple d'utilisation, l'étudiant serait capable d'écrire une phrase et obtiendrait des explications détaillées sur sa structure, ou des solutions alternatives dans les contraintes syntaxiques lui suggérant différentes significations.

En cliquant sur différentes fenêtres, l'utilisateur pourrait obtenir des informations sur les règles syntaxiques sous-jacentes, sur les différents sens des mots, sur les formes et structures linguistiques alternatives, les expressions archaïques, formelles ou de jargons spécifiques à certains domaines, ainsi que des détails de terminologie et des informations du glossaire. La qualité et le niveau des informations fournies pourraient être adaptés à chaque étudiant selon son âge, ses intérêts ou la matière étudiée.

En comparant la phrase d'input écrite par l'étudiant aux différentes phrases possibles et correctes, il sera possible d'évaluer l'exactitude de la phrase écrite du point de vue morphologique, syntaxique et sémantique grâce à des mesures spécifiques.

L'objectif principal de ce travail est de démontrer comment il est possible de créer un environnement de travail dynamique et personnalisable en support à l'apprentissage d'une langue seconde de la part d'un étudiant adulte et dans lequel ce serait l'apprenant lui-même qui déciderait des délais et des méthodes de travail et même d'approfondir l'apprentissage de structure morphosyntaxiques ou encore de réaliser un approfondissement sémantique et conceptuel à propos d'expressions linguistiques particulières, dans les délais et de la manière qu'il jugera plus opportuns. Ceci sera possible grâce à la création d'un learning desk (bureau d'apprentissage) qui permette d'activer des ressources linguistiques diversifiées et /ou des processus selon les besoins d'approfondissement de l'élève.

L'étudiant considéré pour notre scénario est une personne adulte, devant s'expatrier et qui a donc l'intérêt et la nécessité d'apprendre la langue du pays de destination pour des raisons d'intégration ou plus simplement pour des raisons professionnelles. Un système d'apprentissage électronique fondé sur une méthodologie TALN permet d'assister ce genre de travailleurs en les aidant à s'intégrer avec rapidité dans leur nouvel environnement professionnel grâce à un instrument d'acquisition de connaissance linguistique et professionnel très pratique ceci dans des domaines spécifiques tout en facilitant leur travail de remise à niveau et leur enrichissement culturel quel que soit leur âge.

Afin qu'il puisse employer un outil d'apprentissage d'une langue seconde (du genre de celui que nous proposons), le niveau de compétence requis en ce qui concerne les instruments d'élaboration automatique est un niveau moyen, voire élémentaire, c'est à dire que l'on suppose qu'il sache guider l'instrument d'apprentissage électronique plutôt que de se laisser guider dans le processus d'apprentissage et d'approfondissement de la connaissance linguistique. En effet, la diffusion de l'alphabétisation informatique de masse nous pousse à concevoir des systèmes simples d'emploi, voire utilisables en amateur, qui permettent un apprentissage complet (morphosyntaxique et sémantique) de la nouvelle langue grâce à des sessions indépendantes d'étude sans parcours ou au-delà d'un parcours organisé et prédéfini. En outre dans le scénario suivant, notre système est configuré pour être une aide aux devoirs pour l'étudiant qui désire éventuellement approfondir les notions apprises pendant le cours. Celui-ci aura donc la possibilité de mettre en pratique les enseignements reçus de la part d'un professeur en s'exerçant à écrire des phrases de diverses typologies, tout en ayant la possibilité d'en vérifier simultanément l'exactitude morphosyntaxique à travers l'activité d'un agent qui lui fournira également toutes les solutions alternatives possibles et en mettant en évidence les liens sémantiques et les approfondissements conceptuels.

Un scénario d'utilisation diversifié est de plus en plus à l'ordre du jour en référence au cas de noyaux familiaux qui s'expatrient pour raison professionnelle. Les parents ont la nécessité d'approfondir non seulement

la connaissance de la langue mais aussi particulièrement le jargon professionnel employé dans leur environnement de travail, alors que les enfants doivent surmonter des difficultés d'intégration scolaire. Ces derniers pourront ainsi affronter à leur rythme et de manière personnalisée les difficultés cognitives rencontrées dans les différentes disciplines scolaires et dues à une connaissance limitée de la langue locale.

Nous envisageons donc un système unique de ressources qui puisse être utilisé dans différents cas de figure d'apprentissage.

1 FONCTIONS DU SYSTÈME

NEED possède un ensemble de fonctions qui peuvent être activées par l'utilisateur ou qui peuvent lui être proposées. En effet, dans la phase initiale et pour chaque module linguistique, il est prévu que l'élève soit guidé dans un processus d'apprentissage qui propose des exemples de constructions linguistiques utilisées dans des phrases isolées et ayant un sens ou à l'intérieur d'un contexte. Les particularités linguistiques sont ensuite commentées et analysées avec précision pour chaque cas différent. Ensuite l'étudiant est invité à produire une phrase qui présente une structure linguistique du même genre, enfin cette phrase est analysée pour en vérifier la conformité aux caractéristiques linguistiques à la langue naturelle, autant sur le plan morphologique et grammaticale que sur le plan syntaxique et structurel. En outre si le cas se présente, le système peut, de manière interactive, proposer à l'utilisateur des suggestions pour la correction du texte tout en expliquant à l'élève la fonction d'assistance à l'apprentissage.

Le contrôle, à différents niveaux, de l'exactitude des textes, l'éventuelle mise en évidence des erreurs, les suggestions pour la correction et la proposition d'expressions linguistiques alternatives (variantes linguistiques) constituent l'élément principal du système (voir détails ultérieurs dans le paragraphe concernant l'évaluation). L'exactitude du texte peut être vérifié au travers du respect de trois contraintes.... Ces typologies permettent d'identifier trois catégories de complexité linguistique progressive à l'intérieur desquelles le système met en action les liens les plus adéquats au niveau d'apprentissage de l'étudiant, et éventuellement, au style, jargon du domaine:

- **Contraintes de type orthographique et morphologique:** ces contraintes sont subordonnées à la fois aux règles de construction et de variabilité des mots et à l'accès à une terminologie dédiée à un domaine, qui puisse éventuellement tenir compte des expressions techniques, des emplois particuliers à souligner à l'attention de l'utilisateur
- **Contraintes de type syntagmatique:** elles restreignent l'emploi de structures de phrases de la langue italienne au travers des critères dont la position correspond à un syntagme ou à une phrase complète. Des liens qui reconnaissent le niveau d'apprentissage de l'étudiant sont prévus et s'y aligneront en suggérant des structures syntagmatiques de plus en plus évoluées.
- **Contraintes de type sémantique:** ce sont ceux qui sont relatifs aux restrictions d'emploi des constructions syntagmatiques, ou à l'usage de formes référentielles anaphoriques ou encore qui font appel à la possibilité d'introduire des spécificateurs à un niveau arbitraire d'imbrication

Alors que les deux premières typologies peuvent être employées en relation avec la grammaire du langage qui fait l'objet de l'apprentissage, la troisième, elle, est plus complexe car il n'existe pas de modèle unique et explicatif des phénomènes sémantiques mais aussi parce que les informations mises à disposition lors de l'analyse du texte sont incomplètes. De plus, la compréhension interprétée par celui qui écrit et véhiculée par le texte qu'il a produit est difficile et requiert des modèles de contrainte capables de gérer les incertitudes (interprétations multiples d'une seule phrase erronée) ainsi que l'absence d'informations.

Le système, grâce à l'utilisation d'une interface graphique qui s'ouvre dans une fenêtre et de sous-systèmes de correction, de suggestion, et d'accès à des ressources linguistiques et ontologiques, permettra un apprentissage progressif et personnalisé.

Le système permet en particulier les fonctions suivantes:

- Écriture de documents en langage naturel de la part de l'utilisateur.
- Analyse grammaticale des phrases écrites (parser); le parser utilisé est le système CHAOS qui consent la dérivation de graphes syntactiques (XDG) à partir de phrase en cours d'écriture.

- Aide aux corrections qui donne la possibilité (à l'intérieur du même environnement d'*editing*) de recevoir les erreurs (communication du type de structure non acceptable et de sa position) et de proposer les modifications nécessaires.
- Approfondissement sémantique des termes employés grâce à l'accès à des dictionnaires, des lexiques, ontologies et terminologies, etc pour l'apprentissage et la vérification des diverses interprétations possibles associées à un terme afin de compléter la connaissance de l'emploi des mots, des expressions idiomatiques et des termes appropriés...

L'ensemble de ces fonctions consent à l'utilisateur de prendre conscience de ses erreurs de manière interactive et donc d'améliorer la connaissance du langage grâce aux suggestions fournies par le système. Le système examine la structure syntaxique de la phrase produite par le parser, et vérifie si elle est en accord avec les contraintes du langage. L'analyse se base sur des lexiques, dictionnaires et ontologies (qui peuvent être ou non spécifique au domaine de compétence, selon la personnalisation choisie). Les premiers donnent des contraintes linguistiques, les derniers définissent les propriétés générales essentielles à la validation du processus d'écriture. La possibilité pour l'apprenant d'accéder « à la demande » aux ressources linguistiques (dictionnaires, traducteurs...) et cognitives (ontologies), offerte par OntoLing constitue un atout majeur et de qualité pour un apprentissage « à la carte ». En effet, pour chaque terme examiné il est possible d'obtenir des définitions, des descriptions détaillées de toutes les significations qui peuvent y être associées, des synonymes, antonymes, des références à d'autres termes, des traductions en d'autres langues etc. qui offrent ainsi un vaste panorama de ce que sont les caractéristiques et les propriétés d'un mot et les utilisations possibles du *parser*.

Suivent quelques aperçus des ressources linguistiques et ontologiques auxquelles le *parser* permet d'accéder selon les modalités prévues par ce système que nous avons développé.

2 LE PARSER

Il a été décidé d'utiliser un *parser* robuste, le système CHAOS, développé par le groupe de recherche ART de l'Université de Rome Tor Vergata accessible « en ligne » sur le site :

[http // ai-nlp.info.uniroma2.it/external/chaosproject/index.html](http://ai-nlp.info.uniroma2.it/external/chaosproject/index.html) et au sujet duquel nous fournissons ici quelques détails. (Pour de plus amples informations veuillez vous reporter à la bibliographie de référence disponible également sur le site de CHAOS).

Le parser CHAOS tente de regrouper aussi bien les avantages des méthodologies de parsing superficiel (shallow parsing) que ceux de parsing lexical en proposant une architecture modulaire qui ait comme caractéristique principale la robustesse, ou plus exactement la constance de prestation dans des domaines cognitifs différents. L'architecture qui en résulte, hautement modulaire, permet la création d'analyseurs syntaxiques reconfigurables qui peuvent s'avaloir d'un degré élevé de robustesse. Les modules qui constituent CHAOS sont :

1. *Tokenizer*: qui transforme le texte introduit de suite de caractères en séquence de mots.
2. *Analizzatore morfologico*: qui, à chaque mot associe son rôle syntaxique ainsi que ses propriétés morphologiques.
3. *Yellow Page Tool*: qui recherche dans le texte les lemma présents dans une liste connue de mots et caractéristique dans le domaine.
4. *POS Tagger*: qui effectue le Part-Of-Speech tagging.
5. *Named Entity Recognizer*: qui met en évidence dans le texte les noms propres et les structures identifiables comme Named Entity.
6. *Chunker*: qui reçoit en input des mots isolés et génère des séquences de mots adjacents et considérés comme des unités grammaticales, appelées aussi chunk (par exemples des syntagmes nominaux ou verbaux).
7. *Verb Shallow Analyzer*: qui analyse les structures verbales à l'intérieur des propositions de la phrase. Les liens grammaticaux les plus typiques sont : sujet (SUBJ), objet (OBJ), et modificateur verbal (PPMOD)

8. *Surface Shallow Parser*: qui met en évidence dans le texte toutes les dépendances syntaxiques.

Afin de mieux décrire encore le comportement du système CHAOS voici comment une phrase est analysée par les différents processeurs. Prenons l'exemple suivant : « IBM lavora da 20 anni nella borsa di New York » (*IBM travaille depuis vingt ans à la bourse de New York*). Décrivons le processus d'analyse de la chaîne de parsing.(analyseur grammatical) Chunker-VSA-SSA.

La phrase est tout d'abord découpée en mots analysés selon leur morphologie puis ensuite regroupés en *chunk* par le module adapté, le chunker. Le résultat de cette analyse est décrit dans la figure 1 dans laquelle on peut voir que quelques mots n'ont pas été regroupés, « IBM » et « lavora » alors que d'autres sont regroupés en chunk comme « nella borsa », « da 20 anni » et « di New York ». comme nous pouvons le remarquer, à chaque mot est assignée une interprétation morphosyntaxique préférentielle (l'étiquette de Part-Of-Speech, partie du discours) comme VFT pour « lavora » mais les interprétations morphosyntaxiques alternatives ont été conservées, par exemple « lavora » conserve les quatre interprétations représentées dans le schéma par *id[0]*, *id[1]*, *id[2]* et *id[3]*.

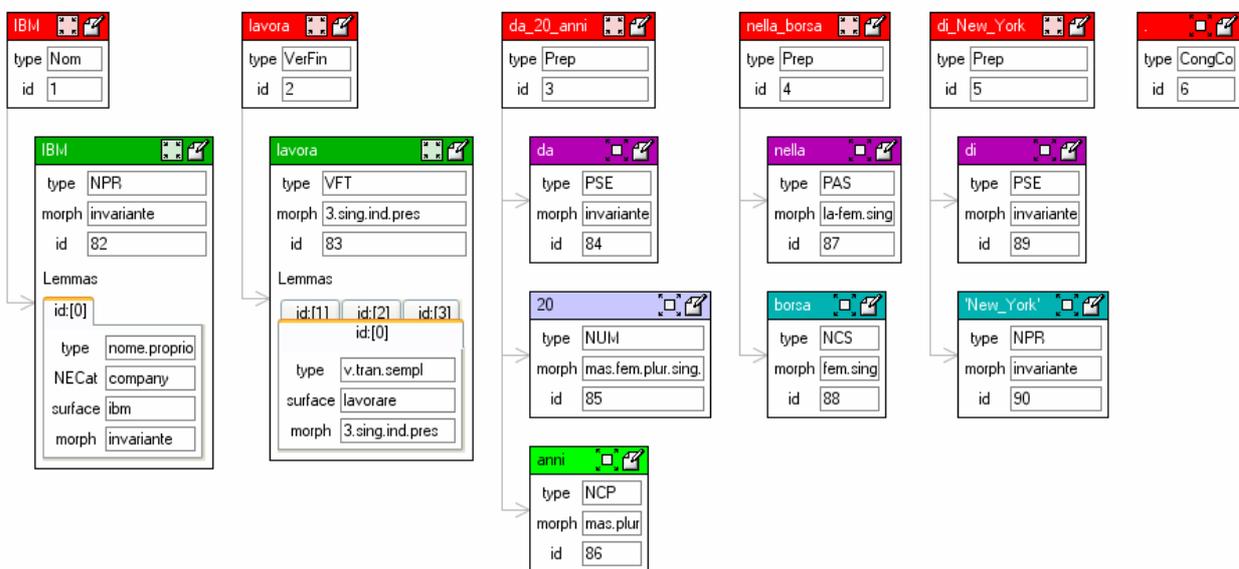


Fig. 1: phrase après utilisation du chunker (analyseur grammatical)

Le résultat de l'application de l'analyseur de liens verbaux (VSA) est présenté dans la figure 2, et comme vous pouvez le remarquer, les structures verbales composées sont reconnues comme des arguments verbaux. Dans le cas du verbe « *lavorare* », une structure de sous catégorisation prévoit qu'il y ait un sujet et un argument... par la préposition « *in* ». ceci est reconnu par le VSA et est proposé comme analyse de la phrase. Cependant, certains arguments verbaux comme « *da 20 anni* » sont négligés car non prévus dans la structure de sous catégorisation. De tels arguments sont parfois appelés *arguments ajoutés* puisqu'ils ne contribuent pas à la définition du sens du verbe.

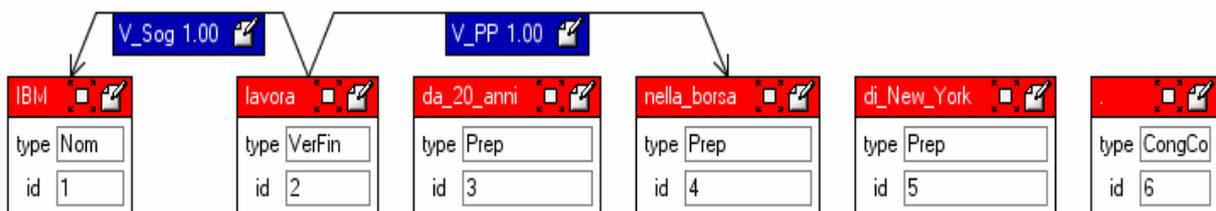


Fig. 2: phrase après utilisation de l'analyseur de liens verbaux

La phase finale de l'analyse faite par le SSA produit la représentation syntaxique de la phrase de départ représentée dans le schéma 3. À ce point, d'ultérieurs modificateurs verbaux sont ajoutés puis les modificateurs nominaux et enfin adjectivaux. Dans le cas spécifique de notre phrase d'exemple, on a ajouté le modificateur verbal « da 20 anni » et le modificateur « di New York », du nom « borsa ». Remarquez que la configuration de l'analyse précédente empêche la liaison de liens tels que celui qui relie « da 20 anni » à « di New York » puisqu'il s'agirait d'une contradiction avec le choix de relier le verbe « lavorare » à son modificateur « nella borsa » fait dans le module précédent.

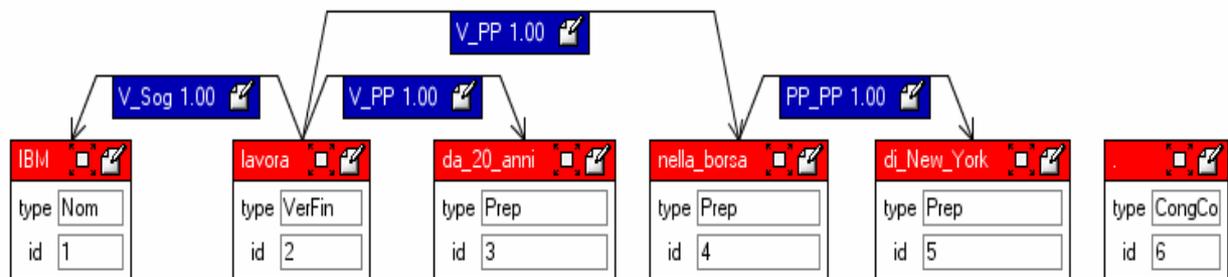


Fig. 3: phrase en fin de processus

3 RESSOURCES LINGUISTIQUES ET ONTOLOGIQUES

Par le terme « *ressources linguistiques* », on entend un ensemble de données linguistiques très variées et regroupées sous forme de base de données lexicales, de dictionnaires bilingues et terminologiques, qui se diversifient d'une part par l'information linguistique spécifique qu'ils offrent, d'autre part par leur format, la granularité de leur contenu et aussi par les motivations (théories linguistiques de référence, tâche à réaliser...). Dans le système proposé, celles-ci seront accessibles après avoir adhérer à la notion de Linguistic Watermark (schéma n°4)

Une ontologie des ressources dans laquelle les caractéristiques, les modalités d'organisation des informations lexicales, ainsi que les méthodes permettant d'y accéder sont décrites de manière formelle. Ce modèle est mis en place dans OntoLing, plug-in de Protégé, basé sur trois composantes:

- Une GUI (cf. Fig. 4), divisée en deux, correspondant au browsing des ressources linguistiques et ontologiques.
- Une composante pour la gestion de l'ontologie.
- Un *core component* qui réalise et fournit toutes les fonctions de OntoLing. Différentes modalités d'accès aux ressources linguistiques et ontologiques ont été mises en place, de manière si transparente qu'elles permettent à l'apprenant d'avoir une vision complète des informations hétérogènes représentées. Le passage de l'une à l'autre est aisé, ainsi l'apprenant peut obtenir des informations de terminologie, chercher une expression équivalente selon le sens, explorer des relations spécifiques de la ressource (synonymes, antonymes, polysémie...), chercher la définition d'un terme, en demander les synonymes, identifier et classer les différents sens et obtenir une traduction dans une autre langue.

L'objectif principal de OntoLing est d'enrichir les ontologies afin d'en augmenter la signification linguistique et la correspondance avec un ou plusieurs idiomes de référence (cf. Fig. 6). Il sert les exigences de NEED sous un autre aspect, celui de la multiplicité des cultures, des ethnies et donc des idiomes qui peuvent caractériser les utilisateurs potentiels d'un environnement multimédia d'apprentissage. Il en résulte que le fait de considérer l'ensemble des thèmes d'étude présents sur une plateforme d'apprentissage électronique

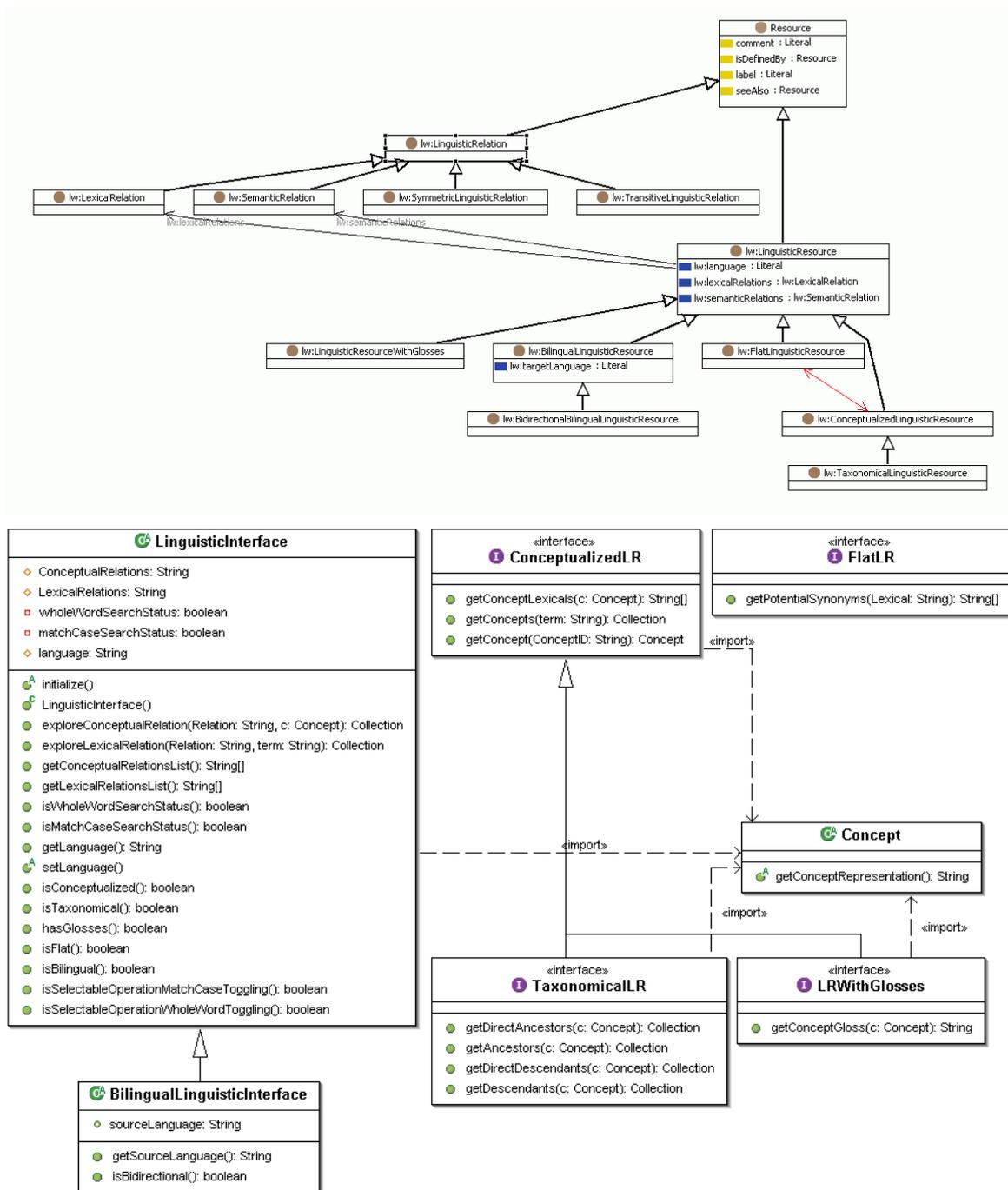


Fig. 4: le Linguistic Watermark : ontologie de référence et diagramme des classes de fontions

comme une structure ontologique rend possible, grâce à OntoLing, l'utilisation sans discrimination des contenus selon les préférences et/ou nécessité de chaque apprenant.

4 PRINCIPALES MESURES D'ÉVALUATION

Quel que soit le scénario d'apprentissage, (institutionnel ou autonome) il est important d'identifier des phénomènes linguistiques fondamentaux dans l'apprentissage qui puissent être associés à des fonctions d'évaluation de niveau et de qualité de ce même apprentissage. Puisque nous sommes dans un environnement TALN, nous pouvons tirer profit de la souplesse des instruments subordonnés au système d'apprentissage électronique, dans leur capacité à reconnaître de manière exacte des phénomènes linguisti-

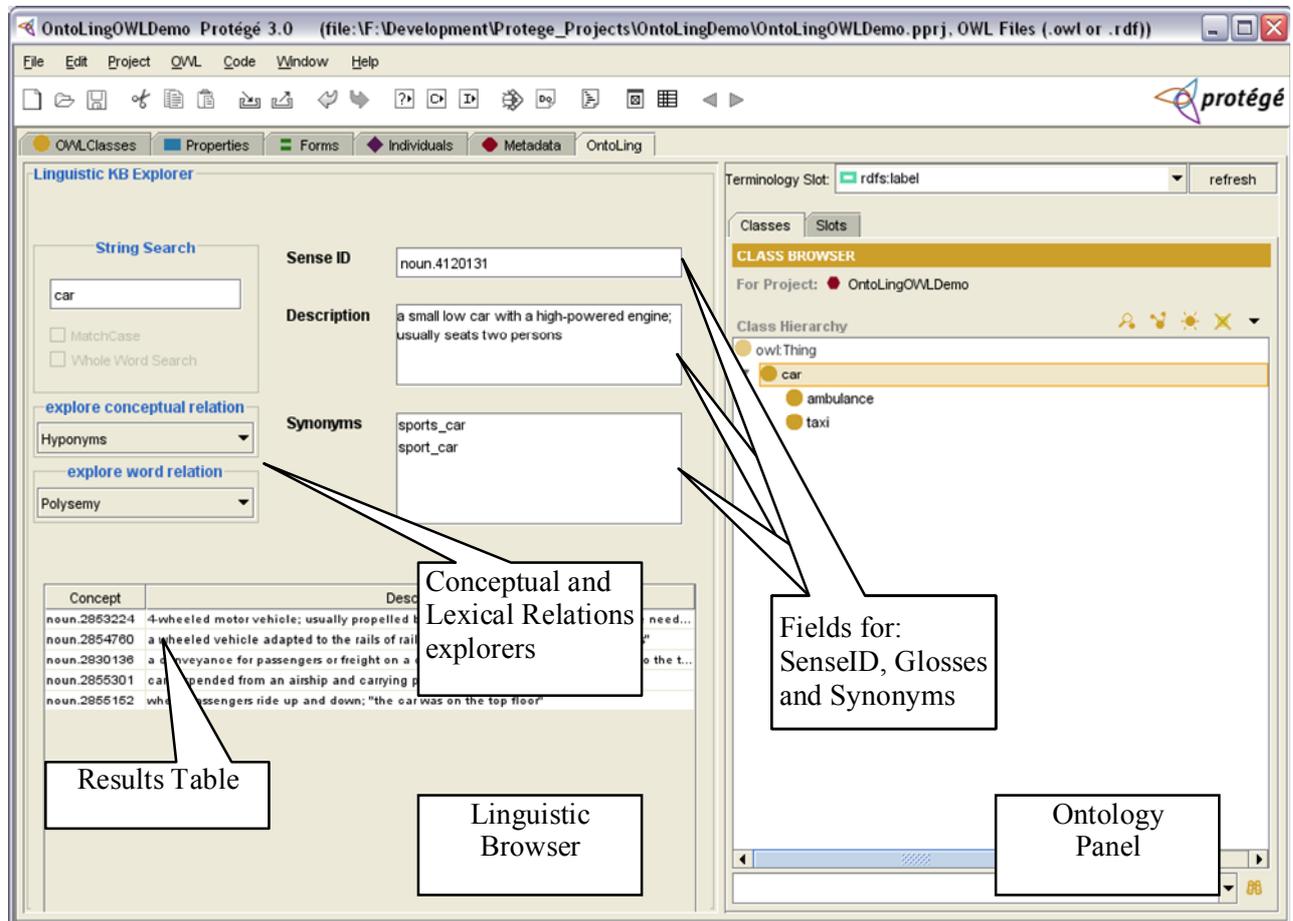


Fig. 5: OntoLing User Interface

ques à l'intérieur de phrases afin de vérifier la capacité de l'étudiant à reconnaître les différentes structures linguistiques et à les utiliser de manière opportune et ceci dans des contextes différents.

À l'intérieur d'un modèle d'apprentissage, nous pouvons définir des paramètres d'évaluation relatifs à:

1. Une syntaxe correcte
2. Une sémantique exacte
3. La conservation du sens

Pour de telles mesures, on peut envisager au moins

1. *Un contrôle de saccords sujets-verbres*: c'est-à-dire de vérifier dans le graphe syntaxique la présence des sujets et des verbes et qu'il n'y ait pas de subordinées sans sujet. (Comme des propositions relatives qui induisent une forte ambiguïté).
2. *Une vérification des modificateurs de syntagmes nominaux*: ceux-ci sont produits par l'analyseur syntaxique superficiel qui met en évidence les dépendances de type non verbal. Il est possible que quelques liens du langage imposent des restrictions sur ce type de structures. Par exemple, le contrôle examinera les structures de link entre chunk nominaux et chunk prépositionnel, et supprimera les interprétations qui ne seront pas cohérentes avec ces liens précis. En cas d'erreur, le système fournit l'ensemble de types d'erreurs associées au segment de phrase en analyse et suggère des possibilités de reconstruction correcte, tout en laissant le choix à l'utilisateur de la modification qu'il juge la plus opportune.
3. *Contrôle des propositions coordonnées et des propositions subordinées*: Le système évaluera la structure interne de l'ensemble des propositions. On pourra par exemple déconseiller l'utilisation de

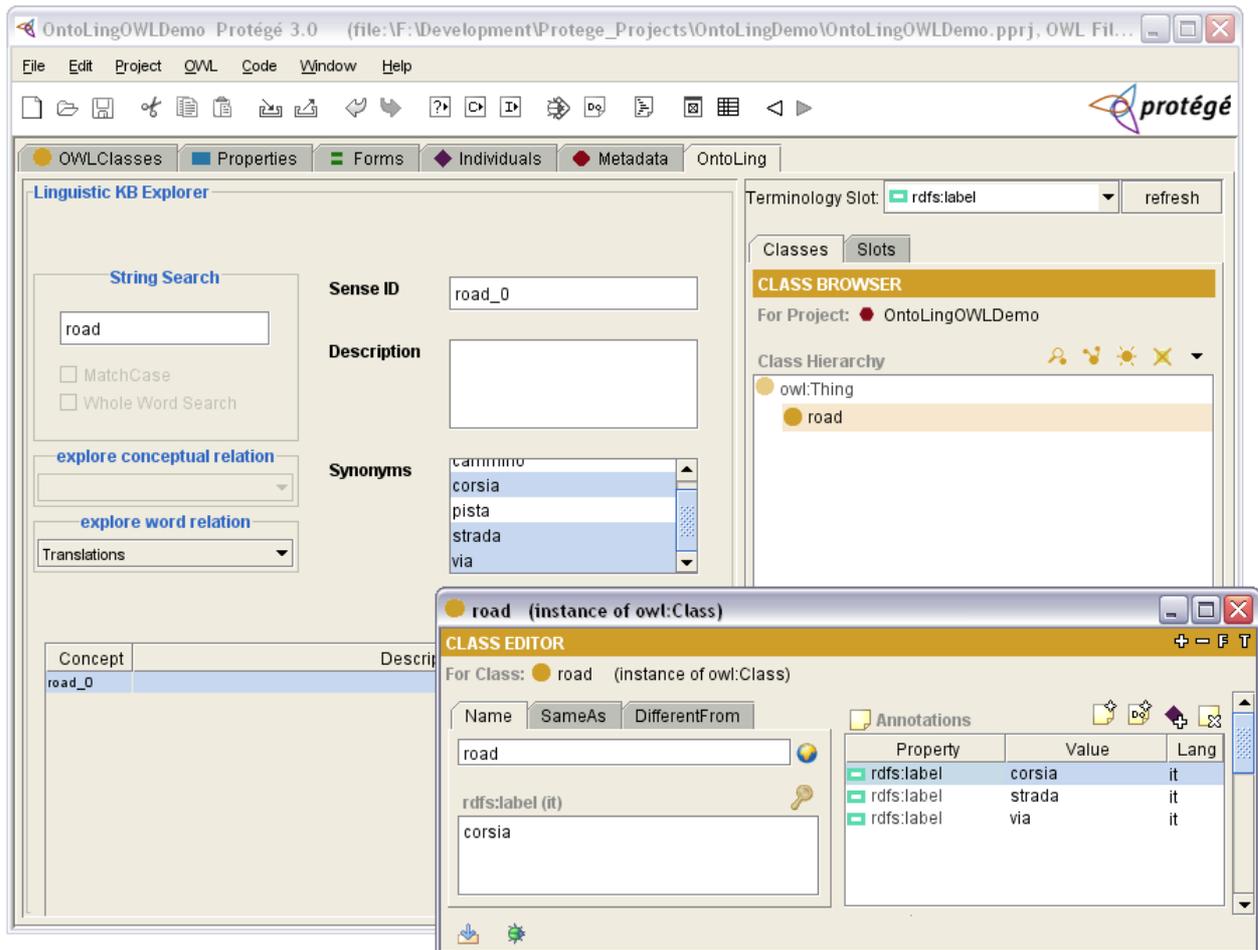


Fig. 6 : utilisation d'une ressource bilingue pour enrichir un concept de l'ontologie exprimé à partir d'un terme anglais en termes de langue italienne

phrases particulièrement complexes ou excessivement structurées et imbriquées qui auraient une compréhension difficile ou ambiguë. D'autres contraintes concerneront la manière dont les différentes subordinées peuvent être exprimées.

Il en découle que la vérification *des compléments verbaux* est fondamentale, c'est-à-dire:

- Contrôle syntaxique de la validité des modificateurs verbaux introduits par des prépositions.
- Contrôle sémantique de la validité des interprétations des modificateurs prépositionnels.

Dans le cas où il serait possible de fournir une sémantique pour le verbe en question, alors le type de modificateur prépositionnel pourrait être soumis à des contraintes de nature sémantique.

5 CONCLUSION

Par ce travail, nous vous avons présenté NEED, environnement pour l'assistance à l'apprentissage des langues naturelles qui permet un contrôle intelligent des activités de productions d'écrits tout en favorisant le processus d'apprentissage de type *learning by doing* (apprendre en faisant). Le système applique des techniques robustes d'élaboration de langage naturel et est fondé sur une représentation explicite des contraintes linguistiques (lexicales, morphologiques et syntaxiques) et du domaine d'application. Il offre une variété de possibilités d'accès aux ressources et aux informations linguistiques et ontologiques.

6 BIBLIOGRAPHIE

- Gibbs G., Learning by Doing: a guide to teaching and learning methods, Further Education Unit, London, 1988.
- Abney S., Partial Parsing via Finite-State Cascades, in Proceedings of the Eight European Summer School in Lgic, Language and Information, Prague, Czech Republic, 1996.
- Basili R., Pazienza M.T., Velardi P., A shallow Syntactic Analyzer to Extract Word Association from Corpora, in Literary and Linguistic Computing, 7, 114-124, 1992.
- Brill E., A simple rule-based part-of-speech tagger, in Proceedings of {ANLP}-92, 3rd Conference on Applied Natural Language Processing, Trento (Italy), 1992, pp. 152-155.
- Federici S., Montemagni S., Pirrelli V., Shallow Parsing and Text Chunking: a View on Underspecification in Syntax, in Proceedings of the Eight European Summer School in Lgic, Language and Information, Prague, Czech Republic, 1996.
- Fellbaum, C., WordNet - An electronic lexical database. MIT Press, 1998
- Basili R., Pazienza M.T., Zanzotto F.M., A Robust Parser for Information Extraction, in Proceedings of the European Conference on Artificial Intelligence (ECAI '98), Brighton (UK), 1998.
- Basili R., Pazienza M.T., Zanzotto F.M., Customizable Modular Lexicalized Parsing, in Proceedings of the 6th International Workshop on Parsing Technology, Trento (Italy), 2000.
- Pazienza M.T., Vindigni M., Identification and Classification of Italian Complex Proper Names, International Conference on Artificial and Computational Intelligence For Decision, Control and Automation In Engineering and Industrial Applications (ACIDCA'2000), Monastir (Tunisia), 2000.
- Basili R., Zanzotto F.M., Parsing Engineering and Empirical Robustness, Journal of Language Engineering, Cambridge University Press, 2002.
- Gennari, J., Musen, M., Ferguson, R., Grosso, W., Crubézy, M., Eriksson, H., Noy, N., and Tu., S. The evolution of Protégé-2000: An environment for knowledge-based systems development. International Journal of Human-Computer Studies, 58(1):89-123, 2003.
- Pazienza, M.T., Stellato, A. The Protégé Ontoling Plugin - Linguistic Enrichment of Ontologies in the Semantic Web 4th International Semantic Web Conference (ISWC-2005) Galway, Ireland, November, 2005
- Pazienza M.T., Stellato A., An Environment for Semi-automatic Annotation of Ontological Knowledge with Linguistic Content, 3rd European Semantic Web Conference (ESWC 2006), Budva, Montenegro, June 2006