



**HAL**  
open science

## Soutenir la coopération par l'indexation semi-automatique d'annotations

Gaëlle Lortal, Amalia Todirascu-Courtier, Myriam Lewkowicz

► **To cite this version:**

Gaëlle Lortal, Amalia Todirascu-Courtier, Myriam Lewkowicz. Soutenir la coopération par l'indexation semi-automatique d'annotations. IC - 17èmes Journées francophones d'Ingénierie des Connaissances, Jun 2006, Nantes, France. pp.61-70. hal-01025839

**HAL Id: hal-01025839**

**<https://inria.hal.science/hal-01025839v1>**

Submitted on 18 Jul 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Soutenir la coopération par l'indexation semi-automatique d'annotations

Gaëlle Lortal<sup>1</sup>, Amalia Todirascu-Courtier<sup>2</sup> et Myriam Lewkowicz<sup>1</sup>

<sup>1</sup> Laboratoire ICD-CNRS équipe Tech-CICO – U.T.T., 12, rue Marie Curie BP 2060 10010 Troyes Cedex, <http://www.utt.fr, {lortal,lewkowicz}@utt.fr>

<sup>2</sup> Linguistique, Langues et Parole - EA 1339 – Université Marc Bloch, 22 rue René Descartes 67084 Strasbourg Cedex, <http://www-umb.u-strasbg.fr, amalia.todirascu@umb.u-strasbg.fr>

## Résumé

Cet article décrit un outil permettant de soutenir des activités coopératives distribuées en conception mécanique à l'aide d'annotations. La conception assistée par ordinateur donne lieu à l'utilisation et à la création d'un grand nombre de documents et de commentaires liés aux documents qu'il faut classifier finement pour permettre leur réutilisation tout au long du projet. L'article se focalise sur les aspects concernant l'indexation des annotations. Notre outil se base sur des ressources termino-ontologiques (R.T.O.) construites semi-automatiquement à l'aide des techniques de Traitement Automatique de la Langue (TAL) sur les documents du projet. Cela permet une reconnaissance de candidats-termes (en utilisant les résultats de Syntex) et leur structuration selon des patrons syntaxiques. Les R.T.O., représentées en format « Topic Maps » aident l'utilisateur à classifier ses annotations et à naviguer dans l'index (en langue naturelle) adapté au partage de ces annotations entre les membres du projet.

**Mots clés :** annotation, indexation, construction de Ressources Termino-Ontologiques.

## 1 Motivations

Dans le contexte actuel du travail en conception mécanique, de nombreux projets sont effectués par des groupes distribués sur plusieurs sites. Dans un tel cadre, les documents numériques sont au centre du processus de travail et il est nécessaire de pouvoir les partager. Plusieurs études ont montré que les phases de travail asynchrones peuvent être soutenues par l'usage d'annotations sur les documents ([17], [30], [38]). Ces annotations sont des commentaires déposés sur un document par les membres du projet, à des fins de communication. Elles permettent de négocier le sens d'un terme ou d'un document, d'élaborer de nouveaux concepts en laissant des commentaires sur un document. Elles ont un rôle important dans le suivi des interactions entre les membres : elles sont les traces d'une

décision, d'une négociation, ou même d'une solution apportée à un problème ; elles jouent un rôle central dans le discours produit par les membres du groupe. Certaines informations contenues dans les annotations pourront être intégrées dans de nouveaux documents. Ces documents et fragments de documents peuvent être qualifiés de DoPAs (DOcuments Pour l'Action) [39]. Les différents auteurs de ces DoPAs ne pourront participer efficacement à leur élaboration que si les fragments de documents sont conservés et organisés dès la première lecture et pour toute version. Le partage ainsi permis participera à la formation d'un référentiel commun (« common ground », [15]) et facilitera la conscience des activités réalisées par les autres participants (« activity awareness », [13]). Pour soutenir ces élaborations de DoPAs, il faut donc rendre possible la gestion des annotations par le groupe projet : visualisation des documents de travail, annotation de ces documents, gestion des modifications, discussions entre les membres du groupe par le biais d'annotations, réutilisation des annotations, création de nouveaux documents sur la base des réflexions au travers des annotations, etc.

L'utilisation et le partage d'annotations ont fait l'objet de différents travaux dont des outils sont issus. On peut citer par exemple Magpie [19] permettant d'annoter des documents par des mots-clés, ou Yawas [18] permettant de laisser des commentaires sur des pages Web. D'autres auteurs proposent des forums de discussion autour des documents ; D3E [35] par exemple traite l'annotation comme un élément du discours, mais ne propose pas de classification des annotations, donc l'utilisateur ne peut pas identifier le processus de prise de décision à travers les annotations. De manière générale, les fonctionnalités de gestion des annotations sont limitées dans la plupart des outils existants. Pour répondre à ce besoin, nous présentons dans cet article les fonctionnalités liées à la classification d'annotations que nous proposons d'ajouter à l'outil d'annotation AnT&CoW. La version actuelle de l'outil AnT&CoW (que nous avons développée pour soutenir les activités documentaires par le biais d'annotations) propose

la visualisation et le partage des annotations, ainsi que la possibilité de répondre à une annotation [28].

L'ajout de fonctionnalités d'indexation d'annotations à AnT&CoW à l'aide d'une structuration des concepts construite par le groupe projet fait référence aux travaux menés en TCAO (Travail Coopératif Assisté par Ordinateur) par C. Simone et M. Sarini qui considèrent que les schèmes de classification (Classification Schemes) définis par une construction collective permettent de structurer les connaissances d'un groupe selon une organisation commune [34]. Les utilisateurs peuvent indexer et retrouver leurs annotations selon les aspects liés au domaine de travail mais aussi selon leur contenu argumentatif (pour reconstituer le processus de décision ou de négociation).

Dans le cas d'un fonds documentaire important (comme c'est le cas des documents numériques produits lors d'un travail coopératif), la construction de ressources pour la classification en partant des textes est longue et fastidieuse, tout comme l'activité d'indexation. Par conséquent, il est possible d'utiliser des techniques de Traitement Automatique du Langage Naturel (T.A.L.N.) permettant de construire semi-automatiquement des Ressources Termino-Ontologiques (R.T.O.) à partir de documents textuels représentatifs et disponibles dès le début du projet. Ainsi, pour aider à construire une classification basée sur des R.T.O. et pour soutenir l'indexation de nouveaux documents, nous avons adapté des techniques et outils d'extraction de termes, d'extraction de relations et de structuration des concepts extraits à partir des documents textuels utilisés par le groupe. De même, nous proposons de soutenir l'indexation par des outils de T.A.L.N. qui scrutent le nouveau document ou fragment de document pour proposer à l'utilisateur un lien vers la classification déjà existante. Cela permet aussi de mettre à jour ces ressources au fur et à mesure que la base de documents du projet est enrichie.

Cet article est organisé de la manière suivante : nous exposons en premier lieu notre contexte d'étude, un projet de conception mécanique aéronautique ainsi que le corpus fourni par l'observation de ce domaine. Nous présentons ensuite l'architecture de l'outil AnTCoW et ses fonctionnalités actuelles. À partir de scénarii inférés de l'étude du corpus, nous présentons les fonctionnalités en cours d'implémentation qui concernent l'organisation d'annotations. Nous nous concentrons sur les outils du T.A.L.N. ré-utilisables pour une tâche de construction de R.T.O. du domaine à partir du texte. Nous détaillons alors nos choix de formalisme et de structuration de concepts conservant une place centrale pour le document et exposons les prémices de notre R.T.O.. Nous concluons enfin par un exemple d'indexation réalisée par notre outil, et des perspectives d'enrichissement de notre outil développé itérativement.

## 2 Contexte d'étude

### 2.1 Le projet Air Campus

Dans le cadre d'une association aéronautique universitaire, un projet de conception mécanique aéronautique a été mis en place entre le bureau de cette association et une équipe de conception mécanique, des chercheurs du domaine ainsi que des techniciens. L'équipe d'ingénierie mécanique travaille de manière asynchrone (pendant leur temps libre) et distribuée (ils sont localisés sur trois sites). Leur objectif est de concevoir un réducteur pour permettre de monter un moteur automobile diesel sur un avion Delvion essence. Le projet est innovant bien qu'ayant déjà été testé par au moins une autre équipe de mécaniciens aéronautique. Les participants au projet se connaissent pour certains, et se rencontrent parfois d'une façon relativement informelle. Les membres du projet fondent leur travail sur des plans, des maquettes 2D ou 3D, des documents qu'ils s'échangent en format numérique par pièce jointe de mél ou en format papier par courrier. Ces documents sont la base de l'évolution du projet puisqu'ils expliquent ou représentent le travail à effectuer en production. Ils sont modifiés au fur et à mesure des échanges ou rencontres et il est important que les intéressés soient tenus informés des mises à jour et modifications des documents. Pour cela, l'annotation est un élément clé. Les membres de l'équipe de conception annotent les documents pour les modifier, pour informer les autres membres de l'importance d'une modification ou expliquer une telle modification, accompagnent leurs documents numériques d'une note dans un corps de mél, expliquent par un mél le contenu d'un document joint,...En clair, le passage au travail médiatisé, et donc au document numérique, supplée l'annotation par différentes techniques agissant sur la communication et la révision de document. Dans le cadre d'une médiatisation de leurs activités autour de documents, il est fondamental que les participants au projet puissent partager et gérer collectivement leurs productions documentaires (documents et messages annotatifs). Il faut qu'ils aient la possibilité de communiquer autour de documents et de conserver les communications et les modifications de documents et de produits. Pour tracer la logique de conception du produit à travers les documents, la problématique de classification des documents est centrale. Celle-ci s'appuie sur une indexation fine des documents et des messages par des mots-clés ou des expressions qui « parlent » au groupe projet.

Notre outil doit donc soutenir ces deux objectifs inscrits dans l'annotation : celui de production (de document, de message pour une communication asynchrone), et celui d'indexation de ces productions pour les réutiliser ou comprendre la logique de conception du moteur au travers des échanges du groupe. Nous nous focaliserons ici sur la problématique d'indexation des annotations basée sur des

R.T.O. représentatives du domaine du projet. L'indexation des annotations et la constitution de R.T.O. sont des tâches liées à l'activité de conception du groupe-projet dont nous présentons maintenant les échanges.

### 2.2 Le corpus de messages « annotants »

Afin d'observer la médiatisation possible de l'activité d'annotation (production de fragments de documents et d'indexation), nous avons regroupé des documents échangés au cours de ce projet. Pour pouvoir tracer les échanges au sein du groupe, nous avons effectué un relevé de documents représentant la communication au sein du groupe. Le seul outil disponible au sein de ce groupe pour la communication asynchrone médiatisée est le mél.

Objet : réducteur air campus : nouvelle pièces ...

Bonsoir Léon,

Ci-joint les CAO des nouvelles pièces :

\* Une version définitive de "platine moteur" sur laquelle des perçages ont été ajoutés, les dimensions des orielles inférieures ont diminuées, il y a une poche sur la face d'appui pour passage d'un joint cf doc "montage cage réducteur".

\* Deux supports de réducteur en acier à tailler dans la masse, (25CD4 si possible). Ces deux pièces sont symétrique. Donne moi rapidement ton avis sur l'usinabilité de ces deux pièces. Le plan incliné de la nervure n'est pas fonctionnel, il faut simplement soigné l'état de surface. Les rayons sont importants pour éviter les concentrations de contraintes. Je joins également la mise en plan en cas de souci. Merci.

Pierre.

**DOC.1** - Texte lié à une pièce jointe

>>>Salut Alphonse,  
>>>As-tu reçu mon mail, pourras-tu usiner la clé de boitier et si oui dans quel délai ?

>>>Merci de me répondre.

>>>Pierre.

>>

>Salut Pierre,

>Je vais usiner la clé de boitier lundi.

>Bonne journée

>A +

>Alphonse

>

Merci en de problèmes sur la FAO contacte moi.  
pierre.

**DOC.2** - Texte lié à un autre texte (réponses à)

Nous avons donc choisi de conserver 27 méls (2200 « mots » hors entêtes et signatures), tous expéditeurs et

destinataires confondus, dans lesquels nous constatons que les participants « annotent » par leur message une pièce jointe au mél, ou dans lesquels le participant répond en citant un mél précédent ancrant son message dans une conversation [3]. Ainsi les corps de ces méls représentent :

1. un texte lié à une pièce jointe : 18 méls (Doc.1) ;
2. un texte lié à un autre texte (mél en « réponse à », en « faire suivre ») : 9 méls liés, 17 méls une fois dépilés (Doc.2) ;

Suite à l'observation de ces échanges par méls, on constate que les participants utilisent un langage propre à leur domaine et des termes propres à leur projet. Il est donc difficile de réutiliser des ressources existantes en mécanique (thésaurus de mécanique, norme PLIB [31]). En effet, très génériques, elles regroupent de nombreux domaines de l'ingénierie mécanique en plus du domaine de la conception aéronautique. De plus, puisque l'on a affaire à un projet innovant, la classification nécessaire n'est prévisible qu'en partie. Ainsi, pour être adaptées au projet, les ressources termino-ontologiques doivent être construites sur la base du corpus étudié et sur la base des corpora de référence.

Comme notre objectif est d'améliorer la gestion des annotations par l'outil AnTCoW, nous présentons l'architecture de cet outil et nous nous focalisons sur les scenarii possibles d'indexation des annotations.

## 3 L'outil AnT&CoW

L'outil d'annotation collaboratif développé sur une plateforme Zope est destiné aux échanges autour de documents entre différents utilisateurs d'un groupe projet par le biais d'annotations. AnT&CoW (Annotation Tool for Collaborative Work) est basé sur une architecture client-serveur qui permet le partage des documents liés au projet. L'outil permet le partage des annotations, la visualisation, et la possibilité de discuter entre les membres du groupe (répondre à une annotation). Le serveur conserve les documents de l'équipe projet et la partie du document sélectionnée pour ancrer l'annotation d'une part, et d'une autre part, les annotations d'un document. Cette architecture permet à tous les utilisateurs de voir le document avec ou sans apparat critique.

Les annotations sont accessibles via un client, une extension d'Annozilla, plug-in d'annotation du navigateur Mozilla [37]. Ce client est une fenêtre incluse dans le navigateur, qui affiche les annotations déjà postées sur une page Web ouverte, ou sur tout document de la famille ML (Markup Language). Il permet de rédiger des annotations suite à la sélection d'une partie d'une page web (par surlignage), de poster des annotations et de répondre à une annotation postée sur la page. Les fonctionnalités d'organisation des annotations proposées par Annozilla se limitent à une indexation par des métadonnées telles que la date, l'auteur, ou le document sur lequel l'annotation porte.

Nous avons donc proposé de rendre cet outil plus performant en développant un module d'indexation basé sur des techniques de T.A.L.N. et sur l'utilisation de R.T.O. pour indexer ces annotations.

Les annotations peuvent être indexées selon plusieurs points de vue. Suivant [39], nous souhaitons une indexation par trois points de vue, (1) organisationnel, (2) argumentatif, (3) de domaine. Les points de vue organisationnel (rôle de l'acteur dans le projet) et argumentatif (trace de l'orientation de l'argumentation entre les participants humains) sont construites manuellement. La première est fondée sur une analyse sociale des membres du projet et la seconde sur une analyse cognitive et pragmatique des interactions dans le corpus. Elles ne seront pas détaillées dans cet article.

La création de la R.T.O. du domaine, elle, est réalisée hors-ligne, à l'aide d'un corpus textuel de référence. Ainsi, l'administrateur applique les outils d'extraction de termes et de relations pour obtenir un ensemble de candidats termes et de relations. Le module d'indexation est localisé sur le serveur ZTAL. L'administrateur propose une structuration de termes qui servira comme index initial pour les annotations des utilisateurs. La R.T.O. est placée sur le serveur ZOnToM et est accessible via le client Web.

Une fois cette première classification créée, les utilisateurs peuvent déjà la parcourir et choisir d'indexer les annotations qui sont créées. Un module d'indexation semi-automatique propose des candidats termes comme index possibles. Les membres du groupe pourront choisir les index qui leur semblent les plus pertinents mais aussi ajouter des index à la R.T.O.

La R.T.O. du domaine, les modules de T.A.L.N. et de mise à jour de l'ontologie sont décrits en section 4. Pour le moment, seule une R.T.O. du domaine est disponible.

## 4 Construction d'une R.T.O.

Nombre de projets se sont tournés vers la construction de ressources termino-ontologiques adaptées à leurs projets pour pallier la généralité ou les défaillances de ressources déjà constituées. Les ressources les plus utilisées pour une indexation d'un ensemble de documents spécifiques à un domaine sont les ontologies ou les thésaurus. Les thésaurus (EuroWordNet [37]) et ontologies génériques (DOLCE [22]) doivent être adaptés à une application spécifique à un domaine puisqu'elles contiennent peu de définitions de concepts spécifiques à un domaine, tâche très lourde et coûteuse. Les ontologies du domaine sont souvent onéreuses et ne sont pas toujours disponibles pour tous les domaines, malgré l'existence d'ontologies portables représentées à l'aide des standards du W3C (OWL, RDF). Ainsi, il n'est pas toujours possible de réutiliser une ontologie pour classifier et indexer les documents d'un projet spécifique. Les thésaurus, qui regroupent des termes structurés hiérarchiquement représentent une alternative

pour indexer des documents (EuroVoc [42], Thésaurus Eau [41]), mais ils ne sont pas disponibles pour tous les domaines possibles (notamment pour le domaine de la conception mécanique aéronautique). Les descripteurs proposés par un thésaurus sont figés et ne conviennent pas toujours pour indexer les documents.

En raison du peu de disponibilité d'ontologies et de thésaurus du domaine et de l'utilisation difficile d'ontologies génériques, nombre de projets utilisent les techniques de T.A.L.N. pour extraire semi-automatiquement des termes (des instances de concepts) [27], pour créer des agrégations de termes (concepts) [14] ou encore pour extraire des relations entre termes [11]. L'expert doit alors nommer les agrégats (clusters) en tant que concepts et doit éventuellement définir les relations entre concepts. Les différentes techniques de constitution de R.T.O. sont liées entre autres aux différentes sources terminologiques disponibles [30], [23] : bases de données, dictionnaires, corpus annotés ou bruts. Vu que nous n'avons pas trouvé d'ontologie ou de thésaurus spécifique à notre domaine, nous avons adopté une approche qui consiste à construire et à mettre à jour semi-automatiquement (à partir des corpus textuels) la R.T.O. utilisée pour l'indexation des annotations et des documents.

### 4.1 Méthodologie

Notre R.T.O. est construite dans un objectif d'intégration à un système d'annotation collaboratif. La R.T.O. est la classification sur laquelle est fondée l'indexation de l'annotation. Elle a pour rôle de permettre de partager et de regrouper les connaissances communes du groupe. L'utilisateur est au centre du projet puisqu'il s'agit de soutenir son activité innovante de conception mécanique. L'outil doit donc lui permettre de créer une classification spécifique (dont le peuplement continue au fil du projet), et d'indexer ses propres annotations et documents. La construction de la R.T.O. se fait au fur et à mesure du projet, par l'intervention directe des membres du projet (qui ne sont pas des ontologistes ou des linguistes). Les termes et les concepts utilisés doivent être clairs et précis pour retrouver facilement les annotations ou les documents et pour être compris par tous les membres du groupe. L'utilisateur doit avoir la possibilité d'introduire ses propres termes, sans être contraint à utiliser un ensemble de descripteurs figés, proposés par une ontologie ou par un thésaurus. La consistance de l'ontologie est validée par son « utilisabilité » dans le projet par les utilisateurs qui l'ont créée. Nous avons besoin d'une ressource termino-ontologique qui présente une structuration des termes comme une ontologie classique, mais qui ressemble plutôt à un thésaurus, par le partage des définitions de termes qui sont utilisés.

Nous avons donc adopté une méthode de construction de R.T.O. semi-automatique, basée sur des outils légers,

disponibles et encapsulables dans notre outil de travail coopératif. Ces outils doivent être robustes, à même de traiter des textes dans lesquels des approximations orthographiques et syntaxiques sont relevées (méls, textes issus de formes de discours dites « nouvelles », comme les messages des blogs ou des forums).

On se focalisera ici sur la construction de cette première R.T.O. du domaine et sur son utilisation pour l'indexation. Nous avons adopté une méthodologie classique de construction de R.T.O. [1] adaptée à notre application :

1. constitution de corpus de référence du domaine de la conception en aéronautique ;
2. extraction de termes à partir de corpus de référence, après avoir identifié les types de constituants syntaxiques qui vont être des candidats termes possibles ;
3. extraction des relations paradigmatiques (verticales) et syntagmatiques (horizontales) entre les termes
  - a) identification de règles heuristiques ;
  - b) identification automatique des relations selon ces règles ;
  - c) structuration des termes ;
4. choix d'un formalisme de représentation de la R.T.O. adapté à la navigation.

Nous allons présenter en détail le corpus du domaine, les choix des outils d'extractions de termes et de relations, les modifications que nous avons effectuées sur les outils d'extraction pour les adapter à notre application, ainsi que le formalisme adopté pour représenter la R.T.O..

### 4.2 Le corpus de référence

Afin de créer un corpus représentatif du domaine étudié, nous nous sommes servis des documents disponibles au début du projet. Un site Web sur l'adaptation des moteurs, des fascicules de pièces spécifiques et quelques documents de réflexion de base du projet, soient environ 19600 mots. Ensuite, nous avons utilisé l'outil BootCat pour rechercher des pages Web pertinentes pour le sujet [4] ce qui nous a permis d'obtenir un corpus d'environ 89000 mots grâce à d'autres sites Web, identifiés à l'aide de 15 mots clés du domaine de l'aéronautique. Suivant [1], après avoir constitué un ensemble représentatif du domaine par des documents fournis par les experts du projet (i.e. ses membres), il s'agit d'étudier de quelle façon traiter ces documents afin de permettre la structuration d'une carte de termes en langue naturelle. Ce réseau de termes permet la classification de documents pour leur partage et réutilisation durant le projet par les utilisateurs eux-mêmes.

On peut tout d'abord observer que le corpus est constitué de différents genres de texte (site web, documents commerciaux, discussions de la planification de la conception) qui mettent au jour différents types de termes liés à des sous-domaines de la mécanique : orienté fabrication, orienté conception, orienté aéronautique,... Ces genres de textes peuvent être compris comme

différents points de vue sur un projet. Dans la structuration de notre ressource, il faut donc prendre en compte une focalisation de rang supérieur qui n'est pas forcément représentée par les termes des documents. Dans ([2]), B. Bachimont souligne l'importance de la prise en compte d'un niveau supérieur dans la construction d'ontologie, comparable à un point de vue de domaine ou de la tâche.

Pour déterminer des points de vue à représenter dans l'ontologie, nous avons besoin d'identifier des termes, des relations et des concepts spécifiques au domaine, issus des corpus de référence.

### 4.3 L'extraction de termes

L'extraction de termes représente une étape importante dans la construction d'une ressource termino-ontologique et de nombreux outils performants [6] sont disponibles pour plusieurs langues. Cette étape intervient également dans le module d'indexation, qui identifie des candidats termes susceptibles de devenir des index pour les annotations ou les documents. Les textes des annotations et des documents contiennent de nombreux termes du domaine (la plupart du temps retrouvés parmi les groupes nominaux). Ces textes ont des origines variées (méls, documents de référence) et le langage varie du style oral au langage technique. En conséquence, les extracteurs de termes doivent être robustes pour pouvoir gérer les erreurs, mais doivent proposer des résultats pertinents, sans encombrer l'utilisateur avec trop de propositions d'index.

Nous avons étudié plusieurs extracteurs de termes dans un objectif d'intégration à notre plate-forme de travail coopératif, pour la construction de la RTO et sa mise à jour. Les termes qui seront extraits serviront à l'utilisateur pour constituer et structurer l'ontologie du domaine. Les outils tels que Termino (David et Plante, 1990), Lexter [7], FASTR [28], qui se basent sur des principes linguistiques, MANTEX [21] qui se base sur des traitements statistiques ou encore ACABIT [16], qui se base sur des traitements mixtes, identifient les candidats termes dans les corpus de référence. Les extracteurs de textes purement statistiques ont l'avantage d'utiliser très peu de ressources linguistiques (ex. LIKES [33]) pour proposer des candidats termes parmi les séquences des mots les plus fréquentes, et donc peuvent être facilement adaptables pour des applications nouvelles (en cas de changement de domaine). Cependant, les résultats proposés par ces méthodes statistiques contiennent beaucoup de bruit (candidats termes incorrects) et l'utilisateur ne pourra pas les filtrer à la main. Les extracteurs de termes qui utilisent des ressources linguistiques ( patrons morphosyntaxiques pour identifier des candidats termes parmi les groupes nominaux et parmi les groupes prépositionnels) proposent des résultats plus pertinents que les outils statistiques. Parmi les extracteurs de termes utilisant des ressources linguistiques, l'analyseur syntaxique Syntex [10], (utilisé en tant qu'extracteur de termes) identifie les dépendances

syntaxiques qui correspondent aux besoins de notre application. Ces relations de dépendances définissent le contexte des candidats termes dont nous avons besoin à la fois pour identifier les candidats termes mais aussi pour identifier les relations entre les termes. Vu les résultats fournis, Syntex est le plus adapté pour notre application puisqu'il nous permet d'utiliser des éléments syntaxiques dans ses contextes, mais également parce qu'il gère les erreurs de syntaxe (il propose une analyse partielle de la phrase). Une fois le corpus traité par Syntex, il est possible d'utiliser ses sorties et de les coupler à un algorithme extrayant des candidats termes et des relations pertinents au sein du corpus.

Pour extraire la liste des candidats termes qui vont faire partie de l'ontologie, nous avons développé un module qui cherche des constituants syntaxiques susceptibles d'être termes de l'ontologie. Nous avons sélectionné un ensemble de candidats termes du domaine à l'aide des patrons morpho-syntaxiques, parmi les groupes nominaux complexes (groupes nominaux modifiés par plusieurs groupes prépositionnels) et parmi les groupes verbaux simples (verbe suivi d'un groupe nominal simple ou complexe). Parmi les constituants syntaxiques identifiés dans les textes, plusieurs structures sont possibles, mais nous proposons toujours ceux qui ont une couverture maximale (exemple : un groupe nominal suivi de plusieurs modificateurs groupes prépositionnels, des noms modifiés par plusieurs groupes prépositionnels reliés par une conjonction), dans un souci de proposer à l'utilisateur une liste restreinte mais pertinente des index. Au delà des structures syntaxiques complexes identifiés, les candidats termes peuvent être groupés en plusieurs catégories : défauts de fabrication, paramètres de fonctionnement des moteurs ou des pièces/composantes, emplacement des pièces, éléments de solutions, etc. (Doc.3). Les diverses familles de termes sont définies manuellement comme des points de vue du domaine et les utilisateurs vont classer eux-mêmes les candidats termes sous le point de vue concerné : composants/pièces, défauts, solutions, paramètres.

Point de vue	Candidats – termes
Paramètre	« un aperçu des dimensions et charges de base de la série sélectionnée »
	« cale d'épaisseur au sommet du logement »
Défaut	« Caractérisation des défauts de parallélisme »
	« un défaut d'alignement »

**Doc.3** – Exemple de points de vue et candidats extraits

Ainsi, les candidats termes identifiés dans un des documents du corpus de référence sont rattachés à divers points de vue (Doc.3).

Une fois des termes extraits et organisés sous les points de vue identifiés, il s'agit d'extraire des relations syntagmatiques ou paradigmatisques pour relier les différents termes.

#### 4.4 L'extraction de relations

L'identification des relations entre des termes permet de relier des classes de termes au niveau sémantique. Pour cette tâche, il existe beaucoup moins d'outils que pour l'extraction de termes. Certains de ces outils construisent des classes sémantiques sur des bases statistiques, et d'autres sur des principes linguistiques. Nous nous pencherons plus particulièrement sur ces seconds. Zellig ([25], [24]) se fonde sur l'hypothèse harrissienne selon laquelle des régularités syntaxiques dans les contextes des termes permettent de former des classes de termes ayant des sens très proches. Les classes formées par Zellig semblent cependant rester très larges et les regroupements proposés sont plutôt utilisés comme amorce à la constitution de classes sémantiques pertinentes [6]. N. Capponi et Y. Toussaint [12] développent pour leur projet un ensemble de règles de généralisation de structures prédicat-argument qui leur permet de faire des inférences dans le cadre de l'ontologie, mais qui n'est pas adaptée à notre contexte de travail (où les utilisateurs ajoutent des termes sans donner des définitions trop précises). L'outil Caméléon [34] identifie les relations sémantiques entre les termes, à l'aide des marqueurs linguistiques de relations, comme les verbes, qui sont considérés spécifiques au corpus. Caméléon se concentre sur l'identification des relations structurantes d'hyponymie et de méronymie. Autres extracteurs de termes [26] proposent une identification automatique des relations d'hyponymie à l'aide de patrons lexico-syntaxiques génériques (Doc 4) qui est applicable surtout pour l'identification de relations hiérarchiques entre les candidats termes dans la phase de construction initiale de la R.T.O..

(1) NP <sub>0</sub> .such as {NP <sub>1</sub> , NP <sub>2</sub> ,... (and   or)} NP <sub>n</sub> , A tel que X, Y et/ou Z
(2) such NP as {NP ,}* {(or   and)} NP de tels A tels que X, ou Y, et Z.
(3) NP {, NP} * {,} or other NP A, B ou d'autres X
(4) NP {, NP}* {,} and other NP A, B et d'autres X
(5) NP, {,} including {NP,}* {or   and} NP A comprenant, X et   ou Y
(6) NP {,} especially {NP ,}* {or   and} NP A et plus spécifiquement X, et   ou Y

**Doc.4** - Patrons lexico-syntaxiques pour l'hyponymie [26]

Upéry [8], développé pour les résultats de Syntex est proche de nos besoins. Upéry permet de rapprocher des

couples de termes retrouvées dans des contextes syntaxiques identiques. Le module rapproche ensuite les termes, ainsi que les contextes syntaxiques, sur la base de mesures de proximité distributionnelle. De tels résultats sont utilisés pour aider à la construction d'ontologies spécialisées.

L'identification semi-automatique des relations entre les candidats termes est une tâche difficile, mais nécessaire au soutien des utilisateurs pour permettre une navigation facile dans une R.T.O. structurée lors de l'indexation de leurs documents et annotations. En plus d'identifier des candidats termes du domaine, nous avons développé un module qui propose une identification des relations hiérarchiques et horizontales entre les termes du domaine en s'appuyant sur les dépendances syntaxiques établies entre les candidats termes.

Notre module d'identification des relations est basé sur la définition des patrons morpho-syntaxiques qui s'appliquent sur les résultats fournis par Syntex afin d'extraire des relations entre les candidats-termes.

Il faut d'abord identifier les relations hyperonymiques entre les termes. Dans cet objectif, nous utilisons des patrons lexico-syntaxiques de Hearst [26], (Doc.4) et de [34] (Doc. 5).

1. Det N1 et Det N2 (qui, adj, p.passé, p.présent)
2. Tous les N2 sauf det N1
3. Det N1 comme det N2

### **Doc.5 - Patrons de relations conceptuelles [34]**

Nous désirons identifier des relations horizontales entre les candidats termes. Si Caméléon [34] propose d'utiliser le verbe pour structurer d'une manière hiérarchisée les termes, nous utilisons le verbe comme représentant des relations sémantiques entre deux candidats termes. Ainsi, le verbe établit une association (relation prédicative) entre un candidat terme qui joue le rôle de sujet et un autre candidat terme qui joue le rôle d'objet. Le module que nous avons développé propose à l'utilisateur des relations possibles entre les candidats termes que l'utilisateur va choisir d'ajouter ou non à la R.T.O..

Ainsi, nous avons défini une liste de verbes spécifiques au domaine et nous identifions les candidats termes qui jouent le rôle de sujet et d'objet à l'aide des patrons. Entre ces deux candidats termes, le module propose une relation prédicative (Doc 6) que l'utilisateur valide par la suite. Comme dans le cas des candidats termes, les verbes sont liés au point de vue « conception » ou « problème » ou « paramètres-système », même si les termes reliés par un verbe ne font pas partie du même point de vue.

Éviter (Le montage et le démontage, la dégradation du soufflet) – point de vue « problèmes »

Relier (le pivot d'arbre, l'alésage du moyeu) – point de vue « solution »

Déterminer (analyse fonctionnelle, les contraintes dimensionnelles) – point de vue « paramètres du système »

### **Doc.6 - Patrons de relations prédicatives**

Ainsi, « éviter » relie un terme qui fait partie du point de vue « solution » avec un terme qui fait partie du point de vue « défaut ».

Le problème d'identification des relations entre les candidats-termes se posera également si l'utilisateur fait appel au module d'indexation pour indexer ses annotations. Dans un premier temps, le module propose une liste de candidats termes et l'utilisateur va choisir les termes d'indexation qu'il souhaite ajouter à la R.T.O.. Dans une extension future de l'outil, le module pourra proposer des hyponymes/hyperonymes grâce à une méthode distributionnelle comme celle mise en place par Upery. Si deux termes partagent les mêmes contextes dans un corpus, les termes sont considérés comme synonymes ou en relation hyperonymique.

Pour aider l'utilisateur à naviguer dans ces concepts il est crucial de trouver un formalisme adéquat.

## **4.5 Le formalisme**

Le rôle du formalisme choisi pour la RTO dans notre plateforme coopérative est de faciliter la navigation dans la base de documents et d'annotations et de faciliter la maintenance de cette R.T.O. (les membres du projet peuvent la mettre à jour en fonction des besoins pendant l'exploitation du système). Ainsi, les termes doivent être précis, mais compréhensibles par tous les membres du projet et les définitions des termes doivent être partagées par les membres du groupe. Nous avons comparé les formalismes proposés par l'initiative Web Sémantique (OWL, RDF) ainsi que d'autres formalismes issus de l'ingénierie documentaire. La plupart du temps, les formalismes classiques de représentation d'ontologies issus du Web Sémantique demandent l'intervention d'un ou plusieurs experts du domaine pour la mise à jour de ces ontologies. Les définitions proposées sont parfois difficilement compréhensibles par de simples utilisateurs. Or, notre application implique l'intervention de plusieurs membres du groupe dans la mise à jour et la maintenance de l'ontologie. La facilité de navigation, et la possibilité de représenter plusieurs points de vue doivent permettre à ces différents membres d'intervenir. Nous avons donc regardé les formalismes utilisés dans l'ingénierie documentaire.



Le formalisme « Topic Maps » (TM) représente des connaissances sur des ressources (documents, pages Web) grâce à un réseau de termes/thèmes, qui indexe ces ressources. La TM est constituée de thèmes (topic), d'associations entre les thèmes et d'occurrences des thèmes dans les documents. Les occurrences d'un thème sont les documents qui définissent ce thème. Les thèmes ont plusieurs types de caractéristiques telles que : des noms, des occurrences, des rôles (en tant que membres d'associations). L'assignation de ces caractéristiques est considérée valide dans un certain point de vue (scope) ou contexte. Les thèmes sont hiérarchiquement organisés (relations « est un ») et associés par des relations horizontales (« utilisé par »). Le point de vue qui peut être porté par un thème (topic) permet une représentation de la classification par facette.

XML Topic Maps (XTM) [5] reprend les concepts des TM, mais les représente sous format XML, afin de les exploiter et de les partager entre plusieurs applications. Issu des problématiques de gestion de documents, le formalisme XTM est orienté navigation pour l'utilisateur. Les thèmes (topics) sont définis par de simples URL afin que tous les utilisateurs partagent la même définition. Les documents en XTM doivent être clairs et lisibles par des utilisateurs humains.

Cette norme ISO est définie pour constituer des ontologies semi-formelles permettant une grande portabilité et une maintenance simple par l'utilisateur (qui peut alors ajouter des sujets et des ressources). Puisque les TM ne nécessitent pas de définitions précises des concepts et sont conçues pour soutenir l'utilisateur dans sa navigation Web, nous avons adopté ce formalisme pour représenter nos concepts. Dans un premier temps, nous représentons les multiples aspects liés au domaine de l'aéronautique (conception, défauts, utilisation etc.). Dans une extension future d'AnTCoW, nous allons rajouter des points de vue :

- argumentatif : pour des annotations utilisées dans des phases de négociation ou de construction d'une solution.
- organisationnel : prenant en compte le rôle de l'annotateur au sein de l'organisation.

La construction de la R.T.O. du domaine a demandé quelques modifications sur la structure de la TM. Ainsi, les ressources qui stockent les occurrences de chaque topic ne sont pas les documents originaux mais les occurrences de chaque terme avec leurs contextes gauche et droit. Nous avons regroupé les contextes par l'information syntaxique associée (les types de constituants syntaxiques qui se combinent avec le terme). Cette information servira pour une identification semi-automatique des termes synonymes ou hyponymes.

La section suivante illustre la construction de la R.T.O. du domaine de la conception en mécanique aéronautique à l'aide des extracteurs de termes adaptés pour cette tâche et l'utilisation de cette R.T.O. pour l'indexation.

## 5 Scénario d'indexation basé sur l'utilisation de notre R.T.O.

S'inspirant de notre contexte d'étude, nous allons voir comment notre outil peut être utilisé pour indexer les échanges annotatifs entre les membres du projet. Les concepteurs du moteur pour le projet Air Campus ont alors plusieurs documents qu'ils souhaitent mettre en commun. Certains membres ayant déjà travaillé sur des documents intéressants veulent aussi informer les autres membres de problèmes que soulèvent certains documents, que ce soit au niveau de la planification du travail ou de la faisabilité technique.

### 5.1 Un extrait de la R.T.O.

Les membres ayant regroupé un ensemble de documents, ils ont constitué un corpus de référence. Il est alors possible de commencer la construction de la R.T.O. hors-ligne. L'administrateur du système utilise un ensemble de candidats termes et de relations proposées par les outils de T.A.L.N. pour structurer la R.T.O.. Chaque terme qui sera inséré dans la R.T.O. est identifié comme index d'un document ou d'une annotation du système. Il s'agit donc de choisir un ensemble de termes et de relations pertinents pour l'indexation. Chaque terme est identifié par son contexte syntaxique (les constituants syntaxiques qui vont se combiner avec le candidat terme). Cette information sera exploitée si l'utilisateur demande de l'aide pour classer ses propres annotations.

Avec les outils, le corpus de référence et les techniques d'extraction et de structuration exposés précédemment, l'administrateur crée une ontologie constituée de 56 associations prédictives reliant 150 termes regroupés en 3 points de vue: (1) paramètres système, (2) problèmes et (3) solution.

Une fois qu'une R.T.O. existe, les utilisateurs vont pouvoir indexer les annotations qu'ils attachent à un document et permettre ainsi aux autres participants de retrouver les annotations des autres membres du groupe, de naviguer dans la R.T.O. et de visualiser les annotations selon un thème choisi. Les relations entre les termes servent aussi à retrouver des annotations liées au même sujet.

### 5.2 Scénario d'indexation

Si l'utilisateur veut partager ses annotations avec les autres membres du groupe, il lui faut indexer ses annotations à l'aide de la R.T.O. déjà créée. Les outils d'extraction de termes analysent le corps de l'annotation, ainsi que son point d'ancrage (document, paragraphe) et proposent à l'utilisateur des candidats termes possibles comme index (qui se trouvent déjà dans la R.T.O.). L'utilisateur peut choisir un des index proposés qui existe déjà dans la R.T.O. ou décider d'intégrer un nouveau

candidat dans la R.T.O.. Dans tous les cas, c'est l'utilisateur qui décidera de la mise à jour de la R.T.O..

Pour indexer les méls présentés en section 2.2. par exemple, nous avons utilisé l'extracteur de termes qui englobe Syntex et nos outils de structuration.

La liste de candidats proposée parmi les groupes nominaux extraits est la suivante :

1. « les CAO des nouvelles pièces »
2. « Une version définitive de "platine moteur" »
3. « des perçages »
4. « les dimensions des orielles inférieures »
5. « une poche sur la face d'appui pour passage d'un joint cf doc "montage cage réducteur". »
6. « Deux supports de réducteur en acier à tailler dans la masse, ».
7. « l'usinabilité de ces deux pièces ».
8. « Le plan incliné de la nervure »
9. « l'état de surface ».
10. « les concentrations de contraintes »
11. « la mise en plan en cas de souci »

L'utilisateur, en fonction du contenu de son annotation, décide de conserver les termes 2, 5 et 6 (soulignés) pour index de son annotation. Ces termes sont plutôt liés au point de vue « système » et seront donc reliés en tant que tels dans la R.T.O.. Un autre utilisateur recherchant plus tard des informations sur le « système » pourra alors retrouver facilement ces annotations qui peuvent être cruciales dans l'activité de conception.

Pour ce cas, il n'y avait pas de verbe du domaine pour identifier une association entre les deux termes candidats ; on ne pourra donc pas suivre de relations spécifiques pour retrouver ses annotations.

## 6 Conclusion

Dans le cadre d'activités distribuées, il est important de ne pas se contenter de s'envoyer ou de partager des documents ; il faut aussi assister les phases de négociation, de discussion sur ces documents. Ces phases peuvent être matérialisées par des annotations, fragments de documents qui contribueront à la fois à la réalisation de la tâche et à l'élaboration de nouveaux documents. Le fait de conserver les traces des interactions matérialisées par les annotations permettra par ailleurs de tracer les processus argumentatifs de prise de décision collective. Conserver les annotations fait naître la problématique de l'indexation de ces annotations afin de permettre une visualisation aisée des problématiques au cours du projet ou a posteriori.

Nous proposons donc un outil d'annotation qui permet d'échanger autour de documents numériques, mais aussi d'indexer les annotations ancrées à ces documents. Cette indexation peut être fondée sur une classification construite sur les documents du domaine du projet pour être adaptée à l'activité du groupe et au lexique utilisé dans leur activité. Dans cet article, nous avons décrit la structuration d'une

R.T.O. à partir de documents textuels. La R.T.O. structurée selon des patrons lexico-syntactiques, est formatée en Topic-Maps (T.M.).

Il s'agit maintenant d'optimiser l'indexation des annotations en étendant les processus de mises en correspondance et la couverture des index disponibles à de nouveaux points de vue. L'indexation actuellement effectuée par des mises en correspondance simples d'un contexte d'ancrage structuré de l'annotation avec les termes de la R.T.O. peut être effectuée par des mises en correspondance au niveau des concepts de la T.M., des occurrences et du contexte des occurrences. Il est aussi possible d'affiner l'indexation en constituant des R.T.O. de nouveaux points de vue tels que l'argumentation ou l'organisation. Pour le moment, il s'agit d'adapter les extracteurs de termes pour identifier les marqueurs de discours. Ces R.T.O. seront ensuite intégrées à l'outil qui sera testé dans un contexte de conception coopérative.

## Remerciements

Ce projet est soutenu par un financement du projet TCAN Mediannote - département STIC du CNRS et d'un co-financement du Conseil Régional Champagne Ardenne et du Fonds Social Européen.

## Références

- [1] Aussenac-Gilles N., Condamines A., (2004) Documents électroniques et constitution de ressources terminologiques ou ontologiques, in J. Charlet et J.-M. Salaun (eds) *Information-Interaction-Intelligence*, Volume 4, n°1, p.75-93
- [2] Bachimont B., (2000) Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en Ingénierie des connaissances. In *Ingénierie des connaissances, Evolutions récentes et nouveaux défis*, Jean Charlet, Manuel Zacklad, Gilles Kassel, Didier Bourigault (eds), Eyrolles, pp 305-323
- [3] Barcellini F., Détienne F., Burkhardt J.-M., Sack W., 2005, Thematic coherence and quotation practices in OSS design-oriented online discussions in *Proceedings of GROUP 2005* p.177-186
- [4] Baroni M. and Bernardini S., 2004, BootCaT: Bootstrapping corpora and terms from the web, in *Proceedings of LREC 04*
- [5] Biezunski, M., Bryan, M., Et Newcomb, S. R., (1999) *Topic Maps, spécification ISO/IEC 13250*, 3 Décembre 1999.
- [6] Bourigault D. Et Jacquemin C., (2000), Construction de ressources terminologiques, in *Ingénierie des langues*, Jean-Marie Pierrel (ed.), Hermès, pp 215-233
- [7] Bourigault D., (1996) LEXTER, a Natural Language Processing tool for terminology extraction, in *Proceedings of the 7th EURALEX International Congress*, Goteborg.
- [8] Bourigault D., (2002), UPERY : un outil d'analyse distributionnelle étendue pour la construction d'ontologies à partir de corpus, in *Actes de la 9ème conférence annuelle sur le Traitement Automatique des Langues (TALN 2002)*, Nancy, 2002, pp. 75-84

- [9] Bourigault D., Fabre C., (2000) Approche linguistique pour l'analyse syntaxique de corpus, *Cahiers de Grammaires*, n° 25, Université Toulouse - Le Mirail, p. 131-151
- [10] Bourigault D., Fabre C., Frérot C., Jacques M.-P. & Ozdowska S. (2005), Syntex, analyseur syntaxique de corpus, in *Actes des 12èmes journées sur le Traitement Automatique des Langues Naturelles*, Dourdan, France
- [11] Buitelaar P., Olejnik D., Hutanu M., Schutz A., Declerck T., Et Sintek, M. (2004), Towards Ontology Engineering Based on Linguistic Analysis, in *Proceedings of LREC'2004*, Lisbon, may 2004, ISBN 2-9517408-1-6, pp.7-11
- [12] Capponi N. Et Toussaint Y., (2000), Interprétation de classes de termes par généralisation de structures, in *Ingénierie des connaissances, Evolutions récentes et nouveaux défis*, Jean Charlet, Manuel Zacklad, Gilles Kassel, Didier Bourigault (eds), Eyrolles, pp 335-353
- [13] Carroll J., Neale D., Isenhour P., Rosson M, McCrickard D., 2003, Notification and awareness: synchronizing task-oriented collaborative activity, in *International Journal of Human-Computer Studies*, v.58 n.5, p.605-632.
- [14] Cimiano, P. Hotho, A., Et Staab S. (2004), Clustering Concept Hierarchies from Text, in *Proceedings of LREC'2004*, Lisbon, ISBN 2-9517408-1-6, pp. 1721-1724.
- [15] Clark H. 1996 *Using Language*. Cambridge University Press, Cambridge.
- [16] Daille B., Identification des adjectifs relationnels en corpus in *Actes de la Conférence de Traitement Automatique du Langage Naturel (TALN'99)*, Cargèse, 1999
- [17] Darses, F. (2001) Converger vers une solution en situation coopérative de conception : analyse cognitive du processus d'argumentation. In F. Darses (Ed.) Modeling Cooperative Activities in *Design Proceedings of the 10th Atelier du Travail Humain*, 27-28 juin 2001, Paris, France: INRIA.
- [18] David S. et Plante P. (1990), De la nécessité d'une approche morpho-syntaxique dans l'analyse de textes. *Intelligence Artificielle et Sciences Cognitives au Québec*, 3(3):140-154
- [19] Denoue, L., et Vignollet, L. (2000), An annotation tool for Web browsers and its applications to information retrieval, in *proceedings of RIAO 2000*
- [20] Dzbor, M. - Motta, E. - Domingue, J. B. (2004), Opening Up Magpie via Semantic Services. In *3rd ISWC*, November 2004, Japan
- [21] Frath P., Oueslati R. Et Rousselot F., (2000), Identification de relations sémantiques par repérage et analyse de cooccurrences de signes linguistiques, in *Ingénierie des connaissances. Évolutions récentes et nouveaux défis*. Eds. Jean Charlet, Manuel Zacklad, Gilles Kassel, Didier Bourigault. Eyrolles, Paris, pp 291-304
- [22] Gangemi, A., Guarino, N., Masolo, C., Et Oltramari, (2003) A. Sweetening WordNet with DOLCE, *AI Magazine* 24(3): Fall 2003, 13-24
- [23] Gómez-Pérez, A., Manzano-Macho, D., Alfonseca, E., Núñez, R., Blacoe ,I., Staab, S., Corcho, O., Ding, Y., Paralic, J., Troncy, R.,(2003) *Ontoweb, Deliverable 1.5: A survey of ontology learning methods and techniques, IST Project IST-2000-29243 OntoWeb*, juin 2003
- [24] Habert B. & Fabre C. (1999). Elementary dependency trees for identifying corpus-specific semantic classes, in *Computer and the Humanities*, 33(3), 207-219
- [25] Habert, B., Nazarenko, A., (1996). La syntaxe comme marche-pied de l'acquisition de connaissances : bilan critique d'une expérience, in *Actes des Journées d'Acquisition de Connaissances (JAC)*
- [26] Hearst M.A. (1992) Automatic acquisition of Hyponyms from large text corpora. In *Proceedings of the Fourteenth International Conference on Computational Linguistic, (CoLing'92)* Nantes, France, July 1992 pp 539-545.
- [27] Jacquemin C. (1997), Variation terminologique : Reconnaissance et acquisition automatiques de termes et de leurs variantes en corpus. Mémoire d'habilitation à diriger des recherches en informatique fondamentale, Université de Nantes
- [28] Jacquemin C. Et Bourigault D. (2003), Term Extraction and Automatic Indexing, in Mitkov R. (ed), *The Oxford Handbook of Computational Linguistics*, Oxford University Press, 2003, pp. 599-615
- [29] Lortal G., Lewkowicz M., Todirascu-Courtier A., 2005, Annotation: Textual Media for Cooperation, in *Proceedings of Annotation for Cooperation Workshop* November 24-25th (p.41-50)
- [30] Maedche, A, Staab, S. (2001) Ontology Learning for the Semantic Web, *IEEE Intelligent Systems, Special Issue on the Semantic Web*
- [31] Martin, G., Détienne, F., Lavigne, E. (2001) Analysing viewpoints in design through the argumentation process. In *proceedings of INTERACT'2001*, Tokyo, Japan, July 9-13.
- [32] Pierra G., Dehainsala H., Ait Ameer Y., Bellatreche L., (2005) Bases de données à base ontologique. Principe et mise en œuvre, in *Ingénierie des Systèmes d'Information VOL 10/2 – 2005* Hermès Lavoisier
- [33] Rousselot, F., Frath, P., Et Oueslati, R. (1996), Extracting concepts and relations from Corpora. In *Proceedings of the Workshop on Corpus-oriented Semantic Analysis, European Conference on Artificial Intelligence, ECAI 96*, Budapest.
- [34] Séguéla P., Aussenac-Gilles N., (1999) Extraction de relations sémantiques entre termes et enrichissement de modèles du domaine, in *Actes de IC'99 (Ingénierie des Connaissances)*, Paris, pp 79-88
- [35] Simone C. and Sarini M., (2001), Adaptability of Classification Schemes in Cooperation: What does it mean?, in *Proceedings of ECSCW 2001*, 19-38.
- [36] Sumner T., Buckingham Shum S., Wright M., Bonnardel N., Piolat A. & Chevalier A. (2000), Re-designing the peer review process : A developmental theory-in-action. In R. Dieng, A. Giboin, G. De Michelis & L. Karsenty (Eds.), *Designing cooperative systems: The use of theories and models* (pp. 19-34). Amsterdam : I.O.S. Press, 2000.
- [37] Vossen P. (1998), *EuroWordNet A Multilingual Database with Lexical Semantic Networks*. Kluwer Academic publishers.
- [38] Wilson M., Daniels D., Phillips J., (2004), *Annozilla*, <http://annozilla.mozdev.org/>
- [39] Zacklad M., Lewkowicz M., Boujut J-F., Darses F., Et Détienne F. (2003), Formes et gestion des annotations numériques collectives en ingénierie collaborative, in *Actes des journées Ingénierie des Connaissances 2003*, Laval
- [40] Zacklad, M. (to be published) Documentarization processes in Documents for Action (DofA): the status of annotations and associated cooperation technologies, in *JCSCW*
- [41] Le thésaurus EAU - <http://eaudoc.oieau.fr/sie/gedoieau.asp>
- [42] Le thésaurus CELEX: <http://europa.eu.int/celex/>