

Proposition de projet INRIA

INRIA Rhône-Alpes — thème 3A

EXMO

Échanges de connaissance structurée médiatisés par ordinateur

Computer-mediated exchange of structured knowledge

mercredi 23 octobre 2002

1. CONTEXTE ET MOTIVATIONS	3
1.1 OBJET DU PROJET □ INTEROPÉRABILITÉ DANS LA COMMUNICATION DE CONNAISSANCE FORMALISÉE	3
1.2 MÉTHODOLOGIE.....	5
1.3 SITUATION.....	7
2. PERSPECTIVES APPLICATIVES	10
2.1 COMMUNICATION, INTEROPÉRABILITÉ SYNTAXIQUE ET TRANSFORMATIONS	10
2.2 INGÉNIERIE DES SYSTÈMES DE TRANSFORMATIONS.....	10
2.3 PROPRIÉTÉS, INTEROPÉRABILITÉ SÉMANTIQUE ET COMPOSITION	12
2.4 TECHNOLOGIES POUR UN «WEB SÉMANTIQUE□	13
3. AXES DE RECHERCHE.....	14
3.1 REPRÉSENTATION DU CONTENU	14
3.2 PROPRIÉTÉS DES TRANSFORMATIONS.....	14
3.3 INTEROPÉRABILITÉ SÉMANTIQUE	17
3.4 VERS UNE INFRASTRUCTURE SÛRE DE PARTAGE DE TRANSFORMATIONS	20
3.5 SÉMIOLOGIE COMPUTATIONELLE	22
4. DÉVELOPPEMENT LOGICIEL.....	24
4.1 TRANSMORPHER □ EXPRESSION ET EXÉCUTION DES FLUX DE TRANSFORMATIONS.....	24
4.2 DLML : DESCRIPTION LOGIC MARKUP LANGUAGE	25
5. RELATIONS NATIONALES ET INTERNATIONALES.....	27
5.1 RELATIONS LOCALES	27
5.2 RELATIONS AU SEIN DE L'INRIA	27
5.3 RELATIONS NATIONALES	28
5.4 RELATIONS INTERNATIONALES.....	29
6. COMPOSITION DE L'ACTION EXMO.....	31
7. PUBLICATIONS (DEPUIS LA CRÉATION DE L'ACTION EXMO).....	32

Résumé

EXMO part du principe que, dans les futurs systèmes d'information, la connaissance formalisée sera échangée de façon routinière. Dans le processus de communication, l'ordinateur peut introduire une plus-value à son rôle de médium et de mémoire en accomplissant des tâches comme le formatage, le filtrage, la catégorisation, le test de consistance ou la généralisation de la connaissance. Ces manipulations peuvent être vues comme des transformations. Cette approche commence à se développer avec la généralisation de l'utilisation de langages d'échange standardisés (XML) lors de la communication en réseau. En contrepartie, les utilisateurs devront exiger une plus grande assurance sur les transformations opérées. EXMO a pour but le développement d'outils théoriques et logiciels pour aider à l'échange d'éléments de connaissance formalisée.

Notre angle d'attaque principal consiste à examiner les propriétés que doivent satisfaire les transformations appliquées aux représentations formalisées. Parmi ces propriétés, on trouve la préservation du contenu ou de la structure, la traçabilité des sources ou, au contraire, la confidentialité. On cherche à élaborer une «Théorie générale des transformations» fondée sur les propriétés satisfaites par les transformations plutôt que sur les transformations elles-mêmes.

D'une part, nous nous attachons à montrer que cette problématique s'applique à un ensemble varié de contextes dépendant des langages, propriétés et transformations impliqués. Ainsi, nous étudions la préservation de l'information modélisée par des treillis ou la préservation de scénario de documents multimédia durant l'adaptation.

D'autre part, nous nous intéressons plus particulièrement aux propriétés sémantiques lors de la transformation de la connaissance d'un langage à un autre. La question essentielle est-les conséquences d'une représentation initiale sont-elles les transformées de celles de la représentation transformée? Pour cela nous établissons les relations entre plusieurs types de propriétés sémantiques, connues ou non, en rapport avec celles-ci. Nous avons aussi étudié des cadres plus restreints (famille de langages, patrons) permettant d'utiliser ces relations et d'établir plus facilement les propriétés des transformations.

Enfin, sur un plus long terme, nous avons l'ambition d'explorer des propriétés appelées «Émiologiques», c'est-à-dire qui concernent l'interprétation par un utilisateur humain de la représentation communiquée, et leur préservation.

Du point de vue pratique, nous avons décrit la syntaxe et la sémantique de plusieurs langages de représentation de connaissance (TROEPS, logiques de descriptions...) dans un format XML. Nous avons développé un langage capable de décrire des transformations composées et de les exécuter. Notre but est maintenant de tirer parti des descriptions de la sémantique et des transformations pour vérifier des preuves de certaines propriétés des transformations avant de les utiliser. L'autre direction consiste à construire la preuve en fonction de celles des transformations initiales et des propriétés des constructeurs utilisés.

Les applications anticipées sont l'ingénierie des systèmes de transformations (où le système d'information est compris comme un flux de transformations) et l'infrastructure du «Web sémantique».

1. CONTEXTE ET MOTIVATIONS

Cette proposition de projet a pour but d'indiquer et de motiver les directions de recherches d'un éventuel projet EXMO pour les années à venir. On distinguera l'action EXMO, créée en 2000, qui concerne le passé et le présent, et le projet EXMO qui concerne les réalisations à venir dans un éventuel projet.

Les activités du projet EXMO sont motivées par rapport à l'état de l'art et nos expériences passées et situées dans la ligne de recherches à laquelle ils se rattachent et les développements actuels autour du «Web sémantique» (§1.1). Nous introduisons alors succinctement notre approche (§1.2) avant de présenter les enjeux stratégiques, le contexte international et celui propre à l'INRIA (§1.3).

Après cette introduction, nos activités seront motivées dans leur contexte applicatif envisagé (§2). Nous exposerons ensuite les travaux de recherche que nous comptons mener (§3) et les développements logiciels associés (§4). Les relations avec d'autres partenaires seront enfin présentées (§5) suivies par la présentation actuelle de l'effectif de l'action EXMO (§6).

1.1 Objet du projet : interopérabilité dans la communication de connaissance formalisée

L'activité du projet EXMO s'inscrit dans le domaine de la représentation de connaissance dont le but est de produire des modèles et des outils pour représenter et manipuler informatiquement les données et connaissances. Cette discipline, partie prenante de l'intelligence artificielle, est née de la volonté de développer des «Systèmes intelligents».

1.1.1 Problématique

Le but de notre activité est de prendre au sérieux l'idée que la «Connaissance» formalisée est communiquée par le biais d'ordinateur. Ce faisant, l'ordinateur peut apporter une valeur ajoutée à son rôle de mémoire ou de médium en accomplissant des fonctions plus avancées (détection de similarité, de cohérence, généralisation, filtrage et autres transformations). Il n'est donc pas confiné au simple rôle de tuyau.

De plus, nous posons que cela vaut la peine d'exprimer la connaissance de manière formelle sur ordinateur, non pour les besoins exclusifs de l'ordinateur, mais pour communiquer à d'autres cette connaissance (sous une forme adaptée à l'interlocuteur).

Ce qui nous importe est que la sortie du système soit conforme aux attentes et, en particulier, que les manipulations informatiques préservent le sens du message.

Cette problématique de l'utilisation de la représentation de connaissance à autre chose que la résolution de problèmes était déjà présente au sein du projet SHERPA depuis sa création¹. Mais le surgissement du World wide web a donné une importance beaucoup plus grande à la fois à l'utilisation de la représentation de connaissance pour aider à structurer le web et la manipulation collaborative de connaissance sur le web. Nous avons étudié ce dernier sujet

¹ Jérôme Euzenat, Représentation de connaissance par objets, dans Roland Ducournau, Jérôme Euzenat, Gérald Masini, Amedeo Napoli (éds.), Langages et modèles à objets: état et perspectives de la recherche, INRIA, Rocquencourt (FR), pp293-319, 1998

dans SHERPA. L'activité principale d'EXMO se situe donc dans le prolongement de nos travaux antérieurs dans le sens où l'on examine les transformations de représentations et donc le rapport entre deux représentations d'une même situation².

Si beaucoup des résultats obtenus dans le domaine de la représentation de connaissance sont transférables presque tels quels dans le domaine du web, la pression y est aussi très importante (des équipes importantes et des entreprises avec des moyens considérables sont capables d'intégrer des innovations très rapidement). La création de l'action EXMO en 2000 correspondait au besoin de prendre du recul par rapport aux besoins immédiats des utilisateurs potentiels et de développer une problématique utile à plus long terme et sur laquelle nous pourrions prendre de l'avance.

La réelle difficulté du domaine est un problème d'interopérabilité (ou de communication). C'est ceci qu'exprime le titre du projet. Ce problème est lié à la confrontation de plusieurs représentations développées pour répondre à des besoins différents et à leur exploitation conjointe. Le problème principal est d'établir dans quels cas (langages, particularités des représentations), quelles propriétés (préservation du sens, du contenu, de la structure...) peuvent être assurées.

La problématique d'EXMO est donc le développement d'outils (théoriques et informatiques) permettant l'organisation, la manipulation, la présentation et la combinaison d'éléments de connaissance structurée de telle sorte que le résultat soit adapté à et intelligible aux interlocuteurs.

Il y a beaucoup à intégrer en provenance d'autres champs de l'informatique (preuve de programme, théorie de l'information, systèmes de réécriture, systèmes multi-agents...) ou d'autres disciplines (sémantique, pragmatique, sémiologie...). Ce sujet de recherche est l'occasion de poser d'éternelles questions sur la nature de la connaissance, du signe et de la communication. Mais le faisant dans un cadre particulier, contraint par la médiation informatique, il semble possible d'y apporter des réponses plus contextuelles.

1.1.2 Le «Web sémantique»

L'utilisation des technologies d'Internet est l'opportunité pour les entreprises d'accéder à et de partager la connaissance bien souvent difficilement accessible sous forme documentaire. Cependant, les limites de cette approche apparaissent rapidement. L'organisation des sites se révèle une tâche coûteuse et la recherche en texte intégral peu efficace. La recherche et l'interrogation d'un site en s'appuyant sur le contenu des documents sont une nécessité et les formalismes de représentation de connaissance sont de bons candidats pour représenter ce contenu. La représentation du contenu permettra de le manipuler pour faire de la recherche par analogie, par spécialisation, par similitude, etc. [Euzenat2002b].

La proposition d'un «Web sémantique»³ consiste à adjoindre au web actuel (informel) des annotations (rédigées dans des langages exploitables par une machine) liées entre elles comme au web informel. Exploiter ce «Web sémantique» demandera donc d'appréhender ces représentations formelles diversifiées [Euzenat2002a].

L'ensemble des travaux d'EXMO a généralement pour but de concourir à une meilleure appréhension des contenus. La problématique de l'action EXMO se trouve donc au cœur de la

² Jérôme Euzenat, Représentations de connaissance: de l'approximation à la confrontation, Mémoire d'habilitation à diriger des recherches, Université Joseph Fourier, Grenoble (FR), 1999 <ftp://ftp.inrialpes.fr/pub/sherpa/theses/hdr-euzenat.ps.gz>

³ Tim Berners-Lee, James Hendler, Ora Lassila, The semantic web, *Scientific american* 284(5):35-43, 2001, <http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html>

mise en œuvre du «Web sémantique». Cela concerne, bien entendu, les travaux sur l'intelligibilité des connaissances communiquées mais aussi les travaux sur la transformation de représentations formelles. Les travaux sur le «Web sémantique» se basent actuellement sur la notion d'ontologie (que l'on peut rapidement décrire comme un schéma conceptuel de base de connaissance ou une axiomatisation du domaine). Même s'il existe un jour un langage commun de représentation, il faudra importer les ontologies décrites dans d'autres langages d'une façon qui respecte la sémantique des langages manipulés.

Apporter des solutions à ce problème fait partie des ambitions d'EXMO (voir §3.3). Notre contribution au «Web sémantique» se situe donc principalement dans la conception de technologies permettant de le construire.

Nous avons été très actif dans la constitution de la communauté autour de ce thème, en particulier en organisant le premier "semantic web working symposium" à Stanford en 2001 (250 inscrits) et en éditant le livre qui en est tiré [Cruz2002a]. Nous avons par ailleurs contribué à la réflexion stratégique sur ce thème en organisant le workshop stratégique conjoint à l'union européenne et à la National Science Foundation et en éditant le rapport [Euzenat2002a].

1.2 Méthodologie

Cette section introduit le contexte de nos travaux en rappelant les bases sur lesquelles sont construits les langages de représentation de connaissance actuels (§1.2.1), en introduisant l'approche qu'Exmo compte mettre en œuvre dans l'ensemble de ses développements (§1.2.2) et en identifiant les points sur lesquels nous ferons porter notre effort (§1.2.3).

1.2.1 Théorie des modèles

Après les années de développement empirique, une rationalisation du domaine de la représentation de connaissance a eu lieu, principalement liée à l'étude de la sémantique des formalismes utilisés. Depuis une quinzaine d'années, la sémantique des langages de représentation de connaissance (logiques de descriptions, graphes conceptuels et représentations de connaissance par objets) a été étudiée. Elle est définie en général à l'aide de la théorie des modèles, issue elle-même de la logique⁴.

On considérera ici un langage L comme un ensemble d'expressions syntaxiques, souvent définies inductivement en appliquant des constructeurs sur d'autres expressions. Une représentation (r) est un ensemble d'expressions. On définit inductivement sur la structure du langage une fonction d'interprétation (I) vers un ensemble nommé domaine d'interprétation (D). Cette fonction reflète la construction du sens d'une expression en fonction de ses composants. Les assertions de ce langage sont satisfaites par une interprétation si elles remplissent une condition (en général être interprétées dans un sous-ensemble particulier du domaine). Dans ce cadre un modèle est une interprétation dans laquelle tous les axiomes sont satisfaits. Une expression (\square) est alors une conséquence d'un ensemble d'axiomes (r) si elle est satisfaite par tous leurs modèles (ce qui est noté $r \models \square$).

Pour un ordinateur, le but consiste à déterminer si une expression particulière (par exemple une requête) est conséquence des axiomes (par exemple, la base de connaissance considérée). Pour cela on développe des systèmes exécutables, appelés démonstrateurs, qui peuvent être fondés sur l'interprétation de règles d'inférence ou sur des programmes plus classiques. Ils

⁴ Jérôme Euzenat, Sémantique des représentations de connaissance, Notes de cours, université Joseph Fourier, Grenoble (FR), 2002, 145p. <ftp://ftp.inrialpes.fr/pub/exmo/tmp/src.ps.gz>

permettent de déduire des théorèmes (ce qui est noté $r \vdash \square$). Ces démonstrateurs vérifient la propriété de correction s'ils ne répondent positivement qu'en cas de conséquence et la propriété de complétude s'ils répondent positivement pour toutes les conséquences. Cependant, suivant le langage et sa sémantique, la décidabilité — c'est-à-dire la possibilité de créer de tels démonstrateurs — n'est pas garantie. Même dans le cas de langages décidables, leur complexité algorithmique peut être prohibitive.

Pour cela on est amené à réaliser un compromis entre l'expressivité d'un langage et la complexité ou la complétude des démonstrateurs associés. Ce choix s'est traduit par la définition de langages à l'expressivité limitée — comme les graphes conceptuels ou les représentations par objets — ou de familles de langages modulaires dont les démonstrateurs peuvent être étendus — comme les logiques de descriptions.

L'action EXMO s'appuie d'abord sur des langages dont la sémantique est ainsi définie afin d'établir les propriétés des manipulations informatiques appliquées aux représentations.

1.2.2 Transformations et propriétés

Dans notre approche, les traitements appliqués aux représentations sont qualifiés de transformations. Une transformation est un moyen algorithmique d'engendrer une représentation à partir d'une autre (pas forcément dans le même langage). On se concentrera sur des transformations obtenues par assemblage de transformations élémentaires dont on ne se préoccupe pas de la description. On connaît par contre leurs entrées et sorties ainsi que des propriétés liant les entrées aux sorties.

On définit un système de transformations par l'ensemble de ses transformations élémentaires et l'ensemble de ses assembleurs de transformations. Un flux de transformations est un ensemble d'instances de transformations élémentaires assemblées et reliées par des canaux. Un flux de transformations est lui-même une transformation. Plus concrètement, nos travaux concernent des transformations syntaxiques de documents XML ("eXtensible Markup Language") encodant des langages de représentation de connaissance. Nous cherchons à exploiter le langage de transformations XSLT ("XML Stylesheet Language Transformations"⁵) recommandé par le W3C, pour lequel nous avons proposé un langage d'assemblage de transformations (voir §4.1).

Le but d'EXMO est d'étudier les propriétés des transformations et leur combinaison. Une propriété est simplement un prédicat booléen portant sur une transformation (par exemple, « \square préserver l'information \square est un tel prédicat — il est vrai ou faux d'une transformation — et se vérifie s'il existe un procédé algorithmique permettant, pour toute représentation r , de la retrouver à partir de sa transformée $\square(r)$). La question intéressante est alors de vérifier, pour un flux de transformations donné qu'il vérifie bien une propriété. À côté des propriétés liées aux transformations individuelles, on trouve les propriétés des systèmes (étant donné un ensemble de constructeurs de systèmes d'informations, la préservation de l'information est-elle décidable et à quel prix). Nous identifions deux types de propriétés : des propriétés de préservation et des propriétés à motifs qui consistent à tester si la présence d'un motif dans la représentation de départ entraîne la présence de ce motif dans la représentation d'arrivée.

Les propriétés de préservation nous intéressent plus particulièrement. Elles peuvent permettre de garantir la préservation d'un ordre entre la représentation d'entrée (r) et la représentation de sortie ($\square(r)$) comme son anti-préservation. Pour cela on peut identifier des propriétés :

⁵ James Clark (éd.), XSL transformations (XSLT) version 1.0, Recommendation, W3C, 1999 <http://www.w3.org/TR/xslt>

syntaxiques : fondées sur l'organisation des éléments syntaxiques (ordre, encodage...);
sémantiques : fondées sur les notions de modèles ou de conséquence
sémiologiques : fondées sur l'interprétation en tant que signe des objets manipulés.
Dans un premier temps, nous travaillons plus spécifiquement les propriétés sémantiques.

1.2.3 Vers où

Dans le cadre de cette proposition de projet, il semble que trois points particuliers devront être l'objet de notre attention afin de constituer une contribution importante à l'interopérabilité

- la possibilité d'analyser les transformations en unités assemblées par des opérateurs précis
- la proposition d'un ensemble de propriétés utiles et l'étude de leur comportement vis-à-vis de ces opérateurs
- la capacité, en ce qui concerne les transformations élémentaires, à prouver les propriétés sémantiques que nous aurons identifiées.

C'est principalement à ces trois problèmes que nous nous attaquerons dans les années à venir.

1.3 Situation

Nous situons nos travaux dans le contexte de l'INRIA et des institutions des recherche en général (§1.3.1) avant de caractériser l'état de l'art (§1.3.2) et notre originalité (§1.3.3) et nos limites (§1.3.4) vis-à-vis de celui-ci.

1.3.1 Contexte institutionnel

La proposition de projet EXMO s'inscrit dans deux des cinq défis scientifiques du plan stratégique de l'INRIA et de son contrat quadriennal («Concevoir les nouvelles applications exploitant le web et les bases de données multimédia» et «Savoir produire des logiciels sûrs»). Elle est additionnellement au cœur de l'une des six orientations scientifiques du projet d'établissement du CNRS («Information, communication et connaissance»). Par ailleurs, EXMO a pour but de poursuivre une réflexion de nature fondamentale et présente une ouverture vers les sciences humaines.

Cette proposition concerne un domaine en pleine effervescence au niveau international, en particulier dans le «Web sémantique» (programme DAML de la DARPA, réseau ONTOWEB, task force «Web sémantique» d'INTAP, activité «Semantic web» du W3C, programme IST «Semantic-based knowledge systems») et national (action spécifique «Web sémantique» du CNRS). Les thèmes «developing safe transformation methods», «experimenting transformation infrastructure», «checking proofs and policies» font partie des recommandations faites par le workshop stratégique UE-NSF sur le «Web sémantique» que nous avons organisé [Euzenat2002a]. «Relationships between semantic models» en constitue l'une des quatre directions de recherche mises en avant. Le travail accompli ces dernières années nous place en position de jouer un rôle important dans ces domaines. Cependant, les moyens consacrés par l'INRIA à cette problématique restent modestes comparés à ceux mobilisés ailleurs (Karlsruhe, Amsterdam, Stanford, University of Southern California...).

1.3.2 État de l'art

L'état de l'art en matière de transformation est la programmation directe des transformations sans rechercher spécifiquement les propriétés qui en sont attendues. Quelques efforts permettent de prendre en compte la sémantique en embarquant des démonstrateurs dans les moteurs de transformation⁶. Cela permet de disposer de transformations dont le fonctionnement est en accord avec la sémantique mais dépend toujours de la programmation.

D'autres groupes se sont intéressés au typage de XML⁷ puis, par extension, au typage des transformations de documents XML^{8,9,10}. C'est donc à certaines propriétés particulièrement bien étudiées que s'attaquent ces travaux. Ils concourent à la définition de transformations s'exécutant sans problème.

En ce qui concerne les propriétés sémantiques, on trouve surtout des groupes dont le but est de transformer certaines représentations du web sémantique vers leurs outils (en général des moteurs d'inférence) et qui s'attachent à montrer le rapport sur une série de transformations particulières. C'est le cas des travaux de l'université de Manchester sur OWL et le démonstrateur FaCT.

1.3.3 Originalité

En tant que projet intellectuel, nous voulons contribuer globalement à l'amélioration de la garantie d'intelligibilité lors de la communication de connaissance formelle. Nos travaux considèrent différents langages formels, différents types de propriétés et différentes applications. Cette approche globale fait notre originalité et nous pensons qu'elle permet de disposer de résultats généraux et donc transposables d'un domaine vers un autre. Ils seront susceptibles de former la base d'environnements logiciels plus assis.

En tant que projet «logiciel», nous avons une cible plus restreinte qui consiste à considérer les propriétés sémantiques des flux de transformations exprimés dans TRANSMORPHER. Notre objectif à quatre ans est de développer le logiciel décrit au §3.4. Il reste beaucoup de travail car différents langages et propriétés peuvent être considérés. Cette approche permet de restreindre le champ de nos travaux et de nous permettre d'en faciliter l'aboutissement tout en obtenant des résultats utilisables.

1.3.4 Limites

EXMO n'a pas pour ambition de contribuer à la preuve de transformations en général mais à établir la preuve de propriétés particulières de transformations composés de transformations élémentaires. Il ne se préoccupe pas de la preuve des transformations élémentaires à l'exception de quelques domaines très particuliers (voir §3.2.2 et §3.3.2).

⁶ Hans Chalupsky, OntoMorph: a translation system for symbolic knowledge, Proceedings of 7th international conference on knowledge representation and reasoning (KR), Breckenridge, (CO US), pp471-482, 2000

⁷ Denise Draper, Peter Frankhauser, Mary Fernandez, Ashok Malhotra, Kristofer Rose, Michael Rys, Jérôme Siméon, Philip Wadler (éds.), Xquery 1.0 and Xpath 2.0 formal semantics, Working draft, W3C, 2002 <http://www.w3.org/TR/query-semantics/>

⁸ Malcolm Wallace, Colin Runcinan, Haskell and XML: generic combinators or type-based translation? *SIGPLAN notices* 34(9):148-159, 1999

⁹ Philip Wadler, A formal semantics of patterns in XSLT, *Markup technologies*, 1999 <http://www.cs.bell-labs.com/who/wadler/papers/xsl-semantics/xsl-semantics.pdf>

¹⁰ Haruo Hosoya, Benjamin Pierce, Regular expression pattern matching for XML, Proc. 25th PoPL, pp67-80, 2001

Ces limites d'EXMO pourraient être comblées en utilisant les acquis d'autres projets INRIA, comme PROTHÉO pour la preuve de systèmes de réécriture, LOGICAL et LEMME pour les environnements de preuve. Par ailleurs, le travail sur les transformations et leur typage devrait être abordé plus directement par le projet WAM. Les travaux du projet ACACIA sur la traduction de RDF vers les graphes conceptuels pourrait être à la fois outil et champ d'application pour EXMO.

2. PERSPECTIVES APPLICATIVES

Deux contextes applicatifs motivent et suscitent nos travaux. Le contexte dans lequel ils ont été initialement formulés est celui de l'ingénierie des systèmes de transformations. Nous avons proposé ce thème pour la création de l'action EXMO et il semble qu'il rencontre l'écho attendu auprès des communautés liées au web et au document. Ce thème a été rejoint par une thématique liée au «Web sémantique» dont le but est d'étendre la problématique des systèmes de transformations au réseau et de l'appliquer aux représentations manipulées dans le «Web sémantique».

2.1 Communication, interopérabilité syntaxique et transformations

De plus en plus, l'informatisation et la mise en réseau des organisations les conduisent à échanger de l'information sous forme électronique. Par exemple, le commerce électronique engendre un échange permanent de documents structurés, les dates de réunions s'échangent dans toutes sortes de formats, les données de conception d'un artefact sont transmises aux sous-traitants et aux unités de production.

Cependant, l'intégration des données est loin d'être parfaite. Elle est freinée par de nombreux facteurs comme l'hétérogénéité des outils de manipulation de l'information, les formats de données inaccessibles ou la confidentialité de certains éléments d'information. Ainsi, les documents traités par un tableur ne sont pas disponibles dans les agendas de la force de vente, l'unité chargée d'imprimer le matériel promotionnel ne peut lire les données de CAO et les statisticiens ne peuvent avoir accès à certains tableaux car ils contiennent des éléments confidentiels.

Certaines briques technologiques ont été récemment développées afin de résoudre ces problèmes. C'est notamment le cas du langage XML destiné à servir de base à l'encodage de nombreux langages. Le langage XSLT permet pour sa part d'exprimer des transformations de documents XML. Ces éléments sont de natures à résoudre beaucoup des problèmes cités plus haut. Ainsi, XML permettra d'appréhender plus facilement la syntaxe des données et des transformations utilisant XSLT permettront d'adapter ces données à un contexte cible. Les fichiers de tableur et de CAO exportés en XML pourront être transformés dans un format admissible par les agendas ou les outils de mise en pages, les données confidentielles pourront être expurgées des fichiers clients avant d'être traitées par l'unité de statistique.

Cette présentation constitue une instance du cadre plus général auquel EXMO entend contribuer. Elle décrit bien la transformation de données structurées lors de sa communication entre personnes ou entre applications.

Il subsiste cependant des problèmes qui devront trouver une solution adaptée dans les années à venir. Nous identifions et discutons ci-dessous trois problèmes particuliers.

- l'absence de prise en compte de manière globale de ces transformations (§2.2);
- la nécessaire prise en compte des propriétés des transformations et, en particulier, de leurs propriétés sémantiques (§2.3);
- la conception des systèmes de transformations à partir de ressources ouvertes (§2.4).

2.2 Ingénierie des systèmes de transformations

Il est peu raisonnable d'imaginer développer une transformation en XSLT à chaque fois que l'on désire communiquer de l'information entre deux nouvelles cibles. D'autre part, il est peu imaginable de considérer chaque transformation indépendamment des autres. Il semble

inévitable que, dans l'avenir, il faudra gérer des systèmes de transformations complexes. Les transformations doivent donc être appréhendées sous forme de systèmes de transformations intégrant chaque transformation individuelle dans un tout cohérent.

On a besoin de pouvoir assembler facilement des transformations — par exemple, « donnez-moi la liste des publications d'Exmo depuis 2000, dont on a supprimé celles qui ne sont pas encore diffusables, classées par catégories au format HTML ». Pour cela, à partir d'une source de données bibliographique, on aimerait pouvoir assembler des transformations élémentaires (sélection, filtrage, classement et formatage) tout en s'assurant que l'assemblage satisfait certaines propriétés (par exemple, puisque les informations ont été triées, on aimerait que le formatage ne change pas l'ordre).

La figure 1 présente une organisation dans laquelle coopèrent des agents (humains et logiciels). Les agents sont à la fois producteurs et consommateurs de connaissance représentée informatiquement (l'organisation elle-même peut-être considérée comme un agent). Pour être plus concret, on peut envisager que les individus (a) et (b) collaborent à la conception (Conception) d'un document sur un produit et une source extérieure (un sous-traitant) y apporte les informations concernant les caractéristiques sous-traitées. Cette information, suivant les canaux commerciaux peut être traduite (Traduction) en diverses langues, complétée par l'information des prix de ventes pratiqués (Pricing) et formatée (Appauvrissement) pour diverses utilisations, y compris des utilisations informatiques extérieures à l'entreprise.

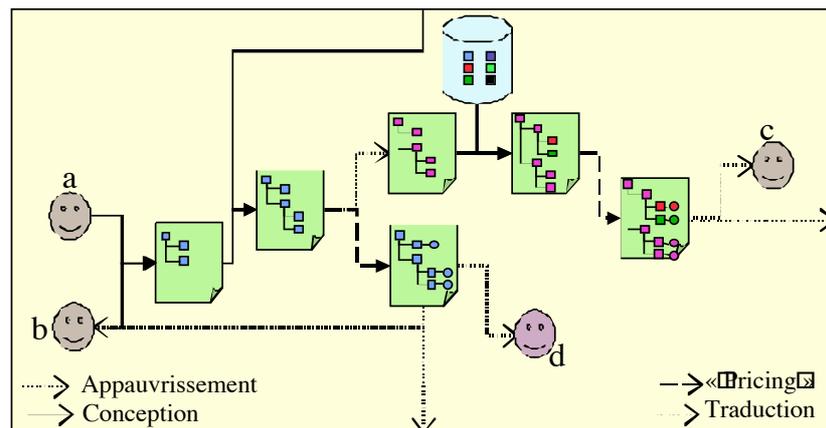


Figure 1 — Transformation de la connaissance structurée dans une organisation.

Ce type de schéma n'a rien de futuriste — il s'incarne déjà dans la gestion de connaissance, la gestion de ressources intégrée, la gestion électronique de documents (ou la maquette numérique) et l'ingénierie concurrente. Il n'est cependant pas explicité dans un schéma permettant à un processeur d'exécuter les transformations, de calculer les dépendances entre sources d'informations ou de garantir qu'aucune information n'a été perdue.

Un système de transformations organise explicitement l'ensemble de ces transformations de flux de données dans un schéma unique permettant de réaliser les tâches de changement de formats à l'aide de transformations standard, souvent réutilisées, les tâches de filtrage à l'aide de transformations simples dont on s'assurera qu'elles ne laissent pas passer ce qui doit être filtré. L'ensemble de ces transformations s'exécute de manière coordonnée afin de s'assurer que l'information fournie est toujours à jour.

Les systèmes de transformations s'appliquent donc à l'échelle du fonctionnement d'un système d'information et non plus à l'échelle de chaque individu. Il y a, à notre avis, une ingénierie à développer autour de la conception de tels systèmes. Nous appelons ingénierie des systèmes de transformations l'activité qui s'attache à développer des transformations, à

les assembler, à définir et à vérifier les propriétés qui doivent être satisfaites par ces transformations.

L'ingénierie de ces systèmes de transformations nécessitera des outils, des méthodologies et des méthodes formelles. En effet, il sera nécessaire de vérifier qu'un système particulier, et non plus une transformation individuelle, ne laisse pas échapper plus d'information qu'autorisé ou que le flux dans sa globalité termine. Pour cela, il faudra disposer d'une description analysable de ce flux de transformations et de caractérisations formelles des propriétés escomptées. EXMO est concerné par les outils et les méthodes formelles et cherche à associer les deux en proposant des solutions pour la création de véritables ateliers de conception de systèmes de transformations (voir §4.1).

2.3 Propriétés, interopérabilité sémantique et composition

L'un des premiers aspects à traiter par le concepteur des systèmes de transformations est la spécification de leur comportement. Cette spécification peut être fondée sur la notion de propriété à satisfaire par une transformation. Les propriétés peuvent être très nombreuses et peuvent concerner des aspects très syntaxiques comme des aspects sémantiques. Des exemples de propriétés sont□

- la terminaison□ la transformation s'arrêtera et donnera un résultat quelle que soit son entrée□
- la non-divulgaration de certaines informations□ les noms de personnes passant commande n'apparaissent plus dans la sortie□
- la préservation des conséquences□ ce qui est vrai dans tous les modèles de l'entrée l'est dans ceux de la sortie.

Dans l'exemple de l'intégration de planifications, on exprimera la propriété que chacune de ces transformations individuellement termine, n'ajoute pas d'information qui n'était présente au départ, préserve les dates et participants de chacun des événements. Il est nécessaire de donner à l'architecte des systèmes de transformations la possibilité d'exprimer ces propriétés et de les attacher aux transformations manipulées. Ainsi, il pourra à tout moment connaître les capacités des transformations qu'il utilise.

Les propriétés sur lesquelles notre attention se porte sont les propriétés sémantiques, c'est-à-dire celles qui concernent les modèles et les conséquences d'une représentation. En effet, il est très pratique d'utiliser des langages différents pour exprimer la connaissance, mais on peut aussi développer des modélisations différentes (par exemple, la représentation de l'espace fondée sur des points ou fondées sur des volumes). Pour passer d'un monde à l'autre il est nécessaire, au-delà des différentes formes syntaxiques que peuvent prendre les représentations, de s'assurer que la signification reste la même en dépit des opérations. L'importance de la sémantique provient du fait qu'elle définit les résultats des opérations qu'un calculateur peut appliquer aux représentations. Elle définit, par exemple, les réponses valides aux requêtes en fonction des modèles des expressions. Nous appliquons l'approche sémantique autant aux représentations de connaissances formelles qu'à l'ordonnement de documents multimédia (voir §3.2.2).

Il sera aussi indispensable d'inventorier le comportement de ces propriétés vis-à-vis des opérations permettant d'assembler les transformations. Par exemple, la non-divulgaration et la préservation des conséquences sont préservées par l'itération mais pas nécessairement la terminaison. Une telle étude permettra de calculer les propriétés des flux de transformations conçus par assemblage des transformations élémentaires de la bibliothèque.

Ensuite, s'il est nécessaire de modifier ce flux de transformations pour y intégrer les informations provenant du calendrier administratif, il faudra vérifier que les propriétés sont toujours préservées par le nouveau flux, modifier certaines propriétés (de l'information

nouvelle est apportée), en ajouter de nouvelles (garantir que deux évènements n'ont pas lieu simultanément).

2.4 Technologies pour un «Web sémantique»

Nous avons développé ci-dessus l'aspect local de l'ingénierie des systèmes de transformations. Nous pensons qu'il est nécessaire de la considérer dans un contexte global. Pour cela nous cherchons à développer une infrastructure permettant de prolonger les solutions évoquées plus haut en tirant parti du web.

Au lieu d'utiliser des transformations disponibles localement dans son unité, l'ingénieur des systèmes de transformations aura la possibilité d'intégrer des transformations disponibles sur le réseau mondial. Il est vraisemblable que toutes les transformations de changement de format (par exemple, du format de calendrier vCal vers HTML) ou d'export pour des dispositifs particuliers (WAP) devraient y être disponibles. Ainsi, il est vraisemblable que les transformations d'un format vers un autre pourront être importés telles quelles, d'une entreprise vers une autre. Un atelier permettant cela améliorera grandement la productivité de ces ingénieurs et, espérons le, la qualité des transformations utilisées. Cependant, il devra faire face au problème du contrôle de cette qualité.

Les problèmes posés ci-dessus se posent alors sans doute de manière encore plus aiguë. On ne fait plus forcément confiance aux transformations disponibles, on va vouloir être certains qu'elles satisfont bien les propriétés désirées. Pour cela, on voudra disposer de la preuve de la propriété à satisfaire. Comme il n'est pas imaginable de développer une preuve à la main à chaque fois que l'on désire intégrer une transformation, on préférera n'avoir qu'à vérifier des preuves préexistantes. C'est la technique du code porteur de preuve — où des preuves des propriétés satisfaites sont associées aux programmes et vérifiées avant que ceux-ci soient exécutés — appliquée aux transformations. Dans le contexte du web, les transformations, leurs propriétés et les preuves de ces propriétés peuvent être mise à disposition sur le réseau.

Dans le cadre du «Web sémantique», il apparaît inévitable d'avoir à intégrer des éléments de connaissance provenant de sources diverses, exprimés dans des langages différents et associés à des axiomatiques différentes. Ainsi, il sera nécessaire de transformer les éléments avant de les assembler pour un usage particulier. On désire proposer au-dessus des outils d'ingénierie des systèmes de transformations des outils capables de prendre en compte les propriétés que les transformations doivent satisfaire.

L'ingénieur du système de transformations peut rapatrier les transformations, les propriétés associées et leurs preuves et vérifier les preuves avant d'intégrer les transformations dans son environnement de développement de transformations. Il va alors pouvoir assembler diverses transformations afin de créer le flux de transformations dont il a besoin. Il pourra chercher à prouver les propriétés qui sont requises. Ensuite, il pourra à son tour publier sa transformation et sa preuve. Ceci constitue un cycle de développement de transformations sûres.

3. AXES DE RECHERCHE

Cinq axes de recherche sont présentés ci-après. Le premier comprend la représentation de connaissance au sens large et sa manipulation sous une forme communicable. C'est ce qui constitue la suite et la base de notre travail (§3.1). Le second étudie l'applicabilité de notre approche à des types de propriétés et de transformations les plus variées possible (§3.2). Le troisième se concentre sur ce qui devrait permettre l'intelligibilité entre ordinateur et les propriétés sémantiques (§3.3). Nous proposons ensuite une infrastructure technologique applicable à la fois au «Web sémantique» et à l'ingénierie des transformations (§3.4). Elle permet d'appliquer les concepts dégagés dans les axes précédents. Enfin, nous évoquons l'extension de nos travaux au domaine sémiologique qui devrait permettre de réintégrer l'humain dans la chaîne d'interopérabilité (§3.5).

Cinq axes peuvent sembler beaucoup. Il faut cependant noter que le premier est une tâche de fond demandant peu d'activité alors que le dernier est pour l'instant très prospectif. Le cœur de notre activité est donc lié aux trois axes centraux. Nous avons préféré présenter un panorama complet.

3.1 Représentation du contenu

Notre origine étant la représentation de connaissance, nous maintenons un intérêt pour les problèmes posés par la représentation de connaissance, en particulier dans le cadre de la représentation du contenu d'une base documentaire. Nous avons développé des comparaisons de différents systèmes de représentation de connaissance (graphes conceptuels, représentations par objets, logiques de description) dans la représentation du contenu de documents pour l'indexation (en liaison avec les équipes ACACIA et ORPAILLEUR). Nous avons, par ailleurs, décrit plusieurs langages de représentation de connaissance en XML (TROEPS, le langage pivot d'ESCRIRE et diverses logiques de descriptions, voir plus bas) [Euzenat2000a, 2001c, 2002b]. On prolongera notre réflexion sur l'échange de connaissance formalisée et, en particulier, sa communication à travers le Web et sa contribution au Web.

Le prolongement de ces travaux est maintenant lié à la représentation de la sémantique des langages de représentation que l'on retrouvera au §3.3.

Cette activité nous permet d'intervenir lors du développement d'OWL (un langage dérivé des logiques de description qui devrait s'imposer dans le «Web sémantique») et de suivre les travaux de Raphaël Troncy doctorant à l'INA sur la liaison entre représentation de connaissance et description de documents audiovisuels (voir §5.3.1).

Objectif Se maintenir à niveau, renforcer notre expérience, exploiter notre savoir faire.

3.2 Propriétés des transformations

La conception de systèmes informatiques en tant que systèmes de transformations impose non seulement de savoir exprimer de tels systèmes mais aussi d'établir leurs propriétés. Nous nous intéressons plus particulièrement aux propriétés de préservation. Elles peuvent permettre de garantir la préservation d'une relation d'ordre entre les représentations d'entrée (r) et de sortie ($\square(r)$) comme son anti-préservation. Pour cela on peut identifier des propriétés

- Syntaxiques comme la complétion ($\square(r) \leq r$, où \leq dénote la subsomption structurelle entre représentations)

- Sémantiques \square comme la préservation des conséquences ($\square(r) \square r$, ou l'équation 2 présentée au §3.3) \square
- Sémiologiques \square comme la préservation des interprétations (soit \square les règles d'interprétation des expressions et \models_i , l'interprétation de l'individu i , $\square \square L$, $\square i, j$, $r, \square \models_i \square \square \square(r), \square(\square) \models_j \square(\square)$).

Notre but est de bâtir une «Théorie générale de ces transformations» fondée non pas sur les représentations ni sur les transformations en tant que telles, mais sur les propriétés qu'elles doivent satisfaire. On s'occupe donc de propriétés diversifiées (préservation du contenu ou de la structure, identifiabilité de la source — traçabilité —, confidentialité) et l'on s'intéresse plus aux transformations qu'aux langages de représentation. On présente ci-dessous les applications de ce cadre à des catégories particulières de propriétés (informationnelles, computationnelles...). Les propriétés purement sémantiques sont l'objet d'une section à part (§3.3).

Notre but ici est de montrer que l'approche par transformations et propriétés s'applique à des contextes variés et que les travaux que nous développons contribuent à la résolution des problèmes posés.

Objectif \square Caractériser, pour un type de propriété particulier, quelles transformations les laissent invariantes et quelle est l'action de la composition de transformations.

3.2.1 Treillis et information

Cette étude est plus particulièrement le travail de doctorat d'Olivier Brunet [Brunet2002b]. Elle a pour point de départ d'autres travaux sur la diffusion de l'information comme la théorie des flux d'information¹¹ ou l'utilisation des treillis comme l'analyse de concepts¹².

L'analyse des flux d'information est centrée autour d'une formalisation à base de demi-treillis (un ensemble E muni d'un ordre partiel \leq et dans lequel tout couple d'éléments a un plus petit majorant pour l'ordre). Les treillis permettent d'avoir une notion d'imprécision dans la description de l'information. L'information disponible en un point sur un objet est un élément de ce treillis. L'ordre permet de préciser qu'un élément du treillis contient plus d'information qu'un autre. On peut ainsi définir l'absence totale d'information (l'élément maximal de l'ordre) et un élément plus grand qu'un autre est une approximation de celui-ci (car il contient moins d'informations, mais ne le contredit pas). Ainsi, considérant deux systèmes échangeant de l'information, on peut écrire de façon rigoureuse qu'un système connaît précisément telle information d'un autre système, ou qu'il ne connaît cette information que partiellement, ou encore pas du tout.

Si l'on envisage les langages que nous utilisons comme des treillis $(\square L, \leq \square)$, une transformation \square d'un tel treillis dans un autre permet de modéliser la modification subie par l'information lors de sa circulation (la transformation) d'un espace de représentation à un autre. Une transformation perdra de l'information si sa transformation adjointe (définie par le plus petit majorant des éléments qui ont pour image un élément de L') peut retourner un élément plus grand que l'élément de départ (elle n'en retournera jamais de plus petit).

On peut définir le cas particulier où les transformations sont des applications dans le même langage $(\square L \square L)$, comme le domaine d'interprétation d'une logique modale. Dans ce cas, l'interprétation d'une formule \square est l'ensemble des représentations (les nœuds du treillis) dans

¹¹ Jon Barwise, Jerry Seligman, Information flow: the logic of distributed systems, Cambridge university press, Cambridge (UK), 1997

¹² Bernhard Ganter, Rudolf Wille, Formal concept analysis, Springer-Verlag, Berlin (DE), 1999

lesquels elle est vraie et l'interprétation de la modalité K_i (connu de i , liée à une transformation \square) s'interprète comme l'ensemble des représentations qui, une fois transformées par \square , satisfont la formule à connaître ($[K_i] = \{x \mid L \square (x) \square [\square]\}$). Comme l'information croît de manière monotone en descendant le treillis, l'interprétation d'une formule est un ensemble clos pour \leq .

L'axiomatisation de cette structure révèle qu'il s'agit de la logique modale IS4 (qui ne satisfait ni le tiers exclus — $\square \square \neg$ —, ni l'axiome 5 — $\square K_i \square \square K_i \square K_i$ — du fait de la spécificité de la négation sur des sous-ensembles clos par \leq) [Brunet2001a].

L'intérêt de ce travail est de permettre de poser des questions liées à la connaissance après transformation (par exemple, $\square \square \square K_{ij} \square \square \square K_i \square \square \square K_j$, est-il possible en rassemblant les connaissances de i et j de déduire des informations dont aucun des deux ne dispose, avec $\square \square \square (K_i \square \square K_{ij}) \square (K_j \square \square K_{ij})$). Ces techniques s'appliquent donc si l'on veut s'assurer que certaines informations ne sont pas dérivables de celles que l'on laisse filtrer.

Il faudra raffiner le type de transformations utilisées (en particulier, les termes peuvent être des documents XML) pour pouvoir appliquer ces travaux sur l'information aux systèmes de transformations.

3.2.2 Préservation de scenarii dans les documents multimédia

Lorsqu'un document multimédia doit être exécuté sur des plates-formes dotées de possibilités diverses (un téléphone ne pouvant afficher qu'une image à la fois, une borne interactive sans clavier...), il est nécessaire de le transformer afin de l'adapter au dispositif cible. Nous nous intéressons à décrire les propriétés des documents qui doivent être préservées par cette adaptation. Pour cela nous utilisons l'approche sémantique qui consiste à considérer un modèle d'un document multimédia comme l'une de ses exécutions possibles (une exécution satisfaisant sa spécification) et en définissant des classes de modèles.

Par exemple, la structure temporelle des scenarii multimédia est exprimée en Smil 2.0 à l'aide d'opérateurs qui peuvent se transcrire dans les relations temporelles qualitatives¹³ entre les éléments apparaissant dans la présentation. Les modèles peuvent être rassemblés en classes en fonction de propriétés qu'ils satisfont ensemble (par exemple, modulo que leur description qualitative saturée, c'est-à-dire dans laquelle il n'existe qu'une seule relation qualitative entre chaque couple d'éléments, soit la même).

L'adaptation va, en première approximation, réduire l'ensemble des modèles du scenario (puisque'elle doit supprimer les exécutions impossibles). Mais l'approche sémantique permet d'exiger, par exemple qu'au moins une instance de chaque classe des modèles de la spécification initiale soit un modèle de la spécification transformée. L'un des premiers intérêts de ce cadre a été de révéler que l'approche d'adaptation utilisée généralement se bornait à distordre le document initial sans que cela soit forcément nécessaire (il peut y avoir des modèles du document qui satisfassent les contraintes d'adaptation).

Nous comptons nous intéresser au type de structure de navigation dit «non déterministe» (par exemple, un bouton qui, lorsqu'il est pressé, interrompt totalement la présentation en cours et passe à une autre section). Nous nous intéressons à la classe de modèles correspondant à un parcours de cette structure et à sa préservation (scenarii). En effet, si la structure navigationnelle est transformée, il est possible de se trouver avec des éléments de documents non atteignables ou des ordres de présentations non exécutables (ce qui peut être très important dans le cadre d'un jeu par exemple).

¹³ James Allen, Maintaining knowledge about temporal intervals, *Communications of the ACM* 26(11):832-843, 1983

Cependant, lorsqu'il n'existe pas de modèles satisfaisant les contraintes d'adaptation, il est nécessaire que la transformation propose une exécution proche d'une des exécutions possibles. Pour cela nous employons encore l'approche sémantique en cherchant à définir une distance entre les modèles.

Nous pouvons être aidé en cela par les transformations élémentaires applicables ou la structure de la représentation. Nous avons déjà exploré la préservation de la structure temporelle des documents multimédia en tirant parti de la notion de "voisinage conceptuel" dans l'ensemble des relations qualitatives. Mais, nous avons montré que l'évaluation des transformations en fonction de la préservation du maximum de contraintes temporelles n'était pas forcément la meilleure dans tous les cas. Il est donc nécessaire de décrire les scénarii temporels avec des primitives de plus haut niveau (comme les structures rhétoriques de la "rhetorical structure theory"¹⁴).

Ce travail, commencé par le DEA de Victor Dias co-encadré avec Nabil Layaïda, ouvre de très vastes perspectives. Le principal groupe travaillant sur ce type de sujet est le CWI (Amsterdam) qui utilise des représentations rhétoriques pour les parcours thématiques dans les documents multimédia¹⁵.

3.3 Interopérabilité sémantique

Dans le cadre de la communication de représentations formalisées, on va chercher à garantir la préservation du sens entre deux représentations. Ceci permet l'intelligibilité d'une représentation dans le contexte cible et donc l'interopérabilité entre les deux contextes. Plus précisément, entre deux langages L et L' , une représentation r de L va être transformée en une représentation $\alpha(r)$ dans L' (où $\alpha: L \rightarrow L'$). On adopte donc une approche consistant à étudier l'intelligibilité sémantique via une transformation.

Assurer l'intelligibilité sémantique peut être décrit à l'aide des deux équations complémentaires

$$\forall \alpha: L \rightarrow L', r \models_L \alpha \iff \alpha(r) \models_{L'} \alpha \quad (1)$$

$$\forall \alpha: L \rightarrow L', \alpha(r) \models_{L'} \alpha \iff r \models_L \alpha \quad (2)$$

La garantir dans toute sa généralité est, bien entendu, hors de portée. Par conséquent, nous nous intéressons à étudier des cas particuliers de ces équations [Euzenat2001b].

Objectif Caractériser, pour un certain ensemble de techniques, comment celles-ci peuvent permettre de garantir l'intelligibilité (chacune des deux équations ci-dessus). Dresser les conditions dans lesquelles les techniques proposées fonctionnent.

3.3.1 Hiérarchie de propriétés sémantiques

Notre premier travail a consisté à inventorier l'ensemble des manifestations de l'équation (1) que l'on peut utiliser dans les transformations de langages à langages. Cela permet d'envisager différents cas de figure

- L' contient tous les constructeurs de L , dans ce cas, la traduction α est l'identité

¹⁴ William Mann, Sandra Thompson, Rhetorical structure theory toward a functional theory of text organization, *Text* 8(3):243-281, 1988

¹⁵ Lloyd Rutledge, Jim Davis, Jacco van Ossenbruggen, Lynda Hardman, Inter-dimensionnal hypermedia communicative devices for rhetorical structure, Proceedings of the Multimedia Modeling conference, Nagano, (JP), 2000, pp89-105, <http://www.cwi.nl/~media/publications/>

- L contient strictement plus de constructeurs que L' (il est plus expressif). Dans ce cas, la seule solution n'est pas comme on pourrait penser l'appauvrissement syntaxique. Il faut, en toute théorie, mettre en œuvre un démonstrateur pour déduire de r tout ce qui est exprimable en L' .
- Il existe une transformation de L vers L' préservant les interprétations des assertions. Nous avons exploré le cas où l'on disposait de preuves d'équivalence entre L et L' d'un type assez particulier□ pour chaque constructeur d'un langage, on montre que toute expression l'utilisant est exprimable avec les constructeurs de l'autre. On a alors la définition constructive d'une transformation préservant les interprétations et que l'on peut tenter de développer en XSLT.
- Il existe une transformation de L vers L' préservant les modèles. Ce cas correspond à la définition de l'expressivité proposée par Franz Baader¹⁶ appliquée à notre cadre (on a une transformation et non une fonction de renommage).
- Il existe une transformation de L vers L' telle que tout modèle de la représentation image inclue un modèle de la source modulo la transformation. Ceci généralise la propriété précédente et permet d'autoriser des représentations avec un nombre de modèles plus importants tout en préservant les conséquences.

Ces propriétés se retrouvent dans des contextes divers (preuve d'équivalence entre logiques de descriptions, "interprétation" en logique du premier ordre) et ont bien souvent été étudiées de manière séparée. Les travaux de Claudio Masolo¹⁷, Ciocoiu et Nau¹⁸ ou Franz Baader¹⁶ peuvent y être associés.

Inventorier ces propriétés et les relations d'implications qu'elles entretiennent permet de déterminer la propriété satisfaite par une succession de transformations utilisées pour passer d'un langage à un autre.

3.3.2 Propositions sémantiques des transformations sur les graphes étiquetés

Au lieu de se focaliser sur les propriétés sémantiques, en général, on peut fixer plus précisément le langage et étudier les propriétés satisfaites par des transformations naturelles sur les constructions de celui-ci. On considère ici les langages à base de graphes encodant des connaissances exprimées dans différents formalismes de représentation (dont RDF, au même titre que les graphes conceptuels, peut être considéré comme une instance). Nous nous interrogerons sur la façon dont des transformations élémentaires sur ces graphes préservent la sémantique au sens proposé des équations (1) et (2).

Au cours de sa thèse, Jean-François Baget¹⁹ a proposé plusieurs réductions de problèmes de déduction dans divers formalismes de représentations de connaissances utilisant des graphes (plus précisément, des graphes conceptuels). Ces réductions reposent sur plusieurs transformations élémentaires telles que□

- la transformation d'une étiquette en un sommet (utilisée pour réifier les types);
- la transformation d'un arc en un sommet (utilisée pour réifier les relations);

¹⁶ Franz Baader, A formal definition of the expressive power of terminological knowledge representation languages, *Journal of logic and computation* 6(1):33-54, 1996

¹⁷ Claudio Masolo, Criteri di confronto e costruzione di teorie assiomatiche per la rappresentazione della conoscenza: ontologie dello spazio e del tempo, Tesi di dottorato, Università di Padova, Padova (IT), 2000

¹⁸ Mihai Ciocoiu and Dana Nau, Ontology-based semantics, Proceedings of 7th international conference on knowledge representation and reasoning (KR), Breckenridge, (CO US), pp539-546, 2000 <http://www.cs.umd.edu/~nau/papers/KR-2000.pdf>

¹⁹ Jean-François Baget, Représenter des connaissances et raisonner avec des hypergraphes□de la projection à la dérivation, Thèse d'informatique, Université de Montpellier II, Montpellier (FR), 2001 http://www.inrialpes.fr/exmo/people/baget/publications/baget_01_these.pdf

- la transformation d'un contexte (utilisés dans les graphes emboîtés) en un sommet et un ensemble de relations;

Ces réductions de problèmes (de complexité polynomiale, et souvent linéaire) ne s'attachent pas seulement à la conservation des réponses positives, mais conservent l'ensemble des modèles des graphes transformés (en fait, des classes d'équivalence pour ces modèles). Bien que ressemblant à des réductions de problème, au sens algorithmique du terme, elles peuvent en différer de manière importante. En effet, la polynomialité de la transformation, si elle est souhaitable, n'est pas obligatoire, nos transformations peuvent être incomplètes et la réponse à un problème de décision est insuffisante□ nous nous intéressons à la façon dont sont conservés les modèles.

Ces travaux serviront de base à l'étude d'une bibliothèque de transformations de graphes (sémantiquement définis), qui devraient alors être analysées en tenant compte des propriétés suivantes□

- Caractéristiques du domaine: langage de représentation de connaissance utilisé dans le document r que l'on souhaite transformer, et propriétés structurelles (du graphe) requises pour pouvoir appliquer la transformation □
- Caractéristiques du co-domaine: langage de représentation de connaissance utilisé dans le document cible $\square(r)$, et propriétés structurelles du graphe obtenu.
- Transformation des données: étude de complexité, et développement de l'algorithme permettant de transformer r en $\square(r)$.
- Transformation des inférences: y-a-t-il une fonction \square qui, à partir des modèles de $\square(r)$, permette de retrouver exactement/tous/une partie des modèles de r ? Il faudra de même prendre en compte la complexité de cette fonction.

3.3.3 Structures et interopérabilité sémantique

S'il n'est pas immédiat de produire les propriétés sémantiques cités dans les sections précédentes, il est possible d'en tirer parti dans des structures particulières liées aux langages ou aux théories utilisées. Nous avons pour l'instant considéré deux types de structures qui, appliquées aux représentations manipulées, permettent de faciliter l'établissement des propriétés.

Familles de langages. Dans le cas d'une famille de langages, il existe un langage L L' tel que toute formule de L ou de L' soit une formule de L L' . De plus, les opérateurs sont interprétés de la même façon dans L et dans L L' (et par conséquent, les opérateurs présents dans les deux langages ont la même interprétation). Un cas de famille de langages bien connu est celui des logiques de descriptions où toute une hiérarchie de langages a été définie²⁰. Nous l'avons exploitée dans la construction du langage DLML (voir §4.2). Nous avons expérimenté avec elle l'établissement de propriétés sémantiques en utilisant des transformations entre deux langages dont on connaît les propriétés sémantiques [Stuckenschmidt2001a]. Affecter les propriétés aux transformations permet de rechercher un chemin (matérialisé par la composition d'une séquence de transformations) dans les langages de la famille de telle sorte de préserver une propriété particulière.

Transformations fondées sur les patrons. Dans le cas général, il est aussi possible que les deux langages disposent de structures communes. C'est en particulier vrai lorsque ces langages sont issus d'une souche commune. Une tendance actuelle est d'adapter la notion de patron ("pattern") utilisée en conception logicielle pour l'appliquer aux langages de

²⁰ Franz Baader, Deborah McGuinness, Daniele Nardi, Peter Patel-Schneider (eds.), The description logic handbook, Cambridge university press, Cambridge (UK), 2002 to appear

construction d'ontologie. L'instanciation d'un patron p dans un langage L consiste à fournir un morphisme $\llbracket p \rrbracket : L \rightarrow \mathcal{O}$. L'un des buts de ces patrons est de représenter des notions communes et ainsi faciliter la traduction d'un langage vers un autre. Formellement, si le morphisme $\llbracket p \rrbracket$ est réversible (c'est-à-dire que les éléments utilisés dans L pour manifester la notion utilisée dans p ne peuvent être utilisés pour autre chose), il est alors possible de construire la transformation $\llbracket p \rrbracket^{-1} : \mathcal{O} \rightarrow L$. Cette transformation va permettre de traduire fidèlement la notion décrite par p .

Pour pouvoir obtenir des résultats tangibles sur la notion de patron, il faut décliner plus avant la description des patrons et leurs instanciations possibles. En effet, les langages concrets correspondent à l'instanciation de plusieurs patrons à laquelle sont ajoutées des constructions particulières. Les types de langages pertinents seront dépendants des propriétés à satisfaire.

Ces techniques sont des techniques de génie logiciel appliquées à la construction de représentations formalisées. Elles s'appliquent assez naturellement au partage de connaissance. La construction par famille de langages a été proposée avec Heiner Stuckenschmidt de TZI-Brème. L'utilisation de patrons a reçu sa proposition la plus aboutie à Karlsruhe²¹, mais a attiré l'attention d'autres auteurs.

3.4 Vers une infrastructure sûre de partage de transformations

L'une des applications précises des travaux présentés ci-dessus combine l'ingénierie des systèmes de transformations avec le «Web sémantique». En effet, on peut s'attendre, dans le cadre du «Web sémantique» à ce que des connaissances soient représentées de manière indépendante par différentes organisations. Certaines organisations éprouveront le besoin de combiner la connaissance provenant de sources hétérogènes. Si la connaissance est exprimée dans des langages différents avec des axiomatisations différentes, il est nécessaire d'opérer des transformations avant de l'utiliser et se pose alors la question des propriétés des transformations vis-à-vis du sens des représentations initiales.

Nous avons proposé pour cela une infrastructure [Euzenat2002d] qui repose sur les travaux précédemment décrits et qui s'appuie sur le développement logiciel de TRANSMORPHER (voir §4.1). Elle permet d'assurer ces propriétés par l'échange de transformations, de propriétés et de preuves.

Objectif Développer le système proposé et l'utiliser comme preuve de concept. L'intégrer à l'environnement de TRANSMORPHER.

3.4.1 Composition de propriétés

L'un des buts d'EXMO est d'envisager des flux de transformations qui résultent de l'assemblage de transformations plus élémentaires (qui peuvent être vues comme des fonctions, l'assemblage étant le plus souvent la composition). Si chacune des transformations plus élémentaires est dotée d'assertions spécifiant les propriétés qu'elles satisfont, il faut construire la propriété concernant l'assemblage. Un exemple simple est la propriété de terminaison sur des entrées finies qui est préservée par la composition mais pas forcément par

²¹ Steffen Staab, Michael Erdmann, Alexander Maedche. Engineering ontologies using semantic patterns. In Asuncion Gomez Perez, Michael Gruninger, Heiner Stuckenschmidt, Michael Uschold (eds), Proc. IJCAI 2001 workshop on ontology and information sharing, Seattle (WA US), pp174-185, 2001 <http://ceur-ws.org/Vol147/staab.pdf>

l'itération saturée (jusqu'à ce que la transformation ne soit plus applicable). Le sens, par contre, est préservé par la composition et l'itération.

Notre but est d'inventorier à la fois les constructeurs de transformations (composition, itération, séparation des données) et les propriétés que l'on peut associer à une transformation (classement, privilèges, pertinence, niveau de granularité, traçabilité de l'information). On développe aussi une typologie des transformations (analysables par une machine) qui permette de prouver aisément certaines propriétés (principalement syntaxiques). Alors on pourra élaborer la preuve qu'un flux de transformations satisfait une propriété par assemblage. Il est aussi intéressant d'étudier l'interférence entre propriétés de différents types (par exemple, comment préserver la traçabilité face aux exigences de classement de l'information□).

Ce travail est en cours. Il a donné lieu à la définition du langage de TRANSMORPHER (voir §4.1). Nous avons commencé à observer la composition de propriétés sémantiques décrites au §3.3.1, qui permettent de déduire la propriété satisfaite par la composition de transformations dans une famille de langages (§3.3.2). Un certain nombre de travaux peuvent être considérés comme comparables dans le contexte du typage^{9,10}.

3.4.2 Transformations porteuses de preuves

On a vu plus haut que les transformations pouvaient être bâties sur des preuves. Il est utile de pouvoir disposer de ces preuves. Ainsi, pour l'équivalence entre deux logiques de descriptions, on montre que les expressions de l'une sont exprimables en fonction des expressions des autres et que la transformation préserve les modèles (au sens de la théorie des modèles). Disposer de la description de la sémantique des langages et de la preuve d'une transformation permet d'appliquer à l'importation de connaissance exprimée dans un langage différent le principe du code porteur de preuve ("proof-carrying code"²²) et, par conséquent, d'être certain de disposer de la transcription de la représentation initiale dotée des propriétés annoncées.

Afin de parvenir à vérifier la preuve des propriétés sémantiques du type de (1) et (2), il est nécessaire de disposer (a) de la représentation de la transformation, (b) de la représentation de la sémantique des langages (ou spécification de patrons) et (c) de la représentation de la preuve. Ces trois éléments seront à développer afin de pouvoir communiquer et vérifier des preuves□

- Pour représenter les transformations, on s'appuie sur le langage de TRANSMORPHER décrit au §4.1. L'un des objectifs de ce langage est d'encapsuler le langage XSLT utilisé dans les transformations de manière à présenter un langage plus aisément analysable et à décomposer hiérarchiquement les transformations de telle sorte que l'on puisse réaliser des preuves à partir de « Π mmes□ qui seront des contraintes pour les sous-transformations.
- DSD (Document Semantics Description) est un langage en cours de développement de description de la sémantique (en théorie des modèles) fondée sur XSLT, Xpath et MathML. Il sera expérimenté sur plusieurs formalismes de représentation de connaissance et par des transformations entre formalismes.
- La représentation des preuves (de telle sorte que l'on puisse vérifier facilement des preuves par induction sur la structure de termes dans un premier temps) reste à

²² George Necula, Compiling with proofs, PhD thesis, Carnegie Mellon university, Pittsburgh (PA US), 1998

développer. Une telle représentation devra tirer parti des travaux sur MathML et OpenMath mais aussi d'initiatives plus adaptées telles que OMDoc²³.

Actuellement, nous avons fait des expériences avec le langage de description de la sémantique. Celui-ci a été utilisé dans la description de DLML (où nous proposons une DSD de chaque logique construite automatiquement à partir de la sémantique des constructeurs de termes). Nous disposons de quelques exemples de preuves (principalement d'équivalence) dans les logiques de descriptions qui seraient un très bon premier exemple de l'application de ces concepts. Nous avons aussi des exemples de transformations entre représentations hétérogènes □ syllogistique et logiques de descriptions.

Dans le contexte du web, les transformations, leurs propriétés et les preuves de ces propriétés peuvent être mises à disposition sur le réseau. Il faut noter que données, transformations et preuves n'ont pas à être issues de la même source. Le langage RDF permet d'annoter une transformation par une propriété et une propriété par une preuve. L'ingénieur du système de transformations peut rapatrier les transformations et leurs preuves et vérifier les preuves avant d'intégrer les transformations dans son environnement de développement de transformations. Il va alors pouvoir assembler diverses transformations afin de créer le flux de transformations dont il a besoin. Il pourra chercher à prouver les propriétés qui sont requises. Ensuite, il pourra à son tour publier sa transformation et sa preuve. Ceci constitue un cycle de développement de transformations sûres.

3.5 Sémiologie computationnelle

Même en ce qui concerne la connaissance formalisée, les propriétés sémantiques ne sont pas suffisantes pour garantir l'interopérabilité. Nous avons mis en évidence la difficulté, même pour les individus qui connaissent parfaitement la sémantique des langages utilisés, d'appréhender correctement les contenus. Certains collègues de Manchester, étendant au web leur éditeur pour l'un des langages du «[Web sémantique](#)», ont mis en évidence indépendamment ce problème²⁴.

L'activité «[Sémiologie computationnelle](#)» fait partie des thèmes de recherche du projet EXMO. Cependant le manque de temps et de moyens ne permet pas pour l'instant de s'y consacrer correctement. C'est donc plutôt une activité de veille amont et de réflexion.

Son but est de faire prendre en compte, par la machine, la dimension sémiologique dans l'interprétation des représentations (même formalisées) par les humains. La sémantique formelle utilisée dans les systèmes de représentation de connaissance est valable pour une machine ou pour un utilisateur en présence d'un système car on peut supposer qu'il interprète de manière cohérente les termes (identificateurs) qu'il a utilisés. Mais lorsque plusieurs utilisateurs communiquent, cette bonne compréhension est difficile à obtenir.

Les travaux, menés ailleurs, sur la construction d'«[Ontologies](#)» contribuent à ce but²⁵. Mais d'autres problèmes interviennent. Par exemple, un utilisateur peut exprimer sa connaissance à l'aide de hiérarchies de classes et de clauses, puis communiquer celle-ci en la traduisant vers un langage destiné à assurer l'interopérabilité. Mais si ce langage exprime la totalité de la connaissance sous forme de clauses (tout en préservant la sémantique des assertions), l'émetteur risque de ne plus reconnaître le résultat pourtant sémantiquement équivalent. Ainsi, lorsqu'il y a traduction entre plusieurs systèmes de représentations, la

²³ Michael Kohlhase, OMDoc □ an open markup format for mathematical documents, SEKI report SR-00-02, Universität des Saarlandes, Saarebrücken (DE), 2000
<http://www.mathweb.org/src/mathweb/omdoc/doc/omdoc/omdoc.ps>

²⁴ Sean Bechhofer, Carole Goble, Ian Horrocks, DAML+OIL is not enough, in [Cruz2001a], pp151-160, 2001

²⁵ Nicola Guarino (éd.), Formal ontology in information systems, IOS press, Amsterdam (NL), 1998
<http://www.ladseb.pd.cnr.it/infor/ontology/papers/FOIS98.ps>

bonne compréhension ne peut être garantie. La raison principale est que la compréhension humaine est tributaire de la forme alors que le traitement informatique ne dépend que de la sémantique. Il faut ajouter un traitement sémiologique (ou pragmatique dirait-on en linguistique) pour faciliter la reconstruction du sens. Ce traitement complète celui purement sémantique présenté ci-dessus.

Nous avons élaboré quelques exemples convaincants de ce problème (en particulier sur les logiciels d'évaluations de communications scientifiques qui ne permettent pas au destinataire d'interpréter correctement le jugement du lecteur) [Euzenat2000d]. L'exemple de la représentation de connaissance sous divers formalismes se pose de manière cruciale lorsque les utilisateurs de logiciels différents construisent une même ontologie de manière collaborative.

D'une manière générale, les travaux proposés sont au confluent de la sémantique des représentations de connaissance, de la sémiologie et de la sémantique des langues naturelles (car le but est qu'au travers des chaînes de traitement, l'information conserve son sens). Mais l'état de l'art dans le domaine de la sémiologie laisse peu d'espoir pour la résolution de problème à court terme. Il existe cependant un début de cadre exploitant de manière raisonnée les méthodes structuralistes dans la prise en compte de ces problèmes²⁶ qui doit être exploité et approfondi.

Une ébauche de solution consisterait à ajouter, à côté des descriptions de syntaxe et de sémantique, les règles d'interprétation des signes utilisés par l'émetteur de telle sorte que le message soit présenté au récepteur de telle manière que celui-ci interprète correctement le message (connaissant les règles d'interprétation du récepteur). Ce type de techniques applique des idées de profils de l'utilisateur à l'interprétation des langages.

L'une des pistes est proposée par Joseph Goguen, dans la continuité de ses travaux sur les institutions, qui développe une approche baptisée sémiologie algébrique visant à considérer les représentations en liaison avec leur usage²⁷. La sémiologie algébrique consiste à intégrer dans un système (de signes) non seulement la syntaxe et la sémantique des représentations mais aussi — autant que faire se peut — les règles d'interprétation qu'un utilisateur doit appliquer pour interpréter la représentation. Un morphisme transforme un tel système (la source) en un autre (la cible) destiné à être plus adapté à une utilisation particulière. La sémiologie algébrique étudie les systèmes de signes et leurs transformations dans le cadre de la théorie des catégories. Ce travail est un très bon point de départ pour envisager la manipulation de l'information structurée d'un point de vue unifié.

Cette thématique serait l'occasion de se rapprocher du domaine des sciences humaines et sociales et plus particulièrement des études sur la sémiologie et la communication.

Objectif □ Développer une conception de l'interprétation des systèmes de signes qui puisse être décrite informatiquement et la manière de les utiliser de telle sorte de garantir l'interopérabilité à ce niveau.

²⁶ Peter Bøgh Andersen, A theory of computer semiotics, Cambridge university press, Cambridge (UK), 1990 (rev. 1997)

²⁷ Joseph Goguen, An introduction to algebraic semiotics with application to user interface design, *Lecture notes in computer science* 1562:242-291, 1999 <http://www-cse.ucsd.edu/users/goguen/ps/as.ps.gz>

4. DÉVELOPPEMENT LOGICIEL

Le développement logiciel est d'abord pour EXMO l'occasion de mettre en œuvre les concepts dégagés et d'expérimenter les résultats de notre travail. Le développement de logiciel peut aussi être l'occasion de valorisation sous forme de transfert technologique. Nous sommes très attachés à l'idée que nos développements puissent rester utilisables à des fins de recherche. Pour cela nous les implémentons dans un cadre ouvert et, si possible, standard (nous avons eu de mauvaises expériences avec des logiciels propriétaires). De même nous sommes très attaché à ce que nos réalisations soient ouvertes et puisse être appropriées par toutes les parties intéressées, que ce soit à des fins d'utilisation, de recherche ou de développement complémentaire. Dans le cadre de TRANSMORPHER, cette politique semble être un facteur important dans l'intérêt porté par d'autres acteurs.

4.1 TRANSMORPHER ☐ expression et exécution des flux de transformations

Comme on a pu le voir plus haut, pour établir ou vérifier des propriétés sur des transformations, il faut en disposer d'une représentation. Le langage XSLT, en exprimant les transformations en XML, permet de faire cela à peu de frais, mais reste verbeux et délicat à analyser. Afin de résoudre ce problème, nous avons conçu et développons en collaboration avec la société FLUXMEDIA le logiciel TRANSMORPHER qui se présente comme une sur-couche à XSLT permettant d'exprimer des flux de transformations tels que celui présenté à la figure 1.

TRANSMORPHER est un environnement permettant de définir et d'exécuter des transformations génériques sur des documents XML. Il a pour but de proposer des compléments du langage XSLT afin de ☐

- décrire simplement des transformations simples (suppressions d'éléments, remplacements de noms d'attributs, assemblage de documents, substitutions d'expressions régulières sur le contenu...) ☐
- assembler des transformations en connectant leurs entrées et sorties (multiples) ☐
- appliquer des transformations jusqu'à leur inapplicabilité ☐
- intégrer des moteurs de transformations externes.

TRANSMORPHER [Euzenat2001d, 2002c] permet de décrire en XML des flux de transformations, c'est-à-dire des ensembles de transformations connectées par des canaux d'entrée-sortie. Les canaux véhiculent l'information à transformer, principalement du XML. Les transformations peuvent être d'autres flux de transformations ou des transformations élémentaires. TRANSMORPHER définit un ensemble d'abstractions de transformations élémentaires dotées d'un modèle d'exécution et des instanciations de ces transformations. On peut décliner ainsi les transformations élémentaires ☐ appel externe (incluant XSLT), dispatcheur, agrégateur, sérialiseur, générateur, moteurs de requêtes, itérateurs et systèmes de règles.

Les systèmes de règles permettent de décrire des transformations dans un langage plus simple qu'XSLT. Actuellement, il est exécuté en le transformant directement en une transformation XSLT (étendue aux substitutions d'expressions régulières).

Une extension de ce logiciel consiste à attacher des assertions à chaque transformation (élémentaire, externe ou composée) permettant de signifier si une propriété est prouvée, supposée, à vérifier ou à démontrer. Ceci permettra de montrer des exemples réels de détermination de propriétés de transformations composées (voir §3.4).

L'étude de la concurrence (Cocoon, ☐xKit, Xweb) montre d'une part un réel besoin et d'autre part une relative absence d'outils ciblant spécifiquement l'aspect transformation

(indépendant de l'aspect formatage les travaux comparables à TRANSMORPHER sont très liés à la création de site Web). Le positionnement de TRANSMORPHER est plus général et destiné à la circulation et adaptation de l'information (qui peut être destiné à des applications très variées).

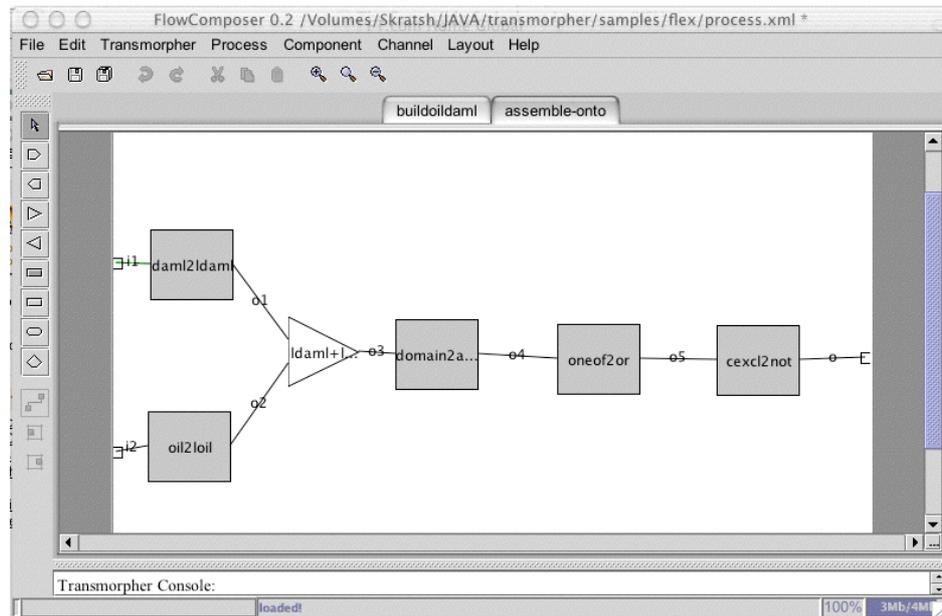


Figure 2 FlowComposer, l'éditeur graphique de Transmorpher.

Le développement de TRANSMORPHER répond pour EXMO à la nécessité de disposer d'un logiciel de transformation d'XML à la fois portable et analysable afin de développer ses activités de recherche. S'il correspond à un savoir-faire, il n'est pas réellement dans l'axe des objectifs du projet proposé.

TRANSMORPHER est développé en partenariat avec la société FLUXMEDIA. TRANSMORPHER est distribué depuis l'été 2001 sous forme ouverte (GNU General Public License). Sa version 0.9.7 est disponible sur le CDROM des logiciels libres de l'INRIA. Malgré une faible publicité (un site web et une présentation à une conférence), il a suscité une dizaine de contacts d'utilisateurs (TZI Bremen, VUAmsterdam, LIMBIO, LISI), développeurs (Lagoon, Xweb, ID, LRI/ENS, OASIS, Sax Machines) ou intégrateurs (ThinkAlong) de TRANSMORPHER. Le logiciel fonctionne correctement, mais continue à faire l'objet de développements pour faciliter son utilisabilité.

Le système peut être obtenu depuis <http://transmorpher.inrialpes.fr>.

4.2 DLML : Description Logic Markup Language

DLML [Euzenat2001e] n'est pas un langage mais un système modulaire de description de types de documents (DTD) permettant de décrire la syntaxe et la sémantique de nombreuses logiques de descriptions (voir §3.3.1). Une importante motivation de DLML est de pouvoir incorporer des représentations en langage formel (ici les logiques de descriptions) dans des documents XML en particulier pour l'action ESCRIRE (voir §5.2.1).

DLML tire parti de la conception modulaire des logiques de descriptions pour définir les logiques par assemblage de constructeurs élémentaires. Le système contient la spécification de la syntaxe (DTD) et de la sémantique (DSD) de plus de 40 opérateurs élémentaires. Ainsi, la définition d'une logique est donnée par la spécification, en XML, des opérateurs qu'elle autorise. À partir de cette déclaration, DLML est capable d'assembler les spécifications

élémentaires pour engendrer les DTD et DSD de la logique elle-même. Plus de 25 logiques répertoriées sont décrites sur le site.

Le langage DLML est associé à un ensemble de transformations XSLT permettant de transformer des descriptions DLML d'une logique vers une autre (soit en normalisant, soit en appauvrissant l'information), d'un langage (par exemple, de syllogismes) vers une logique de descriptions, d'afficher une description à l'aide de LaTeX ou d'importer des descriptions d'un autre système (FACT de l'université de Manchester).

Le système peut être utilisé depuis <http://co4.inrialpes.fr/xml/dlml/>.

5. RELATIONS NATIONALES ET INTERNATIONALES

5.1 Relations locales

Nous avons des relations de longue date et continues avec les partenaires non INRIA locaux. Nous présentons les deux principales ci-dessous. Ces relations cependant ne sont pas de nature à donner lieu, en l'état, à une collaboration étroite entre EXMO et ces partenaires en raison d'une différence de thématique (agents, raisonnement à partir de cas) et d'un déséquilibre entre les tailles des équipes respectives. D'autres contacts avec le CLIPS ou le centre de recherche de Xerox ont fait l'objet, dans le passé, de propositions communes.

5.1.1 Leibniz (Grenoble)

Le projet SHERPA était équipe fondatrice du laboratoire LEIBNIZ. Il est donc naturel que des liens aient persisté. L'équipe avec laquelle nous avons le plus de contact est MAGMA (Yves Demazeau). Ces contacts se sont principalement limités à des échanges de vues sur la communication entre agents sans avoir été concrétisés par une collaboration effective. Jérôme Euzenat a par ailleurs assuré la coordination du profil «*Inférence et connaissance*» du DEA «*Informatique, systèmes et communication*», puis du séminaire du profil dans lequel LEIBNIZ est très impliqué.

5.1.2 LISI (Lyon)

Nous participons avec l'équipe d'Alain Mille au projet ARDECO (associant aussi le projet Eiffel de l'INRIA Rocquencourt et Dassault Systèmes). Le but du projet est d'exploiter les traces d'utilisation du logiciel de conception assistée par ordinateur CATIA pour proposer à l'utilisateur des pistes de réutilisation de raisonnements passés. Il s'agit essentiellement de raisonnement à partir de cas. Ce travail, mené par Pierre-Antoine Champin pour la partie informatique, s'appuie sur une représentation d'épisodes de conception reconnus dans le flux d'utilisation et une appréhension de la similarité entre les épisodes.

Eu égard à la relative différence d'approche et d'objectifs entre l'équipe «*Modélisation des données et des connaissances*» du LISI et EXMO, la coopération n'est pas actuellement très développée. Mais nous avons des préoccupations communes autour du «*Web sémantique*», de l'expression du contenu et de l'adaptation de représentation. Nous travaillons par ailleurs avec plusieurs membres de l'équipe du LISI (Yannick Prié, Mohand-Saïd Hacid) au sein de l'action spécifique «*Web sémantique*» (voir §5.3.3).

5.2 Relations au sein de l'INRIA

L'INRIA dispose d'assez peu de ressources liées à l'échange (et même la manipulation) de connaissance formalisée. Aussi nous collaborons avec les équipes dont la tâche est liée à cette activité. Mais d'autres relations sont possibles avec des projets ayant une contribution sur la réécriture (PROTHÉO) ou les assistants de preuve (LEMME, LOGICAL).

5.2.1 Action de recherche coopérative *ESCRIRE* (ACACIA, ORPAILLEUR, EXMO)

L'ARC *ESCRIRE*, pilotée par EXMO, impliquait en outre les équipes ACACIA et ORPAILLEUR. Son but était de tirer parti de savoir faire complémentaires dans les trois équipes pour comparer trois types de représentations de connaissance (graphes conceptuels, représentations de connaissance par objets et logiques de descriptions) du point de vue de la représentation du contenu de documents et de sa manipulation.

La mise à l'épreuve de ces différents formalismes pour le traitement d'un jeu de documents a nécessité une réflexion méthodologique sur le passage des textes à leur représentation formelle (de façon suffisamment indépendante des formalismes employés) en lien avec le type d'accès que l'on veut avoir sur ces documents [Euzenat2001c].

On a défini en XML un ensemble de langages pivots (c'est-à-dire indépendants des représentations de connaissance utilisées) pour décrire des ontologies et des documents, des requêtes et des réponses à ces requêtes. On a par ailleurs développé une interface permettant de poser graphiquement la même requête à plusieurs systèmes et de collecter les réponses à ces requêtes [AlHulou2002a].

On trouvera plus d'informations sur *ESCRIRE* à <http://www.inrialpes.fr/exmo/cooperation/escrire>.

Par ailleurs, les membres des trois équipes ainsi que d'AXIS participent au réseau thématique IST ONTOWEB dont EXMO est le correspondant (voir §5.4.3). Nous avons aussi collaboré avec ACACIA dans le club CRIN consacré à la «*Mémoire d'entreprise*» et avec ORPAILLEUR dans diverses activités liées à la représentation de connaissance par objets [Napoli2000a].

5.2.2 Projet WAM (ex-OPÉRA)

Si EXMO s'intéresse à l'interopérabilité, WAM s'est toujours intéressé aux transformations de documents structurés. Cette activité prend de la vigueur avec la notion d'adaptation de ces documents.

Il y a un point de contact non négligeable entre les travaux des deux équipes sur le sujet des transformations. Cette activité est destinée à prendre une place centrale dans les travaux du projet WAM alors que le travail d'EXMO à ce sujet, n'est lié qu'à la conception du logiciel TRANSMORPHER destiné à résoudre un besoin non satisfait par ailleurs. Il est tout à fait envisageable que cette partie de l'activité d'EXMO passe progressivement sous le contrôle d'un futur projet WAM.

Nous avons par ailleurs entamé une collaboration avec Nabil Layaida sur l'application de l'approche sémantique à la transformation de documents multimédia (voir §3.2.2). Il serait bon que ce travail puisse se poursuivre (voir §5.3.1).

5.3 Relations nationales

Nous avons quelques relations particulièrement développées au niveau national, mais en général, nous participons à l'animation de la communauté. Jérôme Euzenat est membre élu du bureau de l'Association Française pour l'Intelligence Artificielle, ce qui nous met au contact de l'ensemble de la communauté française du domaine.

5.3.1 Institut National de l'Audiovisuel

Nous collaborons avec le département recherche et expérimentation de l'INA (Bruno Bachimont) sur la représentation du contenu de documents audiovisuels. C'est plus précisément le travail de thèse de Raphaël Troncy que nous co-encadrons qui consiste à développer un dispositif d'intégration de représentation de connaissance au sein de la représentation de la structure du document audiovisuel. Ces travaux s'appuient sur les développements actuels des langages développés pour le «Web sémantique» et sur les méthodologies de développement d'ontologies proposées à l'INA.

D'autres collaborations, associant sans doute WAM, sont à l'étude concernant les langages de description de documents audiovisuels et la prise en compte a priori de descripteurs facilitant l'adaptabilité des documents audio-visuels

5.3.2 Laboratoire d'Informatique Médicale et Bio-Informatique (Paris 13)

Nous avons entamé une collaboration avec l'équipe de Jean-Daniel Zucker alors au LIP6 et maintenant au LIMBIO. Cette collaboration qui vient juste de commencer a plusieurs buts. Le premier est de développer autour de TRANSMORPHER une bibliothèque d'opérateurs d'abstraction permettant d'obtenir une représentation d'un problème d'apprentissage adaptée à un algorithme efficace. Cela permet donc d'expérimenter TRANSMORPHER. Le second but est de caractériser, pour ces opérateurs, les propriétés préservées et celles qui ne le sont pas de sorte de caractériser la représentation obtenue en sortie. Cela constituerait une autre exploration de la problématique des transformations et propriétés (§3.2). Le dernier but de cette collaboration est de construire, autour de TRANSMORPHER, un moteur d'exploration de l'espace des transformations (guidé par les propriétés satisfaites) de telle sorte d'évaluer la meilleure stratégie d'apprentissage.

5.3.3 Action spécifique «Web sémantique»

Le département STIC du CNRS a lancé une série d'«actions spécifiques» dans le but de proposer des lignes d'actions. EXMO est impliqué dans celle liée au «Web sémantique» pilotée par Jean Charlet, Philippe Laublet et Chantal Reynaud. À cette action collaborent aussi les membres d'ACACIA et d'ORPAILLEUR ainsi que les membres du LISI cités plus haut. Le but du travail de prospective mené à l'heure actuelle consiste à dégager les points forts des intervenants français et leur apport dans le «Web sémantique» ainsi que leur insertion dans les travaux en cours au niveau international.

5.4 Relations internationales

Dans le cadre de nos activités, nous sommes amenés à côtoyer l'ensemble des acteurs du «Web sémantique» que ce soit informellement, ou dans le cadre d'actions que nous avons organisé comme le 1er Semantic Web Working Symposium (<http://www.semanticweb.org/SWWS>) et le livre qui s'en suit ou le workshop stratégique UE-NSF (<http://www.ercim.org/EU-NSF/SemWeb.html>). Nous collaborons plus étroitement avec les centres suivants.

5.4.1 Technologie-Zentrum Informatik Bremen / Vrije Universiteit Amsterdam

Nous avons collaboré informellement l'année passée avec Heiner Stuckenschmidt alors à TZI sur la définition de la notion de famille de langages et son implémentation dans DLML (voir §3.3.2 et §4.2). Cette collaboration a fait l'objet d'un financement du programme d'action intégré Procope du ministère des affaires étrangères.

Nous poursuivons actuellement cette collaboration toujours avec Heiner Stuckenschmidt, maintenant à Amsterdam et ses collègues à Brème (Ubbo Visser).

5.4.2 Université de Montréal

Nous avons un projet de collaboration avec Petko Valchev, maintenant à l'université de Montréal sur l'alignement d'ontologies. Ce sujet est devenu très important dans la conception de traducteur entre ontologies. En effet, afin d'assurer l'interopérabilité entre deux services de commerce électronique, par exemple, il est nécessaire de savoir mettre en correspondance les concepts présents dans ces ontologies. Les travaux actuellement consacrés au sujet sont tous parcellaires. Petko a développé un savoir faire important sur ce sujet dans le courant de sa thèse en définissant une mesure de similarité entre classes de différentes taxonomies prenant en compte de nombreux facteurs (types de données externes, collecteurs, circularités multiples entre objets). Nous avons pour but d'étendre ces techniques aux langages actuellement développés pour représenter la connaissance sur le «Web sémantique».

5.4.3 ONTOWEB: Ontology-based information exchange for knowledge management and electronic commerce

EXMO a été l'un des promoteurs et est impliqué dans le réseau thématique ONTOWEB ("Ontology-based information exchange for knowledge management and electronic commerce") financé par l'union européenne. Le nœud INRIA d'ONTOWEB est constitué du projet ACACIA et des actions AXIS, EXMO et ORPAILLEUR. Il a en charge le module dévolu à promouvoir la collaboration internationale sur les thématiques d'ONTOWEB. EXMO représente l'INRIA au sein du bureau exécutif du projet.

Dans le cadre d'ONTOWEB, nous avons organisé la conférence SWWS l'année passée et participons à l'organisation d'ISWC cette année. À la suite de l'organisation de SWWS, la "Semantic web science association" a été créée dont le but est de donner à caractère permanent à ces manifestations. Jérôme Euzenat en est un membre fondateur.

Nous réfléchissons, en liaison avec ERCIM, à un prolongement d'OntoWeb dans le 6^e programme cadre de recherche et développement de l'union européenne.

On trouvera plus d'informations sur ONTOWEB à <http://www.ontoweb.org>.

5.4.4 W3C: WebOnt

Nous allons représenter l'Inria auprès du groupe de travail WEBONT chargé de développer le langage pour le «Web sémantique» pour l'instant dénommé OWL.

On trouvera plus d'informations sur WEBONT à <http://www.w3.org/2001/sw/webont>.

6. COMPOSITION DE L'ACTION EXMO

L'action EXMO est réduite mais déterminée. Le recrutement cette année d'un chargé de recherches et d'un ingénieur contractuel permet d'avancer sur les plans de la recherche, du développement logiciel et de la promotion de nos travaux. Cela devrait aussi permettre d'être plus présent dans le domaine, en particulier en participant à l'activité «Web sémantique» du W3C, ce à quoi nous avons dû renoncer jusqu'à présent faute de temps.

Notre préoccupation pour l'instant est d'encadrer un flux raisonnable d'étudiants de DEA et de doctorat pour développer nos thématiques et présenter des candidats de qualité.

Chercheurs INRIA

Jean-François Baget [CR INRIA, depuis le 1/9/2002]
Jérôme Euzenat [CR INRIA, responsable scientifique]

Ingénieur associé

Fabien Triolet [Ingénieur ODL TRANSMORPHER, depuis le 1/7/2002]

Doctorants

Olivier Brunet [ATER Université Joseph Fourier]
Pierre-Antoine Champin [CIFRE (Dassault Système) avec A. Mille (LISI)]
Raphaël Troncy [CIFRE (Institut National de l'Audiovisuel) avec B. Bachimont]

Assistante de projet

Marion Ponsot [INRIA]

7. PUBLICATIONS (DEPUIS LA CRÉATION DE L'ACTION EXMO)

Journal articles/Articles de revues

- [napoli2000a] Amedeo Napoli, Jérôme Euzenat, Roland Ducournau, **Les représentations de connaissances par objets**, *Techniques et science informatique* 19(1-3):387-394, 2000
- [cerbah2001a] Farid Cerbah, Jérôme Euzenat, **Traceability between models and texts through terminology**, *Data and knowledge engineering* 38(1):31-43, 2001
- [euzenat2001g] Jérôme Euzenat, **Granularity in relational formalisms with application to time and space representation**, *Computational intelligence* 17(4):703-737, 2001
- [euzenat2002b] Jérôme Euzenat, **Eight questions about semantic web annotations**, *IEEE Intelligent systems* 17(2):55-62, 2002
<http://computer.org/intelligent/ex2002/x2055abs.htm>
- [euzenat2002c] Jérôme Euzenat, Laurent Tardif, **XML transformation flow processing**, *Markup languages: theory and practice* 3(3):285-311, 2002

PhD theses/Thèses de doctorat

- [brunet2002b] Olivier Brunet, **Étude de la connaissance dans le cadre d'observations partielles et la logique de l'observation**, Thèse d'informatique, université Joseph Fourier, Grenoble (FR), 25 octobre 2002
<ftp://ftp.inrialpes.fr/pub/exmo/theses/these-brunet.pdf>

Book chapters and collected papers/Chapitres de livres

- [euzenat2001a] Jérôme Euzenat, **Construction collaborative de bases de connaissance et de documents pour la capitalisation**, In: Manuel Zacklad, Michel Grundstein (éds), *Ingénierie et capitalisation des connaissances*, Hermès Science publisher, Paris (FR), 2001, pp25-48
- [euzenat2002d] Jérôme Euzenat, **An infrastructure for formally ensuring interoperability in a heterogeneous semantic web**, In: Isabel Cruz, Stefan Decker, Jérôme Euzenat, Deborah McGuinness (eds), *The emerging semantic web*, IOS press, Amsterdam (NL), 302 p., 2002, pp

Book coordination/Coordination d'ouvrages

- [cruz2002a] Isabel Cruz, Stefan Decker, Jérôme Euzenat, Deborah McGuinness (eds), **The emerging semantic web**, IOS press, Amsterdam (NL), 302 p., 2002
<http://www.inrialpes.fr/exmo/papers/emerging/>

Conference editing/Publication d'actes

- [cruz2001a] Isabel Cruz, Stefan Decker, Jérôme Euzenat, Deborah McGuinness (eds), **Semantic web working symposium**, 597p., 2001
<http://www.semanticweb.org/SWWS/program/full/SWWSProceedings.pdf>
- [euzenat2002f] Jérôme Euzenat, Asuncion Gomez Perez, Nicola Guarino, Heiner Stuckenschmidt (eds), **ECAI 2002 workshop on Ontologies and semantic interoperability**, , 597 p., 2002
<ftp://ftp.inrialpes.fr/pub/exmo/reports/ECAI2002-OIS-ws.pdf>

Conference papers/Communication à des conférences

- [bougé2000a] Patrick Bougé, Dominique Deneux, Christophe Lerch, Jérôme Euzenat, Jean-Paul Barthès, Michel Tollenaere, **Localisation des connaissances dans les systèmes de production: approches multiples pour différents types de connaissance**, Jacques Perrin, René Soenen (éds), Actes journées Prosper sur Gestion de connaissances, coopération, méthodologie de recherches interdisciplinaires, Toulouse (FR), pp31-50, 2000 <http://www.univ-valenciennes.fr/PROSPER/serveur/doc/toulouse/Prosper-loc-formatted.zip>
- [brunet2000a] Olivier Brunet, **Lattice approach to classifications**, Proc. 12th European summer school on logic, language and information (ESSLLI), Birmingham (U K) , p p 3 4 - 4 4 , 2 0 0 0
<ftp://ftp.inrialpes.fr/pub/exmo/publications/brunet2000a.ps.gz>
- [brunet2000b] Olivier Brunet, **Classifications et treillis**, Actes 5^e Rencontres nationales sur de jeunes chercheurs en intelligence artificielle (RJCIA), Lyon (FR), pp29-38 , 2 0 0 0
<ftp://ftp.inrialpes.fr/pub/exmo/publications/brunet2000b.ps.gz>
- [cerbah2000a] Farid Cerbah, Jérôme Euzenat, **Using terminology extraction techniques for improving traceability from formal models to textual requirements**, Proc. 5th international conference on applications of natural language to information systems (NLDB), Versailles (FR), (Mokrane Bouzeghoub, Zoubida Kedad, Elisabeth Métais (eds), Natural Language Processing and Information Systems, *Lecture notes in computer science* (1959), 2001), pp115-126, 2000
<ftp://ftp.inrialpes.fr/pub/exmo/publications/cerbah2000a.ps.gz>
- [cerbah2000b] Farid Cerbah, Jérôme Euzenat, **Integrating textual knowledge and formal knowledge for improving traceability**, Proc. ECAI workshop on Knowledge Management and Organizational Memory, Berlin (DE), pp10-16, 2000
<http://www-sop.inria.fr/acacia/WORKSHOPS/ECAI2000-OM/Papers/ecai2000-cerbah.ps>
- [cerbah2000c] Farid Cerbah, Jérôme Euzenat, **Integrating textual knowledge and formal knowledge for improving traceability**, Proc. 12th international conference on knowledge engineering and knowledge management (EKAW), Juan-les-Pins (FR), (Rose Dieng, Olivier Corby (eds), Knowledge engineering and knowledge management: methods, models and tools, *Lecture notes in computer science* 1937, 2000), pp296-303, 2000

- [euzenat2000a] Jérôme Euzenat, **XML est-il le langage de représentation de connaissance de l'an 2000 ?**, Actes 6^e journées sur langages et modèles à objets (LMO), Mont Saint-Hilaire (CA), pp59-74, 2000 <ftp://ftp.inrialpes.fr/pub/exmo/publications/euzenat2000a.ps.gz>
- [euzenat2000c] Jérôme Euzenat, **Problèmes d'intelligibilité et solutions autour de XML**, Paul Kopp (éd), Actes séminaire CNES sur Valorisation des données, Labège (FR), 2000 <ftp://ftp.inrialpes.fr/pub/exmo/publications/euzenat2000c.ps.gz>
- [euzenat2000d] Jérôme Euzenat, **Towards formal knowledge intelligibility at the semiotic level**, Proc. ECAI workshop on applied semiotics: control problems, Berlin (DE), pp59-61, 2000 http://www.iitp.ru/asc2000/ps/12_EUZEN.PS
- [brunet2001a] Olivier Brunet, **A Model for Knowledge Representation in Distributed Systems**, Proc. KI workshop on Modal Logic in AI, Wien (AT), pp3-12, 2001 <ftp://ftp.inrialpes.fr/pub/exmo/publications/brunet2001a.ps.gz>
- [euzenat2001b] Jérôme Euzenat, **Towards a principled approach to semantic interoperability**, Asuncion Gomez Perez, Michael Gruninger, Heiner Stuckenschmidt, Michael Uschold (eds), Proc. IJCAI 2001 workshop on ontology and information sharing, Seattle (WA US), pp19-25, 2001 <http://ceur-ws.org/Vol-47/euzenat.pdf>
- [euzenat2001c] Jérôme Euzenat, **L'annotation formelle de documents en (8) questions**, Jean Charlet (éd), Actes 6^e journées sur ingénierie des connaissances (IC), Grenoble (FR), pp95-110, 2001
- [euzenat2001d] Jérôme Euzenat, Laurent Tardif, **XML transformation flow processing**, Proc. 2nd conference on extreme markup languages, Montréal (CA), pp61-72, 2001 <http://transmorpher.inrialpes.fr/paper/>
- [euzenat2001e] Jérôme Euzenat, **Preserving modularity in XML encoding of description logics**, Deborah McGuinness, Peter Patel-Schneider, Carole Goble, Ralph Möller (eds), Proc. 14th workshop on description logics (DL), Stanford (CA US), pp20-29, 2001 <http://ceur-ws.org/Vol-49/Euzenat-20start.ps>
- [euzenat2001f] Jérôme Euzenat, **An infrastructure for formally ensuring interoperability in a heterogeneous semantic web**, Proc. 1st international on semantic web working symposium (SWWS), Stanford (CA US), pp345-360, 2001 <http://www.semanticweb.org/SWWS/program/full/paper16.pdf>
- [stuckenschmidt2001a] Heiner Stuckenschmidt, Jérôme Euzenat, **Ontology Language Integration: A Constructive Approach**, Proc. KI 2001 workshop on Applications of Description Logics, Wien (AT), 2001 <http://ceur-ws.org/Vol-44/StuckenschmidtEuzenat.ps.gz>
- [alhulou2002a] Rim Al-Hulou, Olivier Corby, Rose Dieng-Kuntz, Jérôme Euzenat, Carolina Medina Ramirez, Amedeo Napoli, Raphaël Troncy, **Three knowledge representation formalisms for content-based representation of documents**, Proc. KR 2002 workshop on Formal ontology, knowledge representation and intelligent systems for the world wide web (SemWeb), Toulouse (FR), 2002

- [bachimont2002a] Bruno Bachimont, Raphaël Troncy, Antoine Isaac, **Semantic Commitment for Designing Ontologies: A Proposal**, Proc. 13th international conference on knowledge engineering and knowledge management (EKAW), Siguenza (ES), 2002
- [brunet2002a] Olivier Brunet, **A modal logic for observation-based knowledge representation**, Proc. FLoC workshop on Intuitionistic modal logic and applications (IMLA 2002), Copenhagen (DK), (Rajeev Goré, Michael Mendler, Valeria de Paiva (eds), , 61, Bamberger Beitrage zur Wirtschaftsinformatik und Angewandten Informatik, Bamberg (DE), 2002), p p 6 9 - 8 1 , 2 0 0 2 <ftp://ftp.inrialpes.fr/pub/exmo/publications/brunet2002a.pdf>
- [euzenat2002e] Jérôme Euzenat, Heiner Stuckenschmidt, **The `family of languages' approach to semantic interoperability**, Borys Omelayenko, Michel Klein (eds), Proc. ECAI 2002 workshop on Knowledge Transformation for the Semantic Web, Lyon (FR), pp92-99, 2002 <ftp://ftp.inrialpes.fr/pub/exmo/publications/euzenat2002e.pdf>
- [troncy2002a] Raphaël Troncy, Antoine Isaac, **DOE : une mise en oeuvre d'une méthode de structuration différentielle pour les ontologies**, Actes 13^e journées francophones sur Ingénierie des Connaissances (IC), Rouen (FR), pp63-74, 2002

Prospective reports/Rapports de prospective

- [euzenat2002a] Jérôme Euzenat (ed), **Research challenges and perspectives of the Semantic web**, EU-NSF Strategic report, ERCIM, Sophia Antipolis (FR), 82 p., 2002 <http://www.ercim.org/EU-NSF/semweb.html>

Deliverable/Rapports de contrats

- [euzenat2001h] Jérôme Euzenat (ed), **1st international semantic web working symposium (SWWS-1)**, deliverable 7.6, IST OntoWeb thematic network, INRIA Rhône-Alpes, Montbonnot (FR), 30 p., 2001

Booklets and manuals/Manuels

- [exmo2001a] action Exmo, société FluxMedia, **Transmorpher 1.0**, Reference manual, INRIA Rhône-Alpes/FluxMedia, Grenoble (FR), 2001 <http://transmorpher.inrialpes.fr/refman>

Master theses/Mémoires de DEA

- [troncy2000a] Raphaël Troncy, **Intégration texte-représentation formelle pour la gestion de documents XML**, DEA d'informatique, Université Joseph Fourier-INPG, Grenoble (FR), 2000 <ftp://ftp.inrialpes.fr/pub/exmo/reports/dea-troncy.ps.gz>

Miscellaneous/Divers documents publics

- [charre2002a] Bruno Charre, **Intégration texte-représentation formelle pour la gestion de documents XML**, DESS d'intelligence artificielle, Université Pierre et Marie Curie, Paris (FR), 2002
<ftp://ftp.inrialpes.fr/pub/exmo/reports/dessia-charre.pdf>
- [troncy2000b] Raphaël Troncy, **Intégration texte-représentation formelle pour la gestion de documents XML**, Magistère d'informatique, 2000
<ftp://ftp.inrialpes.fr/pub/exmo/reports/troncy2000b.ps.gz>
- [euzenat2002g] Jérôme Euzenat (éd.), «**Semantic web**» special issue, *Ercim news* 51, 2002
<ftp://ftp.inrialpes.fr/pub/exmo/reports/ErcimNews51.pdf>
- [euzenat2002h] Jérôme Euzenat, **Research challenges and perspectives of the semantic web**, *IEEE Intelligent systems* 17(5):86-88, 2002