

*Projet L&D**Langue et Dialogue**Lorraine*

THÈME 3A



*R*apport
*d'**A*ctivité

2002

Table des matières

1. Composition de l'équipe	1
2. Présentation et objectifs généraux	2
3. Fondements scientifiques	2
3.1. Fondements scientifiques	2
3.1.1. Bases théoriques	2
3.1.2. Sémantique et Inférence	3
3.1.3. Multimodalité	4
3.1.4. Ingénierie linguistique	4
3.1.5. Etudes empiriques	4
4. Domaines d'application	4
4.1. Analyse	4
4.2. Génération	5
4.3. Inférence et discours	6
4.4. Dialogue multimodal	6
4.5. Ressources linguistiques et multimédias	7
4.6. Recherche d'information dans les sources de données génomiques	8
5. Logiciels	8
5.1. Liste des logiciels	8
5.2. SOAPical	9
5.2.1. Soapmeter	9
5.2.2. Soapcaster	9
5.3. TAXI : The Adjustable XML edItor	9
5.4. API Feature Structure	9
5.5. Analyseur LTAG	9
5.6. Compilateur de Méta-grammaire	10
5.7. CURT	10
5.8. Indigen	10
5.9. plate-forme CINERGIE	10
5.10. XMAP	11
5.10.1. Xmap_INTERACTIVE	11
5.10.2. Xmap_AUTO	11
5.10.3. Xmap_DB	11
5.10.4. Xmap_SHOW	11
5.10.5. Xprom	11
6. Résultats nouveaux	11
6.1. Logique Hybride	11
6.2. Génération	12
6.3. MMIL, un langage de représentation de contenus multimodaux	12
8. Actions régionales, nationales et internationales	13
8.1. Actions industrielles	13
8.1.1. EADS/S&DE(Vulcain)	13
8.2. Actions régionales	13
8.2.1. Diatelic	13
8.2.2. Projet « Ingénierie des langues, du document et de l'Information Scientifique, Technique et Culturelle » dans le cadre du plan État-Région Lorraine	14
8.3. Actions nationales	14
8.3.1. CINERGIE	14

8.3.2.	ARC RLT	15
8.3.3.	ARC GENI	15
8.3.4.	ANANAS	15
8.3.5.	GDR Sémantique et Modélisation : Groupement de recherche 2521 (CNRS)	16
8.4.	Actions européennes	16
8.4.1.	Université d'Amsterdam	16
8.4.2.	Projet Indigen	16
8.4.3.	Projet MIAMM	17
8.4.4.	Projet OZONE	17
8.5.	Actions internationales	17
8.5.1.	Common Ref	17
8.5.2.	Vietnam	18
8.5.3.	Taiwan	18
8.5.4.	ISO	18
9.	Diffusion des résultats	18
9.1.	Animation de la communauté scientifique	18
9.1.1.	Direction	18
9.1.2.	Présidence	18
9.1.3.	Conseils, Comités, Commissions, Missions...	18
9.1.4.	Responsabilités et coordinations scientifiques	19
9.1.5.	Expertises	19
9.1.6.	Revue	19
9.1.7.	Colloques	19
9.1.8.	Organisations	20
9.1.9.	Séminaires et conférences invitées	20
9.2.	Enseignement universitaire	20
9.3.	Autres enseignements	21
10.	Bibliographie	21

1. Composition de l'équipe

L&D est un projet du LORIA (UMR 7503) commun au CNRS, à l'INRIA, à l'Université Henri Poincaré Nancy 1, à l'Université Nancy 2 et à l'Institut National Polytechnique de Lorraine. Pour plus de détails sur les activités de L&D, nous invitons le lecteur à consulter le site web de l'équipe à <http://www.loria.fr/equipes/led>.

Responsable scientifique

Laurent Romary [directeur de recherche INRIA]

Responsable permanent

Patrick Blackburn [directeur de recherche INRIA]

Ingénieurs, assistants et techniciens

Isabelle Blanchard [technicienne CNRS (à mi-temps), assistante de projet]

Nadia Viscogliosi [ingénieur d'étude CNRS]

Chercheurs titulaires

Marie-Dominique Devignes [chargée de recherche CNRS (département SDV)]

Bertrand Gaiffe [chargé de recherche CNRS (département SPI)]

Claire Gardent [chargée de recherche CNRS (département SHS)]

Matthieu Quignard [chargé de recherche CNRS (département SHS)]

Enseignants-chercheurs titulaires

Jean-Marie Pierrel [professeur à l'UHP - Nancy 1]

Daniel Coulon [professeur - Ecole des Mines - INPL]

Nadia Bellalem [maître de conférences - IUT Nancy-Charlemagne - Université Nancy 2]

Samuel Cruz-Lara [maître de conférences - IUT Nancy-Charlemagne - Université Nancy 2]

Christine Fay-Varnier [maître de conférences - Ecole de Géologie-INPL]

Jean-Luc Husson [maître de conférences - IUT de Nancy-Brabois - UHP Nancy 1]

Azim Roussanaly [maître de conférences - UFR MI - Université Nancy 2 (en détachement à l'INRIA)]

Malika Smaïl [maître de conférences - UFR STMIA - UHP Nancy 1]

Chercheurs doctorants

Evelyne Jacquey [ATER UHP - Nancy 1 jusqu'au 1/09/2002 puis associée]

Frédéric Landragin [bourse MENRT]

Hélène Manuelian [bourse INRIA]

Djamé Seddah [bourse INRIA]

Ti Minh Nguyen [bourse du MAE et enseignante à l'Université d'Hanoi]

Dalila Beckhouche [bourse CIFRE EADS/MSI]

Benoît Crabbé [bourse INRIA/Région - Moniteur - Université Nancy2]

Chercheurs associés, post-doc et invités

Susanne Alt [Chargée de recherche CNRS, ATILF]

Balder ten Cate [invité, ILLC, Université Amsterdam (NL)]

Johan Bos [invité, Université d'Edimbourg (GB)]

Laurence Danlos [associée professeur à l'Université de Paris 7]

Gérald Duffing [associé, enseignant à l'ICN, Université Nancy 2]

Nancy Ide [invité, Professeur au Vassar College (NY)]

Joshua Jen-Shing Hong [invité, "Chi-Nan University" de Taiwan]

Marteen Marx [invité, ILLC, Université Amsterdam (NL)]

Koichi Takeuchi [post-doc INRIA (projet HOTARU)]

Amalia Todirascu [post-doc jusqu'au 1/9/2002 puis associée (Maître de conférences à l'Université de Troyes)]

Kristina Striegnitz [Univ. Saarbruck (D)]

Ingénieurs experts

Charles Beiss [coordinateur de projets européens]

Pierre Dussaulx [Ingénieur expert sur contrat (projets MIAMM, SALT)]
Eric Kow [Ingénieur expert (projet MIAMM)]
Ashwani Kumar [Ingénieur expert sur contrat (projet MIAMM)]
Etienne Petitjean [Assistant ingénieur (projet DIATELIC)]
Annalisa Ricci [Ingénieur expert (projet CINERGIE)]

2. Présentation et objectifs généraux

L'objectif du projet Langue et Dialogue est de définir des modèles et des techniques permettant de mettre en œuvre à court, moyen ou long terme des systèmes de dialogue homme machine finalisés reposant sur une forte composante langagière. Dans ce cadre, notre activité se développe dans trois directions complémentaires :

- l'étude des mécanismes fondamentaux de la communication en langue naturelle seule ou accompagnée d'une désignation gestuelle (dialogue multimodal). Cette recherche s'effectue dans un contexte pluri-disciplinaire alliant linguistique et informatique principalement ;
- la réalisation de systèmes de dialogue effectifs dans le cadre notamment de collaborations industrielles. Cette activité nous permet par ailleurs de disposer d'une plateforme d'expérimentation pour la validation des différents modèles que nous concevons ;
- la définition d'outils et de méthodes génériques permettant d'étudier de façon fine des situations de dialogues réels, issus de la transcription d'expériences de simulation ou d'observations directes. Ce travail s'appuie sur une expérience acquise depuis plusieurs années sur la normalisation et la manipulation de ressources linguistiques (en particulier des « corpus »).

Nous restreignons notre ambition à du DHM finalisé, c'est-à-dire dédié à la commande, à l'apprentissage ou à la recherche d'informations dans un domaine clairement spécifié ; cela suppose que l'on peut expliciter complètement l'action - ou plus généralement l'intention - visée par l'utilisateur, de manière à ce qu'elle puisse être réalisée par le système ¹.

Il ne s'agit, cependant, pas de se limiter à l'emploi de quelques bribes de langage naturel pour apporter une consonnance naturelle. Les recherches que nous menons, visent à long terme, à mettre en œuvre une communication en langue naturelle entre un usager et une machine non contrainte à un contexte d'utilisation.

Le choix de partir de l'hypothèse d'un dialogue oral - même si nous n'effectuons pas de recherche sur la parole proprement dite - résulte de la volonté de gérer au plus près la dynamique de la tâche et correspond à l'expérience acquise par les chercheurs du groupe dans la prise en compte des aspects temporels dans la langue ainsi que de phénomènes plus particuliers tels que la déixis, que l'on peut observer par exemple dans des expériences de simulation de dialogues oraux. De fait, il nous semble important, dans un souci de respect de la spontanéité des utilisateurs potentiels de nos systèmes, de savoir modéliser les modes de référencement immédiate aux différentes composantes d'une tâche donnée, notamment quand celle-ci est visualisée sur un écran.

3. Fondements scientifiques

3.1. Fondements scientifiques

Mots clés : *langue, dialogue, syntaxe, sémantique, pragmatique, inférence, corpus, multimodalité, multimédia, ingénierie.*

3.1.1. Bases théoriques

De quelle façon fonctionne le langage humain ? Quels sont les processus impliqués lorsque deux personnes engagent un dialogue ? Comment font-ils exactement pour se comprendre ? Est-il possible de modéliser de

¹Dans la suite, nous parlerons d'application pour désigner ce domaine clairement spécifié

telles interactions sur ordinateur, et dans l'affirmative, quels sont les méthodes et outils à définir ? De telles questions soulèvent tout un ensemble de problèmes scientifiques, relatifs par exemple au fonctionnement sonore des langues humaines (phonologie), ou à l'organisation de leurs structures grammaticales (syntaxe). Au sein de l'équipe Langue et Dialogue, nous nous intéressons plus particulièrement aux aspects sémantiques et pragmatiques associés à ces questions, et aux problèmes informatiques associés.

De fait, le moment est propice pour aborder de telles questions. L'essentiel des problèmes fondamentaux relatifs aux analyses phonologiques ou syntaxiques a été résolu de façon satisfaisante, et il existe maintenant une large palette d'outils permettant de traiter ces phénomènes langagiers d'un point de vue automatique. Par ailleurs, les recherches menées en intelligence artificielle ont conduit à la définition d'un certain nombre de formalismes de représentation (tels que les logiques de description, les logiques temporelles ou les logiques épistémiques), qu'accompagnent de nombreux outils efficaces de déduction en logique du premier ordre issus du domaine du raisonnement automatique. Enfin, les recherches menées en sémantique formelle ont permis d'élaborer des formalismes de représentation (DRT, SDRT, sémantique dynamique, ou les formalismes de sous-spécification) particulièrement adaptés à la représentation des phénomènes propres au langage humain (anaphores, ambiguïtés de portée etc.), et le liens entre ces formalismes et les logiques classiques est relativement bien maîtrisé.

L'orientation adoptée par Langue et Dialogue vis-à-vis des problèmes sémantiques est ancrée dans ces développements. Le programme de recherche de l'équipe est de développer et d'utiliser de tels outils pour aborder différents problèmes dans le domaine de la sémantique des langues naturelles. Deux aspects sont particulièrement pris en compte au sein de l'équipe :

- l'importance de l'inférence en sémantique des langues naturelles,
- L'utilisation du langage dans environnements riches, notamment multimodaux.

L'approche envisagée par Langue et Dialogue est d'examiner de tels problèmes au regard des théories sémantiques les plus récentes, afin notamment de développer des implémentations expérimentales qui permettent de tester de telles théories, d'utiliser les méthodes les plus récentes, et au final, de relier nos recherches aux données empiriques.

3.1.2. *Sémantique et Inférence*

Les progrès à venir dans le domaine de la sémantique des langues naturelles dépendent étroitement de la compréhension du rôle joué par les phénomènes d'inférence. Au niveau le plus simple, l'inférence peut être vue comme un mécanisme de désambiguïsation. Les énoncés en langue naturelle sont en effet extrêmement ambigus : des opérations d'inférence, notamment lexicale, permettent ainsi d'éliminer les représentations incorrectes (d'un point de vue sémantique). Mais l'inférence peut aussi être utilisée pour des processus plus subtils, pour par exemple intégrer des informations nouvelles dans un contexte déjà connu. Inversement, dans le cas de la génération d'énoncés en langue naturelle, comment s'assurer que l'énoncé courant convient, étant donné les connaissances préalables du locuteur et du destinataire ?

Les théories les plus récentes en sémantique (DRT, SDRT et sémantiques dynamiques de manière générale) abordent activement ces problèmes. Elles intègrent systématiquement l'inférence comme un élément clef et, ce faisant, redéfinissent les limites effectives entre sémantique et pragmatique. Ainsi, l'algorithme de van der Sandt pour le traitement des présuppositions, reposant sur les principes de la DRT, utilise spécifiquement l'inférence pour garantir que de nouvelles informations sont intégrées de façon cohérente avec les plus anciennes.

L'équipe Langue et Dialogue explore ces problèmes sous plusieurs perspectives (par exemple pour la génération, ou l'analyse du discours) et ce sur la base d'approches méthodologiques complémentaires (modélisation à base de logiques modales, ou études de données expérimentales). Par exemple, certaines applications sont bien adaptées à des méthodes de surface (shallow parsing) combinées à de l'inférence rapide en logique de description. Dans d'autres cas, il peut être plus approprié d'analyser les énoncés ou phrases en profondeur pour ensuite s'appuyer sur des moteurs d'inférence de premier ordre. Notre but est d'explorer ces approches et leurs limitations.

3.1.3. Multimodalité

Les langues s'emploient toujours dans des contextes réels qui sont par essence multimodaux, c'est à dire intégrant d'autre canaux de communication que la parole ou le texte proprement dit. Pour le moins, des interlocuteurs en situation de dialogue sont toujours situés dans un endroit identifiés, à un certain moment dans le temps, et ont une conscience mutuelle de leur environnement. L'exploitation de ces informations dans le dialogue (par exemple à l'aide de désignation gestuelles) leur permet de communiquer avec plus d'efficacité. Les théories contemporaines en sémantique commencent à peine à intégrer ces facteurs (on pensera par exemple à la SDRT ou aux travaux de Henk Zeevat).

L'approche envisagée par Langue et Dialogue de la multimodalité reste fortement ancrée sur la complémentarité des autres informations avec le matériau linguistique. Elle vise à définir une description unifiée des différentes sources d'information qui puisse conduire au développement d'un modèle d'interprétation d'énoncés en contexte qui soit indépendant des modes de réalisation. Par ailleurs, l'utilisation des logiques de description semble un cadre adéquat pour l'intégration des modes sur la base de modèles conceptuels unifiés.

3.1.4. Ingénierie linguistique

Les implémentations réalisées au sein de l'équipe ont un rôle important au delà du simple apport de nouvelles applications. Elles permettent d'appréhender réellement la complexité inhérente à la gestion des phénomènes sémantiques et pragmatiques des langues naturelles. Nous pouvons ainsi expérimenter les interactions entre différentes sources d'information (par exemple entre contenus lexicaux et ontologies propres à un domaine d'application).

L'approche envisagée au sein de l'équipe est de développer des plateformes ouvertes intégrant des composants ou des outils correspondant à l'état de l'art en la matière. Cette approche, qui s'applique tout aussi bien à la définition de système de traitement des langues naturelles qu'à l'intégration de systèmes de dialogue homme-machine, repose à la fois sur des protocoles standards (par exemple SOAP) et sur des données parfaitement spécifiées (pour la représentation des contenus multimodaux par exemple). Par itération, on peut ainsi intégrer des composants de plus en plus riches d'un point de vue conceptuel et tendre véritablement vers une ingénierie linguistique raisonnée. De façon complémentaire, une telle approche ne fait sens que s'il existe des normes internationales pour la représentation des données linguistiques, ce qui explique l'implication forte de l'équipe dans cette direction.

3.1.5. Etudes empiriques

Le rôle des méthodes empiriques (apprentissage de modèles, extraction de connaissances à partir de corpus, évaluation) n'a fait que croître dans le domaine de l'informatique linguistique au cours des 15 dernières années. L'équipe Langue et Dialogue est dans ce cadre reconnue depuis des années pour avoir oeuvré à la création, à la gestion et à la diffusion de ressources linguistiques réutilisables par la communauté scientifique, qu'il s'agisse dans le cadre de la mise en oeuvre de serveurs de données (par exemple au travers du projet SILFIDE), ou de la définition de formats de représentation standardisée (par exemple TAGML). Nos travaux dans ce sens accompagnent constamment notre projet scientifique, qu'il s'agisse au niveau de la description de données lexicales réutilisables (pour la paramétrisation de systèmes de dialogue homme-machine par exemple), l'étude de phénomènes fins en sémantique (par la constitution de suites de test), ou l'annotation de corpus dans le cadre d'études en analyse ou en génération (étude de la référence).

Là encore, ces travaux n'ont de sens que si les données sont rendues pérennes par des cadres de représentation normalisés. Nous reconnaissons ainsi toute l'importance des initiatives récentes en la matière (comité TC37/SC4 de l'ISO) auxquelles sont liés plusieurs projets de notre équipe.

4. Domaines d'application

4.1. Analyse

Depuis plusieurs années [6], l'équipe L&D travaille à l'analyse syntaxique des langues naturelles. Dans ce cadre, nous maintenons un analyseur de grammaires d'arbres adjoints permettant l'analyse partielle d'énoncés.

Cet analyseur est par exemple utilisé dans le projet XMiner (analyse de dossiers médicaux) où on a affaire à des énoncés très incomplets.

Connexe à cet analyseur, nous développons des outils permettant la maintenance de grammaires. C'est dans ce cadre que nous visons l'intégration de logiques modales de description d'arbres, soit en vue de décrire une grammaire TAG de façon plus abstraite, soit en vue de faire directement de l'analyse.

L'analyse syntaxique n'étant pas une fin en soi, notre souci est de travailler sur des modèles permettant le plus facilement possible de construire une interprétation sémantique. De là notre intérêt pour l'utilisation de grammaires d'arbres adjoints dans l'analyse syntaxique intégrée à des systèmes de dialogue homme-machine.

4.2. Génération

Pour de nombreuses applications informatiques, les données de toutes natures qui sont produites sont des données complexes difficiles à interpréter par le non-expert. Un générateur de textes, qui prend en entrée ces données et produit un texte exprimant le contenu de ces données en langue naturelle, permet d'aborder ce problème.

Le problème de la génération de texte est essentiellement un problème de choix : étant donnée une multitude de paramètres linguistiques (ordonancement des phrases et des constituants, choix des unités et des catégories lexicales) et extra-linguistiques (contexte situationnel, connaissance du monde, profil de l'utilisateur), le problème est de fixer les bonnes valeurs pour l'ensemble de ces paramètres. Un texte de bonne qualité sera alors un texte pour lequel le générateur a fait les bons choix.

Comme l'a montré en particulier le travail de Matthew Stone, la réalisation de textes de bonne qualité passe en outre par une interaction étroite entre génération, représentation des connaissances et inférence. On a besoin de cette interaction par exemple : pour vérifier que les présuppositions du texte généré sont satisfaites par le contexte énonciatoire ; pour décider des mots les plus appropriés étant donné ce contexte (problème de la lexicalisation) ; ou pour construire des réponses dites "coopératives" i.e. des réponses qui vont au-delà du simple "oui/non" ou de l'énumération et qui, soit corrigent une erreur de compréhension de l'utilisateur, soit fournissent une réponse partielle lorsqu'une réponse directe à la question posée n'est pas possible.

Cette interaction entre génération et inférence constitue l'un des thèmes de recherche de L&D et est au coeur des préoccupations des deux projets InDiGen et GenI.

Le premier projet (InDiGen) porte (i) sur l'utilisation des contraintes pour la génération de groupes nominaux définis et (ii) sur l'utilisation de démonstrateurs de théorèmes pour la génération d'anaphores associatives. Il a donné lieu au développement d'un générateur (implémenté dans le langage de programmation par contraintes MoZart) qui interface un planificateur de phrases avec des prouveurs de théorèmes pour la logique de premier ordre (Spass, Otter) et pour les logiques de descriptions (Racer). Le générateur InDiGen est utilisable sur le web à l'url <http://www.coli.uni-sb.de/cl/projects/indigen/demo>.

Le deuxième projet (GenI), qui a débuté en Mars 2002, fait intervenir plusieurs équipes françaises et vise à créer des ressources et une architecture pour la génération de textes français ainsi qu'à tester cette architecture sur les problèmes de la lexicalisation, des réponses coopératives et des anaphores associatives. Plus spécifiquement, le projet vise à :

- spécifier un langage de représentation sémantique pour la langue naturelle dans le cadre des logiques de descriptions et logiques assimilées (en particulier, les logiques hybrides de Patrick Blackburn),
- définir, à partir de l'effort de normalisation et de l'infrastructure mise en place par l'arc RLT pour les grammaires TAG, une grammaire TAG adaptée à la génération c'est-à-dire une grammaire spécifiant à la fois la syntaxe et la sémantique d'un fragment du français,
- examiner la possibilité de réutiliser dans l'algorithme de génération les techniques de tabularisation mises en oeuvre dans DyALog pour les TAGs,

- valider l'approche proposée par un générateur intégrant (i) une base de données lexicales et une ontologie modélisant un domaine restreint dans le cadre des LDs, (ii) des démonstrateurs pour les logiques hybrides et éventuellement (iii) des méthodes tabulaires pour la partie "réalisation de surface ».

4.3. Inférence et discours

Pour simplifier, faire une inférence signifie extraire de l'information implicite à partir d'une information explicite. Il existe différents types d'inférences. Par exemple, l'inférence statistique utilise un ensemble large de schémas trouvés dans des corpus afin de déterminer l'information qu'on aura probablement à propos d'un petit exemple. Un autre type d'inférence est l'inférence logique. Par exemple, si notre base de connaissances contient explicitement que "tous les tueurs à gage sont violents" et que "Vincent est un tueur à gages", alors elle contient implicitement l'information que "Vincent est violent". L'inférence logique et l'inférence statistique (et, en réalité, beaucoup d'autres formes d'inférences) sont pertinentes dans le discours. L'équipe Langue et Dialogue étudie actuellement l'utilisation de l'inférence logique. Ces dernières années, les performances des outils de raisonnement automatisé (c'est à dire des logiciels qui traitent les différents types d'inférence logique) ont considérablement augmenté. Les prouveurs de théorèmes atteignent des niveaux de performance inimaginables il y a dix ans. De plus, la technologie de construction de modèles, bien que moins avancée, a atteint un stade où elle constitue un intéressant outil d'expérimentation. Ces progrès ont été faits pour un grand nombre de formalismes de représentation sémantique utiles comme la logique du premier ordre, les logiques de descriptions ou les logiques hybrides.

L'utilisation la plus évidente de l'inférence logique dans le discours est la désambiguïsation. Les énoncés en langue naturelle sont très fréquemment ambigus. En fait, l'interaction entre l'ambiguïté lexicale, l'ambiguïté syntaxique et l'ambiguïté de portée peut amener une phrase à avoir une centaine d'interprétations possibles, même si la plupart de ces interprétations sont absurdes étant donné nos connaissances encyclopédiques d'arrière-plan. Les démonstrateurs de théorèmes et les générateurs de modèles vérifient les différentes interprétations et éliminent celles qui sont incompatibles avec le reste du discours. Parfois, il est aussi intéressant de tester si la représentation nous apporte quelque chose de vraiment nouveau ou si elle provient de ce que nous savons déjà. Les démonstrateurs de théorèmes et les générateurs de modèles sont aussi utiles pour accomplir ce type de tâche. Les programmes CURT (Blackburn et Bos) constituent une série de programmes qui illustrent ces diverses possibilités.

L&D est aussi concernée par des utilisations plus expérimentales de ces technologies. Par exemple, nous essayons d'utiliser les générateurs de modèles pour "deviner" à quoi ressemble une situation décrite (en générant le plus petit modèle possible pour cette situation). De telles approches constituent une façon intéressante d'une part, de traiter des problèmes comme les inférences mises en jeu dans les anaphores associatives et, d'autre part, d'étudier la logique des interactions dans les dialogues réels [49].

4.4. Dialogue multimodal

La conception de systèmes de dialogues multimodaux s'appuie sur une étude fine des phénomènes de répartition du sens entre langue et geste. Dans le type de dialogue considéré (dialogue de commande), la référence aux objets et aux actions constitue l'élément fondamental de toute interprétation. Par ailleurs, il apparaît clairement que l'on ne peut étudier les actes de référence sans se préoccuper de la manière dont l'utilisateur perçoit visuellement les objets dont il parle. Ceci amène tout naturellement à l'étude conjointe des acteurs que sont : la langue, le geste et la perception visuelle.

La langue : on s'intéresse spécifiquement aux énoncés accompagnés d'un geste de désignation et plus précisément aux expressions référentielles servant de support au geste de désignation.

Le geste : l'objectif est de pouvoir à partir de données fournies par un dispositif gestuel (ex. souris, écran tactile, interface haptique) de reconstituer une trajectoire à partir de laquelle les objets susceptibles d'être désignés sont mis en évidence.

La perception visuelle : on s'intéresse ici à la configuration spatiale des objets affichés sur une scène. Il est évident que cette dernière conditionne la forme de la trajectoire mais aussi le choix de l'expression référentielle.

L'étude de ces trois acteurs fournit des informations essentielles quant au rôle de chacun d'entre eux dans l'acte de co-référence, informations à partir desquelles il est possible de proposer une modélisation du processus d'interprétation menant de l'énoncé multimodal à l'identification des référents (résolution des références).

Depuis, environ une dizaine d'année, nous menons des travaux en ce sens (thèse de Duermel, Bellalem, Schang et Wolff) qui ont permis d'apporter des avancées à différents niveaux de conception des systèmes de dialogues multimodaux. Aujourd'hui, nos priorités de recherche peuvent se décliner comme suit :

- Description générale du geste : il s'agit de proposer une modélisation du geste indépendante du dispositif c'est-à-dire de mettre en évidence les mécanismes fondamentaux qui régissent la forme des trajectoires en fonction du contexte visuel.
- Sémantique formelle : l'objectif est ici de proposer une modélisation abstraite des contraintes issues des différents acteurs intervenant dans l'acte de co-référence :
 - Focalisation du geste : Le geste de désignation a pour objet de fournir une focalisation sur un sous-espace de l'espace visuel partagé entre l'utilisateur et la machine. Cette focalisation prend appui sur la forme de la trajectoire et sur les caractéristiques de la configuration spatiale ;
 - Structuration de l'espace visuel : Pour interpréter un geste, il est indispensable de disposer d'une modélisation de ce que perçoit l'utilisateur de la scène affichée sur l'écran de la machine. En effet, un geste vague ne pourra être compris que si cette imprécision est compensée soit par une information précise de l'énoncé langagier qui l'accompagne soit par des caractéristiques de saillance forte de la scène.
 - Classification fournie par la langue : Le type des référents est souvent contenu dans l'expression référentielle langagière et donne un élément de filtrage à appliquer sur les objets candidats à la référence. Quand ce n'est pas le cas, cette information peut aussi être recherchée au niveau du prédicat de l'énoncé.
- Intégration sémantique des modalités : Disposant des différentes contraintes énoncées précédemment (contraintes de focalisation, contraintes de saillance, contraintes de type) se pose la question de déterminer le principe qui va guider l'intégration de toutes ces contraintes pour aboutir à l'interprétation de l'énoncé multimodal. Ainsi, nous nous appuyons sur la théorie de la pertinence [48] qui fournit des heuristiques permettant de réduire l'espace de recherche aux éléments les plus pertinents par rapport au contexte d'énonciation.
- Architecture : L'implantation d'un système de dialogue multimodal repose sur un ensemble de modules dédiés à la réalisation de tâches spécifiques (tels que traitement des entrées : parole, geste, gestion du dialogue) mais qui sont néanmoins coopératifs. Cette coopération suppose la circulation d'un flux d'information qui pour être possible doit s'appuyer sur un langage homogène d'interfaçage entre les modules tel que le langage MMIL (cf. 6.3).

4.5. Ressources linguistiques et multimédias

Gérer des ressources linguistiques normalisées représente pour nous à la fois un moyen de valider, par l'observation, nos modèles théoriques et une source générique d'information (par exemple lexicale) pour les prototypes que nous développons. Plus globalement, nous pensons qu'il est nécessaire, d'une part de contribuer activement à la définition de normes plus approfondies dans le domaine de l'ingénierie linguistique et, d'autre part de participer à la dissémination des cadres normatifs existants. C'est à ce titre que projet Langue

et Dialogue occupe ainsi une place active au sein de la communauté nationale et internationale dans le domaine de la normalisation des ressources linguistiques et de leur utilisation.

Par ailleurs, les méthodes développées au sein du projet, notamment dans le cadre de la définition de la norme ISO 16642 (qui a obtenu 100% de vote favorable au stade DIS qui est actuellement en cours de publication à l'ISO), nous semble une base pour l'étude de classes de documents semi structurés possédant des propriétés similaires. Parmi les points que nous développons plus particulièrement, nous pouvons mentionner :

- Développement d'un corpus de dialogue homme-machine complètement normalisé et libre d'accès (projet ASILA) ;
- Mise au point de méthodes reproductibles d'annotation et d'évaluation de la référence sur des corpus de dialogue et de textes (projet ANANAS) ;
- Développement d'un modèle générique de représentation des contenus multimodaux pour le dialogue homme-machine, avec l'ambition d'en faire une contribution normative internationale ;
- Contribution à la coordination des activités internationales de normalisation au sein du consortium TEI et du comité TC37/SC4 de normalisation ;
- Définition d'un modèle générique de spécification de formats XML permettant de spécifier à la volée différents types d'annotations linguistiques, mais aussi culturelles (des expérimentations ont été effectuées sur des descripteurs cinématographiques (projet CINERGIE) et muséographique (collaboration avec Taiwan)).

4.6. Recherche d'information dans les sources de données génomiques

Il s'agit de développer des systèmes visant à assister, évaluer et structurer la collecte des données sur le génome, en vue d'une meilleure exploitation de la masse d'informations disponibles par le Web dans les domaines pharmaceutiques ou médicaux. Deux champs d'applications ont été explorés : la cartographie des gènes humains en relation avec les maladies génétiques et la recherche dans le génome humain de gènes candidats pour des pathologies complexes. Cette activité constitue l'une des opérations du PRST Intelligence Logicielle de Lorraine, thème : Bioinformatique et Applications à la génomique.

Peu de systèmes aujourd'hui prennent en compte les problèmes liés à la collecte des données biologiques (multiplicité des ressources, hétérogénéité et variabilité des formats, mises-à-jour inégales, redondance des nomenclatures, insuffisance de la documentation sur les données...) ou à leur interprétation (recherche d'informations complémentaires, mise en œuvre d'outils d'analyse, intégration de l'ensemble des résultats...). Des solutions sont proposées à partir de méthodes issues du domaine de la recherche d'information et en s'appuyant sur l'utilisation de documents semi-structurés.

Dans chaque domaine d'application, des partenaires biologistes (CNRS FRE2571 et UHP UPRES 3446) ont été amenés à modéliser le scénario de recherche d'information, jusqu'à ce qu'il soit compréhensible et exploitable par les informaticiens. Ce travail sert de base à la conception de systèmes de recherche d'information appropriés. Cette phase de conception a été suivie du développement d'une application interactive, qui a fait l'objet d'une expérimentation par les partenaires biologistes. En retour, diverses modifications ont été introduites pour optimiser l'application et de nouvelles orientations ont vu le jour : création d'une base de données pour stocker les sessions et exploiter les données collectées, automatisation du processus pour traiter un grand nombre d'entrées, extension à des problèmes voisins [46]. De façon plus générale, ce travail nous a conduits à proposer une approche de médiation pour l'inventaire et la navigation à travers les sources de données [45] [15].

5. Logiciels

5.1. Liste des logiciels

Au cours de cette année, l'équipe L&D a développé, amélioré ou tout simplement maintenu un ensemble de logiciels dont la liste suit :

- Logiciels SOAPical

- TAXI : The Adjustable XML edItor
- API Feature Structure
- Analyseur LTAG
- Compilateur de Méta-grammaire
- CURT
- Indigen
- Plate-forme CINERGIE

Les paragraphes suivants décrivent brièvement ces logiciels.

5.2. SOAPical

Contact : Eric Kow et Etienne Petitjean

Site web : <http://soapical.sourceforge.net>

5.2.1. Soapmeter

C'est le visualisateur de l'ensemble des messages SOAP qui passent entre deux services. Il présente à l'utilisateur l'historique de la communication, et offre la possibilité d'examiner les messages XML au format texte ou graphique adapté à la nature du message (TaxiService).

5.2.2. Soapcaster

Dans les architectures flexibles, un médiateur, dont le rôle consiste à faciliter la gestion des communications entre plusieurs composants web-services, est souvent nécessaire.

Soapcaster offre cette fonction de façon simple.

Il s'agit d'un manager central sous forme d'une librairie réutilisable par chaque service. Son rôle est de maintenir une table des services qui y sont inscrits, de permettre les services à s'y inscrire et s'y désinscrire et de régulièrement vérifier si ces services sont toujours disponibles. Cette table est diffusée aux clients et mise à jour à intervalles réguliers. En utilisant la librairie soapcaster, les services peuvent diffuser des messages aux autres services, même s'il arrive que le serveur tombe en panne.

Soapcaster est actuellement distribué avec un exemple d'application dans le domaine de la télémédecine.

5.3. TAXI : The Adjustable XML edItor

Contact : Eric Kow

TAXI est un outil qui permet de définir un éditeur graphique de données semi-structurées adaptable à la nature de l'information éditée. Pour être traité par TAXI, l'information doit être représenté en XML et valide par rapport à une grammaire (DTD ou Schema XML). L'affichage graphique s'appuie sur le format SVG.

L'objectif est d'offrir une visualisation graphique conforme à la nature des objets édités (exemple : arbres syntaxiques, grammaire, structures de traits). On permet, de cette manière, à l'utilisateur de se focaliser uniquement sur l'information et de s'affranchir de coder directement du code XML.

Après définition des paramètres d'adaptation graphique, TAXI peut être utilisé de manière autonome ou comme API à intégrer dans d'autres applications. Il existe également une version utilisable en ligne (web : TaxiService).

5.4. API Feature Structure

Contact : Bertrand Gaiffe

Site web : <http://www.loria.fr/equipes/led/outils.html>

Il s'agit d'une API de représentation et de traitements de structures de traits (FS). Les principaux traitements sont l'unification et la subsomption. La principale nouveauté consiste en l'extension du formalisme aux FS disjonctifs sur les valeurs. Cet API est distribué sous forme d'un package JAVA.

5.5. Analyseur LTAG

Contact : Azim Roussanaly

Site web : <http://www.loria.fr/equipes/led/outils.html>

Une nouvelle version de l'analyseur LTAG est en cours de développement. Les améliorations portent essentiellement sur :

- une meilleure indépendance entre l'analyseur, les ressources, les données et les traitements annexes,
- l'intégration du format TAGML2 pour représenter les ressources (lexique, syntaxe),
- l'intégration d'outils de visualisation et de mise au point d'une grammaire,
- la possibilité d'analyser des phrases partiellement prétraitées,
- la possibilité de traiter des phrases segmentées de manière ambiguë. (plusieurs candidats pour un groupe de mots)

5.6. Compilateur de Méta-grammaire

Contact : Bertrand Gaiffe

Site web : <http://www.loria.fr/equipes/led/outils.html>

Ce logiciel constitue un environnement de gestion de méta-grammaires : il comprend un éditeur spécifique (avec possibilité de visualisation graphique) qui produit une description de méta-grammaire au format XML et d'un compilateur qui construit des arbres TAG au format TAGML.

5.7. CURT

Contact : Patrick Blackburn

Site web : <http://www.loria.fr/equipes/led/outils.html>

Les programmes de CURT combinent construction sémantique et inférence pour les langues naturelles. Ces programmes sont implémentés en PROLOG et fonctionnent sous SICSTUS Prolog et SWI Prolog. Perl, Otter et Mace sont également requis.

5.8. Indigen

Contact : Claire Gardent

Site web : <http://www.coli.uni-sb.de/cl/projects/indigen/demo/>

Le générateur InDiGen adopte les principes de base exemplifiés par le système SPUD de Matthew Stone : comme lui, il permet une forte interaction entre planification de phrases et inférence. Les raisonneurs automatiques utilisés par IndiGen sont : un générateur de modèles (très simple i.e., non optimisé) pour la logique du premier ordre implémenté en Oz, les démonstrateurs de théorèmes pour la logique du premier ordre SPASS et Otter ainsi que les démonstrateurs de théorèmes pour la logique de description RACE et FaCT.

Le générateur InDiGen est implémenté en Mozart Oz

5.9. plate-forme CINERGIE

Contact : Jean-Luc Husson

Site web : <http://www.loria.fr/projets/CINERGIE>

Cette plate-forme permet la gestion et la diffusion de documents multilingues et multimédia.

Tous les composants logiciels sont développés. Dans cette ultime phase du projet, le consortium engage une phase de déploiement, d'enrichissement des bases et d'évaluation.

Parallèlement, les partenaires travaillent à la valorisation (extension des données) et à la diffusion des résultats du projet sous la forme de publications scientifiques et de supports publicitaires.

- Niveau réseau
 - Outils d'administration du réseau de serveurs délocalisés

- Outil de consultation multi-critères des données déclarées publiques du réseau. Cet outil est générique et paramétré par la spécification du modèle abstrait CINERGIE.
- Niveau serveur
 - Administration des utilisateurs, de leurs responsabilités et de leurs droits d'accès aux données.
 - Administration des bases (création du modèle restreint -sous-ensemble du modèle de référence-, construction du comité éditorial, gestion des produits dérivés)
 - Edition de données. Cet outil est générique car totalement paramétré par la spécification du modèle abstrait CINERGIE.

5.10. XMAP

Contact : Marie-Dominique Devignes

Site web : <http://www.loria.fr/projets/Xmap/Index.htm>

5.10.1. Xmap_INTERACTIVE

Assistant pour l'interrogation des sources génomiques, le filtrage des résultats et l'intégration des données collectées. Vise à répondre au problème de la co-localisation sur le génome humain des nouveaux gènes découverts par le séquençage avec les pathologies orphelines (dont on ne connaît pas le gène responsable).

5.10.2. Xmap_AUTO

Version automatisée de Xmap_INTERACTIVE.

Version : 4

Site web : prévu sur <http://bioinfo.loria.fr>

5.10.3. Xmap_DB

Base des données collectées par Xmap_AUTO (gérée sous Postgresql) ; import/export des données au format XML.

Site web : prévu sur <http://bioinfo.loria.fr>

5.10.4. Xmap_SHOW

Outil pour la visualisation des résultats de session Xmap. Convertit le document de session XML en document SVG.

Site web : prévu sur <http://bioinfo.loria.fr>

5.10.5. Xprom

Application basée sur la modélisation générique d'un scénario de recherche d'information et permettant l'exécution de ce scénario. Pour chaque source les fonctionnalités suivantes sont mises en œuvre successivement : formulation des requêtes, soumission des requêtes, filtrage des données retournées, stockage dans un document XML.

6. Résultats nouveaux

6.1. Logique Hybride

Un thème de recherche important pour l'équipe L&D est le développement de formalismes de représentation adaptés à des applications de traitement de la langue naturelle.

L'équipe s'intéresse particulièrement aux logiques de description, aux logiques modales et aux logiques hybrides, toutes trois fortement apparentées. En particulier, nous nous intéressons à l'utilisation des logiques modales pour raisonner sur les structures syntaxiques, à l'emploi des logiques de description pour la représentation sémantique (tout particulièrement pour la génération) et nous conduisons des recherches théoriques quant aux propriétés formelles des logiques hybrides.

Nos travaux confirment que les logiques hybrides partagent avec les logiques classiques plus de bonnes propriétés que les simples logiques modales. Un premier travail [24] montre que les résultats de complétude les plus généraux relatifs aux logiques hybrides peuvent être établis de façon constructive par une méthode de traduction systématique de tableaux classiques en des tableaux hybrides. Un second résultat [13] est la preuve constructive d'interpolants pour de larges classes de logiques hybrides. Ce résultat est probablement le plus important résultat positif d'interpolation dans les logiques modales.

Nous avons donc maintenant confirmation de la possibilité de construire des systèmes automatiques de preuve efficaces pour les logiques hybrides. Nos recherches futures iront probablement vers l'implémentation de tels prouveurs de théorèmes et à leur utilisation pour le traitement de la sémantique des langues.

Les recherches qui ont mené aux résultats précédents ont été conduites en collaboration entre l'équipe LIT (Language and Inference Technology, ILLC, University of Amsterdam) et L&D, grâce au programme d'"Equipes associées" de l'INRIA (<http://www.loria.fr/projets/ledlit/>).

6.2. Génération

La génération d'expressions référentielles est une sous-tâche importante de la tâche de génération. Elle vise en particulier à

déterminer le contenu sémantique nécessaire pour une bonne identification de l'objet désigné.

Partant de l'algorithme standard proposé par Dale et Reiter pour la génération de descriptions définies (e.g., "le chat", "l'homme avec un chapeau noir"), nous avons étendu cet algorithme dans deux directions, l'une utilisant les contraintes et l'autre l'inférence.

Les contraintes ensemblistes permettent en effet une implémentation directe des contraintes régissant la génération des GNs définis. Etant donné un ensemble d'entités E, un ensemble de propriétés P pour une entité donnée x permet d'identifier cette entité de façon unique dans cet ensemble (et partant de produire une description définie) ssi :

- P est un ensemble de propriétés vraies de x et
- pour chaque élément y de E, il y a au moins une propriété dans P qui n'est pas vraie de y

Un résolveur de contraintes basé sur cette idée a été implémenté en Oz (<http://www.mozart.org>) et [33] montre que ce résolveur permet une généralisation de l'algorithme standard pour la génération de GN définis qui permet de l'étendre aux GNs pluriels (e.g., les chats) ainsi qu'aux descriptions disjonctives (les chats et les chiens) et négatives (les chats qui ne sont pas blancs). [21] compare par ailleurs cet algorithme avec l'algorithme incrémental standard pour la génération de descriptions définies et montre que la performance de l'algorithme proposé est meilleure que celle de l'algorithme incrémental.

L'inférence de son côté est nécessaire pour raisonner sur les implications du contexte d'énonciation et du texte construit et plus généralement pour permettre "l'économie textuelle". Par exemple, dans le texte "J'ai vu une belle maison hier soir. La porte était en ivoire.", "la porte" est plus économique que son équivalent sémantique "la porte de cette maison". [47] montre comment inférence et réalisation de surface (la construction de l'arbre syntaxique associé à la sémantique donnée en entrée) peuvent être combinées pour générer de telles anaphores associatives.

6.3. MMIL, un langage de représentation de contenus multimodaux

Dans le cadre du projet MIAMM, nous avons développé le langage MMIL (Multimodal Interface Language), qui permet à différents composants d'une architecture de système de dialogue, d'échanger des informations de manière uniforme. Ce langage, implémenté sous la forme de schéma XML, a été défini en reprenant la méthodologie de spécification introduite dans le cadre des travaux plus généraux menés sur la modélisation de ressources linguistiques. Il permet aussi bien de représenter les structures de dépendances issues de l'analyse syntaxique d'énoncés en langue naturelle que la description de configuration haptiques et graphiques ou encore les requêtes destinés au modèle du domaine. Ces travaux s'intègrent dans une initiative internationale plus large qui, sous l'égide du groupe SIGDIAL de l'ACL (Association for Computational Linguistic), lancée par

Harry Bunt (Univ. de Tilburg), Kiyong Lee (KAIST) et Laurent Romary, vise à définir un cadre général de représentation des contenus multimodaux. La figure 1 montre un exemple typique d'une telle représentation utilisé en sortie d'un analyseur syntaxico-sémantique.

```
<mmilComponent>
  <event id="e1">
    <evtType>Love</evtType>
    <mode>Indicative</mode>
    <tense>Present</tense>
  </event>
  <participant id="indiv1">
    <individuation>singular</individuation>
    <objectType>Human</objectType>
    <refType>Definite</refType>
    <lex>Jones</lex>
    <refStatus>Pending</refStatus>
  </participant>
  <participant id="indiv2">
    <objectType>Human</objectType>
    <refType>Definite</refType>
    <refStatus>Pending</refStatus>
  </participant>
  <relation type="subject" source="indiv1" target="e1"/>
  <relation type="object" source="indiv2" target="e1"/>
</mmilComponent>
```

Figure 1 : Représentation en MMIL de l'énoncé "John loves Mary"

8. Actions régionales, nationales et internationales

8.1. Actions industrielles

8.1.1. EADS/S&DE(Vulcain)

Le travail effectué dans le projet Vulcain a consisté à valider les acquis de la thèse de Patrice Lopez en explorant les possibilité d'utiliser des techniques d'analyse syntaxique partielle à base de grammaires d'arbres adjoints pour la recherche de motifs pertinents à l'intérieur de flux de messages Internet. Les résultats ont permis de valider l'approche sur un corpus de référence fourni par EADS/S&DE

8.2. Actions régionales

8.2.1. Diatelic

Le projet Diatelic est une action transversale du Loria faisant intervenir les équipes Maia, Trio et Langue et Dialogue. Lancée initialement (1996) sous l'égide du programme « habitat intelligent » du département SPI du CNRS, ce projet associe des médecins de l'Altir (Association Lorraine de Traitement de l'Insuffisance Rénale) ainsi qu'un médecin conseil. Il vise à mettre en œuvre une plate-forme de télésurveillance de malades dialysés à domicile permettant de contrôler quotidiennement leur état général et de déclencher si besoin est des alertes préventives de possibles pathologies. En 1999, bénéficiant du soutien du pôle européen de santé, une expérimentation médicale a été lancée pour comparer une population de patients équipés du système que nous avons implanté et une population de référence.

8.2.2. *Projet « Ingénierie des langues, du document et de l'Information Scientifique, Technique et Culturelle » dans le cadre du plan État-Région Lorraine*

Les objectifs que nous nous fixons dans le cadre de ce projet centré sur l'ingénierie des langues, du document et de l'information scientifique, technique et culturelle sont :

- Conforter et créer des synergies entre la recherche (LORIA, INIST, ATILF (ex-INALF), et laboratoires universitaires), les offreurs de contenus (Universités, BRGMS), les développeurs de solutions (INIST, Berger-Levrault), les formations initiales (universités de Nancy et de Metz) et les utilisateurs : le monde économique et celui de la formation professionnelle ;
- Structurer au niveau régional un centre de compétence, pour proposer et diffuser des solutions techniques innovantes et favoriser la valorisation des résultats de recherche ;
- Conforter des activités économiques en ce domaine en donnant des atouts supplémentaires aux industries lorraines dans le cadre de la société de l'information.

Concrètement, une action de mise en place d'une double plate-forme informatique de gestion, de distribution et d'accès à des informations textuelles est en cours ;

- l'une orientée vers l'Information Scientifique et Technique,
- l'autre orientée vers la gestion de ressources linguistiques (textes annotés, lexiques, dictionnaires, outils d'ingénierie linguistique).

Une partie des outils et des logiciels de base serait, bien entendu, commune à ces deux plates-formes. Pour permettre une intégration aisée des divers composants logiciels, il convient de s'appuyer sur des formats d'échange entre les divers composants fondés sur des normes clairement définies (XML, TEI, etc.) et d'intégrer au mieux les compétences des diverses équipes participantes.

8.3. Actions nationales

8.3.1. CINERGIE

Cadre : CINERGIE est un projet national du programme PRIAMM (Programme pour la recherche et l'innovation dans l'audiovisuel et le multimedia) soutenu par le secrétariat d'Etat à l'Industrie et le ministère de la Culture et de la Communication (CNC). D'une durée de 18 mois, le projet a débuté en septembre 2001 et s'achève en février 2003.

Objectif : l'objectif est de développer une plate-forme logicielle de gestion collégiale et de diffusion de ressources cinématographiques à la fois multilingues et multimedia particulièrement riches en Europe.

Une gestion collégiale :

CINERGIE offre un espace virtuel commun sur Internet permettant à un ensemble d'experts (professionnels du cinéma, archivistes, traducteurs, terminologues) de travailler de manière coopérative et d'enrichir ensemble des bases réparties de ressources cinématographiques européennes. La plate-forme CINERGIE repose sur une architecture réseau répartie, interconnectant plusieurs serveurs de données gérés en autonomie. Une typologie spécifique d'utilisateurs et des scénarios robustes d'interaction garantissent la cohérence des données dans les processus d'édition.

Des services innovants : CINERGIE propose un ensemble de services à forte valeur ajoutée pour les utilisateurs finaux du système grâce à l'exploitation d'un thésaurus spécifique, la généralisation du multilinguisme (accès aux traductions du contenu, intégration d'une terminologie multilingue du domaine). L'exploitation de la spécification Dublin Core permet une description fine des documents multimédias référencés dans le réseau. Un outil de consultation dédié permet de construire une requête résultant de la combinaison booléenne de divers critères de recherche. La possibilité de croiser un nombre illimité de critères et l'accès à l'intégralité des champs du modèle de données, dans tout le réseau, permet une recherche d'informations extrêmement puissante. En outre, de nombreux produits dérivés allant du site Internet aux bornes interactives pourront être aisément déclinés à partir d'un format unique.

Les apports de la standardisation : la dynamique internationale en faveur de la standardisation permet de résoudre conjointement les problèmes soulevés par l'extrême fragmentation linguistique de l'Europe, l'hétérogénéité des types de ressources multimédias, le cloisonnement des fonds, les barrières linguistiques. En particulier, les normes en matière de structuration, de codage, de traitement et de diffusion de l'information sur Internet (XML et les directives connexes du consortium W3C) sont autant d'avancées technologiques majeures que CINERGIE met aujourd'hui au service du septième art. L'expérience acquise par l'équipe L&D au fil de divers projets internationaux (ELAN, DHYDRO, SALT) sur les sujets des architectures distribuées pour l'édition collaborative, de la gestion du multilinguisme et de la structuration et du codage de modèles de données a notamment débouché sur l'élaboration d'un modèle abstrait flexible et à très large couverture (combinaison d'un meta-modèle et d'un registre d'environ 160 catégories de données). La collaboration avec la société Lucid-IT a permis de mettre au point un système de Bases de données XML native dont les fonctionnalités réunissent les possibilités des SGBD classiques et des systèmes d'indexation plein texte, exploitant ainsi toute la puissance liée au caractère semi-structuré des données XML.

Consortium :

- Laboratoire informatique : INRIA, équipe L&D
- Laboratoire en sciences de l'information : CERIME (Centre d'Etudes et de Recherches Interdisciplinaires sur les Médias en Europe, Université Robert Schuman)
- Société privée : LucidIT SA

Site Web : <http://www.loria.fr/projets/CINERGIE>

8.3.2. ARC RLT

Cadre : action de recherche concertée de l'INRIA

Partenaires : Projet ATOLL (INRIA Rocquencourt), Loria-L&D (Nancy), Loria-Calligramme (Nancy), TALANA (Paris 7)

Page web : <http://atoll.inria.fr/RLT/arc.html>

Contact : Bertrand Gaiffe (correspondant de l'ARC pour Langue et Dialogue)

Description : Cette action a pour objectif d'examiner les problèmes d'acquisition et de représentation des ressources dans le cadre des grammaires d'arbres adjoints (TAG - Tree Adjoining Grammars) dans un contexte de grammaires à large couverture linguistique et des lexiques associés à ces grammaires.

8.3.3. ARC GENI

Cadre : action de recherche concertée de l'INRIA

Thème : Génération, Inférence et Logique des Descriptions

Durée : Mars 2002 - Mars 2004

Partenaires : ATOLL (INRIA Rocquencourt), ILPL (IRIT), Lattice (Université Paris VII), Orpailleur (INRIA Lorraine)

Page web : <http://www.loria.fr/projets/geni/>

Contact : Claire Gardent (Responsable de l'action)

Description : L'action de recherche concertée GenI poursuit deux objectifs. D'une part, elle vise à développer les ressources et l'architecture nécessaire à une interaction étroite entre la planification de phrases et l'inférence dans un système de génération de textes pour la langue naturelle. D'autre part, elle cherche à déterminer jusqu'à quel point les logiques de descriptions et logiques hybrides peuvent être utiles pour représenter le sens de la langue naturelle et tirer les inférences nécessaires pour la production de textes de bonne qualité.

8.3.4. ANANAS

Cadre : projet soutenu par le Programme Interdisciplinaire « Société de l'Information » du CNRS

Thème : création de corpus annotés pour la référence

Partenaires : ATILF, LORIA, Grenoble III, ISSCO (Genève)

Durée : 01/01/2002-01/01/2005

Contact : Susanne Salmon-Alt (responsable de l'action)

Description : Les corpus français actuellement disponibles pour valider des modèles linguistiques ou des algorithmes de résolution anaphorique sont insuffisants, à la fois quantitativement et qualitativement. Partant de ce constat, ANANAS vise la création d'une base de corpus français annotés en relations anaphoriques qui soit libre d'accès, codée selon un schéma normalisé et qui intègre, si possible, les ressources déjà pré-annotées à d'autres niveaux linguistiques.

8.3.5. GDR Sémantique et Modélisation : Groupement de recherche 2521 (CNRS)

Partenaire(s) :

- Langue et Dialogue (L&D), Loria,
- Institut Jean Nicod, Paris,
- Institut des Sciences Cognitives de Lyon,
- ATILGF, Université de Paris VII
- Logique Interaction langue Computation, Toulouse

Cadre : Sémantique de la langue naturelle.

Période concernée : 2002-2005

Thème : Modèles et formats de représentation pour la sémantique, les modèles à l'épreuve des données, les interfaces de la sémantique linguistique, sémantique computationnelle, sémantique et corpus.

Contact dans l'équipe : Patrick Blackburn

Le GDR comprend onze groupes de travail :

- Anaphores
- Déterminants
- Dialogue
- Logiques de description
- Pragmatique - Sémantique
- Présuppositions
- Psychologie - Sémantique
- Sémantique lexicale - Corpus
- Syntaxe - Sémantique
- Temps - Aspect
- Traductibilité - Expressibilité

Le responsable pour le groupe Logiques de description est Patrick Blackburn.

Page Web : <http://semantique.free.fr>

8.4. Actions européennes

8.4.1. Université d'Amsterdam

Un financement de l'INRIA, pour une durée de 3 ans, en vue de la mise en place d'une équipe associée avec le groupe « Computational and Applied Logic Group » (CALG) de l'Université d'Amsterdam a été obtenue cette année. CALG est le plus grand des 5 groupes qui constituent l'ILLC de l'Université d'Amsterdam. L'objectif de cette association est de mettre en commun l'expertise de L&D et de CALG dans le domaine des logiques pour la représentation des connaissances, les méthodes d'inférence et l'informatique linguistique en vue d'une application aux les problèmes de traitement de langues naturels, avec un intérêt particulier pour l'utilisation des méthodes tabulaires (<http://www.loria.fr/projets/ledcalg/>).

8.4.2. Projet Indigen

Des échanges réguliers ont lieu avec le Département de Linguistique Informatique de l'Université de Sarrebruck (Allemagne) dans le cadre de la cotutelle de la thèse franco-allemande de Kristina Streignitz (Directeurs de thèse : Manfred Pinkal (Sarrebruck), Claire Gardent (Nancy) et Jean-Marie Pierrel (Nancy)) ; ainsi

que dans le cadre du projet InDiGen (Integrierte Diskursgenerierung), projet financé par le DFG (Deutsche Forschungsgemeinschaft) mais dont le responsable scientifique (Claire Gardent) est à Nancy.

8.4.3. *Projet MIAMM*

Cadre : Projet européen dans le cadre du programme IST, coordonné par le Loria

Partenaires : Loria (équipes Langue et Dialogue et Parole, DFKI (DE), SONY (DE), Canon (UK), TNO (NL))

Thème : Spécification et mise en oeuvre de systèmes de dialogue multimodaux comportant une composante haptique.

Période : sept 2001- fev 2004

Page web : <http://www.miamm.org>

Contact : Laurent Romary

Participants : Bertrand Gaiffe, Laurent Romary, Amalia Todirascu, Benoit Crabbé, Ashwani Kumar, Eric Kow, Azim Roussanaly

Description : L'équipe Langue et Dialogue est chargée de différentes actions au sein de ce projet :

- réalisation du module d'analyse syntaxique pour le français. Ce travail reprend les acquis de l'équipe autour de l'analyse à base de grammaires d'arbres adjoints et fournit une sortie exprimée en logiques de description ;
- réalisation du module de fusion d'informations multimodales permettant de fournir une interprétation en contexte des énoncés (parole et geste) de l'utilisateur ;
- définition du langage MMIL (Multimodal Interface Language) dans lequel sont exprimés l'ensemble des échanges d'information à l'intérieur de l'architecture du système MIAMM ;
- intégration logicielle de l'architecture.

Ce projet est considéré comme stratégique pour l'équipe, puisqu'il permet d'une part d'expérimenter les différents modèles développés dans l'équipe au cours de ces dernières années, et, d'autre part, de mettre en place une collaboration étroite avec l'équipe Parole, responsable des aspects reconnaissance au sein du projet.

8.4.4. *Projet OZONE*

Cadre : Projet européen dans le cadre du programme IST

Partenaires : INRIA (plusieurs équipes, dont Langue et Dialogue, MAIA et Parole, Philips (NL, coordinateur), Interuniversity Micro Electronics Center (BE), Epictoid (NL), Technical University Eindhoven (NL), and Thomson Multimedia (FR))

Thème : Mise en place d'un système de dialogue homme-machine multilingue et indépendant de l'application envisagé.

Période : oct 2001- oct 2003

Page web : <http://www.extra.research.philips.com/euprojects/ozone/>

Contact : Laurent Romary

Participants : Bertrand Gaiffe, Laurent Romary

Description : L'équipe Langue et Dialogue intervient dans la définition du système de communication homme-machine au sein de l'architecture OZONE. Le travail effectué au cours de l'année 2002 a consisté à spécifier les différentes composantes de cette architecture.

8.5. Actions internationales

8.5.1. *Common Ref*

En vue d'analyse de corpus anglais, portugais et français et de développement d'un système de compréhension automatique des descriptions définies démonstratives, nous collaborons avec l'Université UNISINOS São Leopoldo (Brésil) dans le cadre du projet « CommonRefs ». Ce projet est co-financé par l'INRIA et le CNPq-ProtémCC pour une durée de 2 ans depuis septembre 2001.

8.5.2. Vietnam

La thèse de Thi Min Huyen, soutenue par l'A.U.F. (Agence Universitaire de la Francophonie), est l'occasion pour nous d'entamer une collaboration avec l'université des sciences naturelles de Hanoï sur la définition d'outils d'ingénierie linguistique communs pour le Français et le Vietnamien.

8.5.3. Taiwan

Une convention de collaboration avec la National Chi Nan University (Taiwan) a été récemment mise en place. Cette collaboration porte sur le développement d'environnements de gestion de « musée numérique » fondé sur la technologie XML.

8.5.4. ISO

9. Diffusion des résultats

9.1. Animation de la communauté scientifique

Le projet Langue et Dialogue est fortement impliqué dans l'animation de la communauté scientifique dans les domaines qui sont les siens. Cela se traduit en particulier par :

9.1.1. Direction

- Co-directeur de l'UMR ATILF/Université Nancy 2 (J-M. Pierrel)

9.1.2. Présidence

- Président du comité TC37/SC4 de l'ISO (L. Romary)
- Vice-Présidente du bureau de l'EACL (European Chapter of the Association for Computational Linguistics) (C. Gardent)
- Président, SIGSEM, ACL Special Interest Group in Computational Semantics (P. Blackburn)

9.1.3. Conseils, Comités, Commissions, Missions...

- Membre du conseil scientifique de la TEI (L. Romary)
- Membre du bureau de l'EACL (European Chapter of the Association for Computational Linguistics) (C. Gardent)
- Membre du conseil scientifique de l'ACI « Travail » du MENRT (J-M. Pierrel)
- Membre nommé au groupe de réflexion sur le traitement informatique du langage auprès du Conseil Consultatif de la Langue Française (J-M. Pierrel)
- Membre du conseil de l'Ecole doctorale de Linguistique de l'Université de Paris 7 (J-M. Pierrel)
- Membre du Conseil Scientifique du Pôle de Recherche Scientifique et Technologique (PRST) Intelligence Logicielle (J-M. Pierrel)
- Membre du conseil scientifique de l'ACI GRID « Grille de ressources informatiques et de données » (J-M. Pierrel)
- Membre du Comité de Pilotage du Réseau Thématique Pluridisciplinaire (RTP) " Communication et Dialogue " du Département STIC CNRS (J.M. Pierrel)
- Membre élu du Conseil scientifique de l'UHP, Université Nancy 1 (J-M. Pierrel)
- Membre élu du Conseil scientifique de l'Université Nancy 2 (S. Cruzlara, D. Coulon)
- Membre élu du Conseil des Etudes et de la Vie Universitaire de l'INPL (C. Fay-Varnier)
- Membre des Commissions de spécialistes des 3 Universités de Nancy (J-M. Pierrel, S. Cruz-Lara, C. Fay-Varnier, D. Coulon)
- Membre du comité de recrutement DR2 pour l'INRIA/Lorraine (C. Gardent)
- Membre élu du conseil d'administration de l'IUT de St Dié (J-L. Husson)
- Membre de la commission de choix de l'IUT de St Dié (J-L. Husson, J.M. Pierrel)
- Membre coopté du conseil scientifique et technique de " Nancy 2005 " (J.M. Pierrel)

- Membre de la commission locale détachement/délégation, INRIA Lorraine (C. Gardent)
- Membre du Comité de rédaction de « la lettre du LORIA » (C. Gardent)
- Chargée de mission pour le suivi des Affaires sociales de l'INPL (C. Fay-Varnier)
- Chargée de mission auprès du service scolarité de l'IUT Nancy-Charlemagne (N.Bellalem)
- Responsable de la formation pour le CNRS et L'INRIA au Loria (L. Romary)

9.1.4. Responsabilités et coordinations scientifiques

- Responsable du projet « Ingénierie des langues, du document et de l'information scientifique, technique et culturelle » dans le cadre du CPER Lorrain (J-M. Pierrel)
- Responsable de l'Action de Recherche Concertée INRIA GenI (Génération et Inférence) (C. Gardent)
- Responsable du projet Franco-Allemand InDiGen. En coopération avec le Department of Computational Linguistics, Universitaet des Saarlandes, Saarbruecken(C. Gardent)
- Responsable Scientifique du Projet "Inference et Generation automatique de la Langue Naturelle". Region Lorraine (C. Gardent)

9.1.5. Expertises

- ACI " jeunes chercheurs " Ministère de la Recherche 2002 (J.M. Pierrel)
- Expert auprès de l'AFNOR X03A-GE1 pour le redéfinition de la norme ISO 1951 " Présentation/Représentation des articles dans les dictionnaires "(J.M. Pierrel)
- ACI " Cognitique " 2002 (J.M. Pierrel)
- Programme CNRS " Société de l'Information " 2002 (J.M. Pierrel)
- Expert auprès du ministère de la recherche dans le cadre d'une réflexion sur les perspectives nationales en matière de calcul et de stockage intensif pour les sciences humaines et sociales (J.M. Pierrel)

9.1.6. Revues

- Éditeur en chef du Journal of Logic, Language, and Information (P. Blackburn)
- Éditrice du bulletin du chapitre européen de l'Association pour la Linguistique Informatique (*EACL Newsletter*) (C. Gardent)
- Co-rédactrice en chef du journal T.A.L. (Traitement Automatique des Langues)(C. Gardent)
- Rédacteur en chef adjoint de la Revue d'Intelligence Artificielle (RIA), Editions Hermes (J.M. Pierrel)
- Membre du Comité de rédaction pour : Journal of Semantics (C. Gardent), Traitement Automatique des Langues (C. Gardent, J-M. Pierrel), Computer and the Humanities (L. Romary), Information-Interaction-Intelligence (J-M. Pierrel)
- Membre du Conseil scientifique d'Hermès Sciences Publications (J-M. Pierrel)
- Membre du Comité de rédaction de TAL, RIA, I3 (J.M. Pierrel)

9.1.7. Colloques

- Présidente de ESSLLI 2002 (14th European Summer School for Language, Logic and Information), Trento, Italy. (C. Gardent)
- Président de de TALN02 (9ème Conférence sur le Traitement Automatique des Langues Naturelles), Nancy, Juin 2002(J-M. Pierrel)
- Président de RECITAL02 (Rencontre des Etudiants Chercheurs en Informatique pour le Traitement des Langues) Nancy, Juin 2002)(A. Roussanaly)
- Membre du Comité de lecture de TALN02 (C. Gardent, L. Romary)
- Membre du Comité de lecture de RECITAL02 (B. Gaiffe, L. Romary)
- Membre du Comité de lecture de PORTAL (C. Gardent)

- Membre du Comité de lecture de TAG+6 (C. Gardent)
- Membre du Comité de lecture de EDILOG (C. Gardent)
- Membre du Comité de lecture de INLG (C. Gardent)
- Membre du Comité de lecture de Existence (C. Gardent)

9.1.8. Organisations

- TALN 02 (Traitement automatique du Language Naturel) Nancy - June 2002
- RECITAL 02 (Rencontre des Etudiants Chercheurs en Informatique pour le Traitement des Langues) Nancy - 2002
- TKE 02 (Terminology and Knowledge Engineering) Nancy - Aug. 2002
- EXISTENCE (Sémantique et syntaxe) Nancy - Sep. 2002 (Co-organisatrice : C. Gardent)

9.1.9. Séminaires et conférences invitées

- Conférence invitée, Hybrid Logic and Description Logic, Workshop for GDR I3, Nancy, 6 December, 2002. (P. Blackburn)
- Conférence invitée, Arthur Prior and the Birth of Hybrid Logic, PILM 2002 (Philosophical Insights into Logic and Mathematics), Nancy, France, 30 September - 4 October 2002. (P. Blackburn)
- Conférence invitée, Hybrid Logic and Description Logic, DL 2002 (International Workshop on Description Logics), Toulouse, France, 19-21 April 2002, (P. Blackburn)
- Conférence invitée, Model Theoretic Syntax, Workshop on the Roots of Pragmasemantics, Szklarska Poreba, Poland, 2-6 March 2002. (P. Blackburn)
- "Les ressources de l'ATILF pour l'analyse textuelle : TLFi, Frantext et le logiciel Stella ", Conférence franco-québécoise " TALN, Corpus et Web ", Université Paris Nord, 27 novembre 2002 (J.M. Pierrel)
- Participation à la table ronde " Les trois révolutions du livre " organisée par le CNAM et la FNAC dans le cadre du " Café des techniques ", 23 octobre 2002 (J.M. Pierrel)
- Présentation du TLFi, Hotel de Ville de Nancy, " Le Livre sur la place ", 18 Septembre 2002 (J.M. Pierrel)
- Présentation du TLFi au CNRS, mars 2002 (J.M. Pierrel)
- "Recherche sur la langue : son analyse et son traitement informatique ", conférence invitée à la journée des professionnels, Salon du Livre, Paris 25 mars 2002 (J.M. Pierrel)
- Histoire de Frantext : constitution d'une base textuelle (1964-2002) et perspective, colloque " L'édition électronique en littérature et dictionnaire : évaluation et bilan ", Rouen, 17 juin 2002 (J.M. Pierrel)
- IST, Langue et Technologie, Journées des documentalistes du CNRS, INIST, Nancy, 18 juin 2002 (J.M. Pierrel)

9.2. Enseignement universitaire

La forte composante universitaire de l'équipe nous conduit à être très impliqués dans les structures d'enseignement universitaire. On peut noter en particulier :

- co-responsabilité de la filière « Perception, raisonnement et Traitement automatique des langues » du DEA d'Informatique de Lorraine (J-M. Pierrel)
- responsabilité du DEA « Modélisation et simulation des espaces bâtis » (J-M. Pierrel)
- responsabilité de la Licence et Maîtrise « Sciences Cognitives » à l'université Nancy 2 (D. Coulon)
- responsabilité de la commission informatique du département SRC à l'IUT de St Dié (J-L. Husson)
- responsabilité du service informatique/multimédia à l'ENSG (C. Fay-Varnier)

- direction des relations industrielles du département SRC à l'IUT de St Dié (J-L. Husson)
- participation à des enseignements de bio-informatique dans le DESS RGTI (Ressources Génomiques et Traitements Informatiques) et la maîtrise MGMC (Mention Génétique Moléculaire et Cellulaire) (M. Smaïl, M-D. Devignes)
- participation à la filière PRTAL du DEA d'informatique (B. Gaiffe, L. Romary)
- participation à la filière « Sciences Cognitives » à l'université Nancy 2 (D. Coulon, A. Roussanaly, E. Jacquy, H. Manuélian, B. Crabbé)
- participation au DESS « TEXTE » à l'université Nancy 2 (J-M. Pierrel, L. Romary, D. Seddah, C. Gardent)
- participation au DESS « Industries de la langue » à l'université de Metz (S. Cruz-Lara, D. Coulon)
- participation au DESS « Systèmes d'Information Distribués » à l'Université Nancy 2 (S. Cruz-Lara)
- participation à la Licence Professionnelle « Concepteur-Intégrateur de Systèmes Intranet / Internet » de l'Université Nancy 2 (S. Cruz-Lara)
- participation à la filière « Sciences du langage » à l'Université Nancy 2 (H. Manuelian, E. Jacquy)
- participation à la filière « Sciences du langage » à l'université de Metz (H. Manuelian, C. Gardent)
- participation à la filière « Mathématiques, Informatique et Statistiques Appliquées aux Sciences Humaines et Sociales » à l'Université Nancy 2 (B. Crabbé, A. Roussanaly)
- participation à la filière « MIAGE » à l'Université Nancy 2 (B. Crabbé, S. Cruzlara)
- participation à la Maîtrise des « Sciences de l'Information et de la Documentation » à l'Université Nancy 2 (N. Viscogliosi)

9.3. Autres enseignements

Parmi les autres enseignements assurés par des membres du projet Langue et Dialogue, il convient de noter :

- Ecole d'été NASSLLI, Lectures on Hybrid Logic, (North American Summer School in Logic, Language and Information), Stanford, USA, 24-30 Juin 2002, (P. Blackburn)

10. Bibliographie

Bibliographie de référence

- [1] P. BLACKBURN, M. DE RIJKE, Y. VENEMA. *Modal Logic*. série Cambridge Tracts in Theoretical Computer Science, volume 53, Cambridge University Press, juin, 2001.
- [2] P. BLACKBURN, M. KOHLHASE. *Inference in Computational Semantics*. juin, 2001, ICoS-3 Workshop Proceedings, June 18-19, Siena, Italy.
- [3] P. BLACKBURN, M. DE RIJKE, Y. VENEMA. *Modal Logic*. Cambridge University Press, 2001.
- [4] D. DUCHIER, C. GARDENT. *Tree Descriptions, Constraints and Incrementality*. éditeurs R. M. H. BUNT, E. T. (EDS)., in « Computing Meaning », série Studies in Linguistics and Philosophy Series, volume 2, Kluwer Academic Publishers, 2001.
- [5] N. IDE, L. ROMARY. *Encoding syntactic annotation*. éditeurs A. ABEILLÉ., in « Treebanks », Kluwer Academic Publisher, 2001.
- [6] P. LOPEZ. *Analyse d'énoncés oraux pour le dialogue homme-machine à l'aide de grammaires lexicalisées d'arbres*. thèse de doctorat, Université Henri Poincaré, Nancy 1, 1999.

- [7] J.-M. PIERREL. *Ingénierie des Langues*. série Traité IC2 (Information, communication et commande, Hermes, octobre, 2000.
- [8] L. ROMARY, P. BONHOMME. *Parallel Alignment of Structured Documents*. éditeurs J. VÉRONIS., Kluwer Academic Press, 2000.
- [9] L. ROMARY. *Langue et dialogues - perspectives de recherche*. Habilitation à diriger des recherches, décembre, 1999.
- [10] G. SABAH, J. VIVIER, A. VILNAT, J.-M. PIERREL, L. ROMARY. *Machine, langage et dialogue*. L'Harmattan, 1998.
- [11] S. SALMON-ALT. *Référence et dialogue finalisé : de la linguistique à un modèle opérationnel*. Thèse d'université, Nancy, mai, 2001.

Articles et chapitres de livre

- [12] C. ARECES, P. BLACKBURN, M. MARX. *Repairing the Interpolation Theorem in Quantified Modal Logic*. in « Annals of Pure and Applied Logic », février, 2002.
- [13] P. BLACKBURN, M. MARX. *Constructive Interpolation in Hybrid Logic*. in « Journal of Symbolic Logic », août, 2002, à paraître.
- [14] P. BLACKBURN, M. MARX. *Remarks on Gregory's « Actually » Operator*. in « Journal of Philosophical Logic », volume 31, juin, 2002, pages 281-288.
- [15] M.-D. DEVIGNES, A. SCHAAFF, M. SMAÏL. *Collecte et intégration de données biologiques hétérogènes sur le web : application dans le domaine de la cartographie du génome humain*. in « Ingénierie des systèmes d'information », numéro 1-2, volume 7, septembre, 2002, pages 45-61.
- [16] C. GARDENT, J.-M. PIERREL. *Aspects linguistiques du traitement automatique du dialogue*. in « Traitement Automatique des Langues », décembre, 2002.
- [17] F. LANDRAGIN, S. SALMON-ALT, L. ROMARY. *Ancrage référentiel en situation de dialogue*. in « Traitement Automatique des Langues », numéro 2, volume 43, décembre, 2002, pages 99-129.
- [18] J.-M. PIERREL. *De l'avantage de structures de laboratoire solides en Sciences du Langage*. in « Lettre du département SHS du CNRS », 2002, à paraître.
- [19] J.-M. PIERREL. *La modélisation en linguistique*. in « Lettre du département SHS du CNRS », 2002.
- [20] L. ROMARY. *De la sémantique des contenus à la sémantique des structures*. in « La Recherche d'information sur les réseaux », série Sciences de l'information, série Etudes et techniques, ADBS Editions, octobre, 2002, pages 203-230.

Communications à des congrès, colloques, etc.

- [21] M. AMOIA, C. GARDENT, S. THATER. *Using set constraints to generate distinguishing descriptions*. in « 7th International Workshop on Natural Language Understanding and Logic Programming - NLULP'02, Copenhagen, Denmark », juillet, 2002.
- [22] P. BERNARD, J. DENDIEN, J. LECOMTE, J.-M. PIERREL. *La structuration du TLFi, codage et balisage*. in « L'édition électronique en littérature et dictionnaire : évaluation et bilan », Editions Champion, 2002, à paraître.
- [23] C. BERNET, J.-M. PIERREL. *Histoire de Frantex : constitution d'une base textuelle (1964-2002) et perspectives*. in « L'édition électronique en littérature et dictionnaire : évaluation et bilan », Editions Champion, 2002, à paraître.
- [24] P. BLACKBURN, M. MARX. *Tableaux for Quantified Hybrid Logic*. in « International Conference on Automated Reasoning with Analytic Tableaux and Related Methods - TABLEAUX 2002, Copenhagen, Denmark », série Lecture Notes in Artificial Intelligence, volume 2381, Springer, éditeurs C. F. U. EGLY., pages 38-52, juillet, 2002.
- [25] P. BLUMENTHAL, P. BERNARD, J.-M. PIERREL. *Quels corpus pour un dictionnaire des collocations ?*. in « 2ème journées de la linguistique de corpus », Lorient », septembre, 2002.
- [26] D. BROEDER, P. WITTEBURG, T. DECLERCK, L. ROMARY. *LREP : A Language Repository Exchange Protocol*. in « Third International Conference on Language Resources and Evaluation - LREC'2002, Las Palmas, Spain », mai, 2002.
- [27] H. BUNT, L. ROMARY. *Towards Multimodal Content Representation*. in « LREC Workshop on International Standards of Terminology and Language Resources Management, Las Palmas, Spain », mai, 2002.
- [28] S. CRUZ-LARA, B.-H. CHEN, J.-S. HONG. *Distributed Content Management Framework for Digital Museum Exhibitions*. in « The Scholars Conference 2002 "Understanding the Future of European e-Content Industries", Tampere, Finland », novembre, 2002.
- [29] S. CRUZ-LARA, J.-S. HONG. *XML-Based Content Management Framework for Digital Museum*. in « 3rd Sino Franco Workshop on Multimedia and Web Technologies, Sophia Antipolis, France », mars, 2002.
- [30] M.-D. DEVIGNES, M. SMAÏL, N. BOUDJLIDA. *Collecte de données biologiques à partir de sources multiples et hétérogènes. Vers une structure de médiation conviviale et orientée source*. in « Journées scientifiques sur le Web sémantique », Paris, France », J. Charlet, P. Laublet, C. Raynaud, octobre, 2002, <http://www.lalic.paris4.sorbonne.fr/stic/octobre/programme0209.html>, Actes électroniques.
- [31] G. DUFFING, M. SMAÏL. *Organising and Searching Partially Indexed Image Databases*. in « 24th BCS-IRSG European Colloquium on Information Retrieval Research », Glasgow, Scotland, UK », Springer Verlag, mars, 2002.

- [32] B. GAIFFE, B. CRABBÉ, A. ROUSSANALY. *A New Metagrammar Compiler*. in « Sixth International Workshop on Tree Adjoining Grammars and Related Frameworks - TAG+6, Venice, Italy », mai, 2002.
- [33] C. GARDENT. *Generating minimal definite descriptions*. in « 40th Annual Meeting of the Association for Computational Linguistic - ACL'02, Philadelphia, USA », juillet, 2002.
- [34] N. IDE, L. ROMARY. *Standards for Language Resources*. in « Third International Conference on Language Resources and Evaluation - LREC'2002, Las Palmas, Spain », mai, 2002.
- [35] E. JACQUEY, J.-M. PIERREL. *Les déverbaux d'action en français : type d'ambiguïté et modélisation en TAL*. in « Journées " Déverbaux, ILF, Silex, Lille », mars, 2002.
- [36] F. LANDRAGIN, N. BELLALEM, L. ROMARY. *Referring to Objects with Spoken and Haptic Modalities*. in « IEEE International Conference on Multimodal Interfaces - ICMI'02, Pittsburgh, PA USA », éditeurs I. C. PRESS., octobre, 2002.
- [37] F. LANDRAGIN. *The Role of Gesture in Multimodal Referring Actions*. in « IEEE International Conference on Multimodal Interfaces - ICMI'02, Pittsburgh, PA USA », éditeurs I. C. PRESS., octobre, 2002.
- [38] P. LOPEZ, C. FAY-VARNIER, A. ROUSSANALY. *Lexicalized Grammar Specialization for Restricted Applicative Languages*. in « Third International Conference on Language Resources and Evaluation - LREC'2002 - Workshop Customizing Knowledge in NLP Applications, Las Palmas de Gran Canaria, Spain », European Language Resources Association (ELRA), mai, 2002.
- [39] H. MANUÉLIAN. *Annotation des descriptions définies : le cas des reprises par les rôles thématiques*. in « Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues - RECITAL'2002, Nancy, France », juin, 2002.
- [40] L. ROMARY, N. REITHINGER, C. LAUER. *MIAMM - Multidimensional Information Access using Multiple Modalities*. in « International CLASS Workshop, Copenhagen, Denmark », juin, 2002.
- [41] A. TODIRASCU, E. KOW, L. ROMARY. *Towards Reusable NLP Components*. in « Proceedings of the Third International Conference on Language Resources and Evaluation, Las Palmas, Iles Canaries, Spain », numéro 2-9517408-0-8, volume III, European Language Resources Association (ELRA), ELRA, éditeurs C. P. S. A. MANUEL GONZALEZ RODRIGUEZ., pages 1116-1123, mai, 2002.
- [42] A. TODIRASCU, L. ROMARY, D. BEKHOUCHE. *Combining Syntax and Ontologies for Information Extraction*. in « Terminology and Knowledge Engineering - TKE'02, Nancy, France », août, 2002.
- [43] A. TODIRASCU, L. ROMARY, D. BEKHOUCHE. *Vulcain - An Ontology-based Information Extraction System*. in « Natural Language for Databases - NLDB'02, Stockholm, Sweden », série Lecture Notes in Computer Science, DSV, Information Technology University, Springer Verlag, éditeurs P. J. BIRGER ANDERSSON., juin, 2002.
- [44] B. WEBBER, C. GARDENT, J. BOS. *Position statement : Inference in Question Answering*. in « Third international conference on Language Resources and Evaluation - LREC'02, Las Palmas, Spain », juin, 2002.

Bibliographie générale

- [45] N. BOUDJLIDA, M. SMAIL-TABBONE, M.-D. DEVIGNES. *Services for a Genomics Open Distributed Environment*. in « XEWA'2000 : IEEE Workshop on XML-Enabled Wide Area Search in Bioinformatics. (League City, Texas) », 2000.
- [46] C. CHELALA, D. M. IMBEAUD, R. ZOOROB, C. AUFFRAY. *Refined localization of twenty-one genes in subregion p13.1 of human chromosome 1*. in « Cytogenet. Cell Genet », 2001.
- [47] C. GARDENT, K. STRIEGNITZ. *Generating Indirect Anaphora*. in « Proceedings of the Fourth International Workshop on Computational Semantics (IWCS-4) », 2001.
- [48] D. SPERBER, D. WILSON. *Relevance : Communication and Cognition. (second edition)*. Harvard University Press, Cambridge, MA, 1995.
- [49] A. TROGNON, D. COULON. *La modélisation de raisonnements générés des interlocutions*. in « Langages, n.144 », décembre, 2001.