

SIGNES

**Signes linguistiques, grammaire et sens:
algorithmique logique de la langue**

**Proposition de projet INRIA-Futurs, Bordeaux
Thème Sym C**

Systèmes symboliques: organisations des contenus et de la langue

En partenariat avec

l'Université Bordeaux 1

l'Université Bordeaux 3 Michel de Montaigne

le Laboratoire Bordelais de Recherche en Informatique LaBRI U.M.R. 5800

l'Équipe de Recherche en Syntaxe et Sémantique E.R.S.S. U.M.R. 5610

l'équipe Textes, Langage, Cognition TELANCO J.E. 2385

Mots-clefs: linguistique informatique, traitement automatique des langues;
langages formels, automates, lambda calcul, logique des ressources;
morphologie, syntaxe et sémantique des langues;
syntaxe générative, sémantique de Montague, représentation du discours;
morphologie dérivationnelle, sémantique lexicale.

Version du 17 octobre 2004

Table des matières

1	L'équipe	3	8	Actions nationales et internationales (2000-2004)	44
			8.1	Actions nationales	44
			8.2	Actions internationales	45
2	Présentation et objectifs généraux	4	9	Diffusion de résultats	46
2.1	Actualité des liens entre linguistique et informatique	4	9.1	Animation de la Communauté scientifique (2000-2002)	46
2.2	L'interface entre syntaxe et sémantique	4	9.1.1	Distinctions et nominations	46
2.3	Résumé des objectifs	6	9.1.2	Comités de rédaction	46
2.3.1	Objectifs théoriques	6	9.1.3	Comités de lecture	46
2.3.2	Objectifs applicatifs	7	9.1.4	FoLLI	47
2.4	Spécificité de ce projet	8	9.1.5	Autres interventions	47
3	Fondements scientifiques	10	9.2	Participation à des jurys de thèse et d'habilitation (2000-2003)	48
3.1	Syntaxe et grammaires formelles	11	9.2.1	Habilitations à Diriger des Recherches	48
3.1.1	Morphologie et automates	12	9.2.2	Thèses	48
3.1.2	Structure syntagmatique	13	9.3	Encadrement (2000-2003)	49
3.2	Morphologie dérivationnelle et sémantique lexicale	18	9.3.1	Maîtrise	49
3.3	Sémantique prédicative	20	9.3.2	DEA	49
3.3.1	Sémantique de Montague	20	9.3.3	Thèses	50
3.3.2	DRT – Théorie de la Représentation du Discours	22	9.4	Enseignements en 2002/2003 et 2003/2004	51
3.3.3	Sémantique des jeux	24	9.4.1	2002/2003	51
3.3.4	Prise en compte de la présupposition	26	9.4.2	2003/2004	51
4	Objectifs	27	9.5	Publications des membres du projet (2000-2003)	52
4.1	Théorie des langages et algorithmique	27	9.5.1	DEA, thèses et habilitations à diriger des recherches	52
4.1.1	Algorithmes d'analyse et complexité	27	9.5.2	Volumes édités	53
4.1.2	Liens entre familles de grammaires	28	9.5.3	Livres	53
4.2	Modélisation linguistique	29	9.5.4	Articles	54
4.2.1	Morphologie et sémantique lexicale	29	9.5.5	Communications à des congrès, colloques, etc.	55
4.2.2	Syntaxe et morphologie	29			
4.2.3	Grammaire de la langue des signes	30			
4.2.4	Résolution d'anaphores	32			
4.2.5	L'interface entre sémantique et discours	33			
4.2.6	Modélisation du discours et du dialogue	34			
4.2.7	Présupposition et modèle partiel	35			
4.2.8	Syntaxe et sémantique des pronoms	35			
4.3	Objectifs de développement	36			
4.3.1	Implantation des modèles proposés en OCAML	36			
4.3.2	Génération d'énoncés en langue des signes	37			
5	Domaines d'application	38			
5.1	Philologie du Sanskrit	38			
5.2	Traitement Automatique de la Langue des Signes Française	39			
6	Principales collaborations	41			
6.1	Nationales	41			
6.1.1	Collaborations Bordelaises	41			
6.1.2	INRIA	41			
6.1.3	Autres	42			
6.2	Internationales	43			
7	Contrats	44			

1 L'équipe

Responsable scientifique

Christian RETORÉ PR U. Bx 1 LABRI

Responsable permanent

Henri PORTINE PR U. Bx III TELANCO

Autres membres permanents

Christian BASSAC MC U. Bx III E.R.S.S.
Joan BUSQUETS MC U. Bx III E.R.S.S.
Kim GERDES MC U. Bx III E.R.S.S.
Patrick HENRY IR CNRS LaBRI
Gérard HUET DR INRIA-Rocquencourt CRISTAL 50%
Renaud MARLET CR INRIA
Richard MOOT CR CNRS LABRI

Collaborateurs extérieurs

Agnès BRACKE U. Bx III TELANCO
Olivier DE LANGHE Édicateur - I.N.J.S. Bx TELANCO
Alain LECOMTE PR U. Grenoble 2 CLIPS
Yannick LE NIR MC EISTI, Pau

Secrétaire

Brigitte LARUE-BOURDON INRIA

Doctorants

		Financement		Depuis:
Maxime	AMBLARD	MENRT U. Bx 1	LaBRI	oct. 03
Roberto	BONATO	cotutelle U. Vérone - U. Bx 1	LaBRI	oct. 02
Pierre	GUITTENY	Chargé de Mission - DRASS	TELANCO	oct. 03
Émilie	VOISIN	alloc. région U. Bx III	TELANCO	oct. 04

Abréviations:

U. Université
Bx Bordeaux
CLIPS UMR 5524 C.N.R.S. *Communication Langagière et Interaction Personne-Système* Université Joseph Fourier et Institut National Polytechnique de Grenoble.
CR Chargé de Recherche
CRISTAL Projet de l'INRIA-Rocquencourt.
DR Directeur de Recherche.
DRASS Direction Régionale des Affaires Sanitaires et Sociales
EISTI Ecole Internationale des Sciences du Traitement de l'Information, Pau.

E.R.S.S. UMR 5610 CNRS, *Équipe de Recherche en Syntaxe et Sémantique*, Universités Bordeaux III et Toulouse II
I.N.J.S. Institut National des Jeunes Sourds de Bordeaux-Gradignan.
LaBRI UMR 5800 CNRS *Laboratoire Bordelais de Recherche en Informatique*, Université Bordeaux I et ENSEIRB
MC Maître de Conférences.
MENRT Ministère de l'Éducation Nationale, de la Recherche et de la Technologie.
PR Professeur des Unversités.
TELANCO JE 2385, *Texte, Langage, Cognition*, Jeune Équipe, Université Bordeaux III

2 Présentation et objectifs généraux

2.1 Actualité des liens entre linguistique et informatique

Depuis les début de l'informatique, la langue est l'un de ses champs d'investigation privilégiés, comme en témoignent les travaux de traduction automatique des années cinquante^[1]. On notera que l'intérêt est ou a été réciproque, puisqu'à cette même période, la linguistique formelle a fourni à l'informatique naissante la théorie des langages qui s'est avérée pertinente pour bien des aspects de cette science comme la compilation ou le parallélisme.^[2]

Aujourd'hui la nécessité de traiter la masse d'informations disponibles, en particulier sur Internet, a conduit à de nombreuses réalisations de traitement automatique des langues, qui sont omniprésentes, en général sous forme textuelle, mais aussi sous forme vocale; en parallèle la modélisation de la langue pose des questions mathématiques et informatiques intrinsèquement intéressantes dont les solutions peuvent conduire à d'autres applications.

Les outils de traitement des langues les plus répandus sont sans doute la recherche d'information, les correcteurs orthographiques, et dans une moindre mesure la traduction assistée par ordinateur, la génération de textes et le résumé automatique.

Si les méthodes statistiques ont connu un développement considérable ces dernières années, les méthodes symboliques, et en particulier logiques, connaissent un véritable renouveau, dû à plusieurs facteurs : les avancées de la logique, l'accroissement des capacités de calcul qui vient pallier la lenteur des algorithmes issus de la logique, et le besoin d'applications plus subtiles où le sens des énoncés est pris en compte.

2.2 L'interface entre syntaxe et sémantique

Bien des applications dont on espère qu'elles verront le jour doivent tenir compte du sens des énoncés, que ce soit pour exprimer une requête en langage naturel, pour synthétiser une phrase, ou pour une aide à la traduction. Dans toutes ces applications il importe de savoir relier le sens d'un énoncé à sa structure syntaxique. Nous nous focaliserons donc sur le lien qu'entretiennent la sémantique et la syntaxe d'un énoncé, et rejoignons ainsi des préoccupations familières à nos collègues linguistes. Nous souhaitons bien sûr étendre cette correspondance au ni-

[1] Jacqueline Léon and Marcel Cori. La constitution du TAL — Étude historique des dénominations et des concepts. *Traitement Automatique des Langues*, 2002.

[2] G. Rozenberg and A. Salomaa, editors. *Handbook of Formal Languages*. Springer Verlag, Berlin, 1997.

veau du discours et du dialogue, mais les énoncés isolés seront notre objet d'étude premier.

Cette équipe souhaite donc se consacrer à l'interface entre syntaxe et sémantique qui présente des intérêts multiples:

- D'un point de vue applicatif, comme on l'a dit, cette interface est la clef de toutes les applications où le sens des énoncés est nécessaire. C'est par exemple le cas dans l'analyse lorsqu'on souhaite obtenir des représentations du sens des énoncés (par exemple en vue de les traduire mais aussi pour étendre le spectre de la recherche d'information) ou produire des énoncés dont on connaît le sens.
- D'un point de vue linguistique, c'est un domaine essentiel que l'on trouve déjà dans les deux faces du signe linguistique de Saussure, et dans les deux formes associées à un énoncé par Chomsky : LF (la forme logique) et PF (la forme phonologique), l'énoncé lui-même étant l'articulation entre ces deux aspects.^[1,2]
- D'un point de vue mathématique et algorithmique, c'est le siège d'interrogations prometteuses sur le lien entre les grammaires génératives fournies par les théories linguistiques dont l'algorithmique est particulièrement efficace et les systèmes logico-déductifs peu efficaces mais riches d'informations pertinentes; ces derniers sont utilisés aussi bien en syntaxe (grammaires catégorielles) qu'en sémantique (sémantique de Montague, DRT). Dans d'autres modèles linguistiques aussi, comme par exemple dans les grammaires de dépendance, on trouve des modélisations plus ou moins directes du lien entre représentations sémantiques et syntaxiques. Des formalisations et implémentations de l'interface dans le modèle chomskyen et, parallèlement, dans un modèle de dépendance, ouvrent un vaste champ de comparaison de l'adéquation linguistique de ces deux approches; c'est là une question essentielle de la linguistique formelle contemporaine.

Suivant la tradition de la grammaire générative la morphologie flexionnelle est incluse dans la syntaxe, tandis que les aspects ontologiques de la sémantique et de la pragmatique n'en font pas partie. Nous gardons cependant à l'esprit que les représentations sémantiques doivent pouvoir communiquer avec la sémantique lexicale, qui relie entre eux les sens des mots et ancre leur interprétation dans l'univers du discours. Le lexique génératif est un modèle pertinent dans la mesure où en plus de la représentation du lexique il fournit un lien avec la sémantique compositionnelle et la syntaxe.

[1] Noam Chomsky. *The minimalist program*. MIT Press, Cambridge, MA, 1995.

[2] Jean-Yves Pollock. *Langage et cognition: le programme minimaliste de la grammaire générative*. Presses Universitaires de France, Paris, 1997.

2.3 Résumé des objectifs

L'un des attraits de ce domaine nommé *linguistique computationnelle* (ou *informatique*) ou *traitement automatique des langues* est, comme le soulignent ces deux expressions, la rencontre de préoccupations pratiques (réalisation d'outils de traitement des langues) et théoriques (linguistiques, informatiques et mathématiques).^[1]

2.3.1 Objectifs théoriques

Objectifs mathématiques et algorithmiques Il s'agit essentiellement d'approfondir le lien entre grammaires de réécriture et grammaires catégorielles. En effet, ces deux familles de grammaires présentent des avantages complémentaires : les premières ont des algorithmes d'analyse particulièrement efficaces (polynomiaux) et rendent aisément compte des théories linguistiques, et les secondes ont une interface facilement calculable avec la sémantique et sont apprenables automatiquement. La démonstration en 92 par Pentus ^[2] de la conjecture formulée par Chomsky en 63 ^[3, p413], qui établit l'équivalence faible entre grammaires hors-contexte (ou algébriques) et grammaires de Lambek est loin de clore la question. À terme, il serait profitable d'avoir une correspondance effective qui permette de compiler les grammaires catégorielles en grammaires génératives, pour en rendre l'analyse plus efficace sans toutefois perdre le lien avec la sémantique.

Objectifs linguistiques La linguistique du XX^e siècle a vu l'émergence de la grammaire générative dont les postulats suggèrent une connexion étroite avec l'informatique.^[4,5,6] En effet, cette théorie affirme que l'activité linguistique humaine est un phénomène calculatoire de faible complexité algorithmique, et d'autre part que toutes les langues sont les instances d'une grammaire universelle innée.

La formalisation de la syntaxe et de la sémantique des langues permet donc au linguiste de confirmer, de raffiner ou d'infirmer ces hypothèses — par exemple,

-
- [1] Jacqueline Léon and Marcel Cori. La constitution du TAL — Étude historique des dénominations et des concepts. *Traitement Automatique des Langues*, 2002.
 - [2] Mati Pentus. Lambek grammars are context-free. In *Logic in Computer Science*. IEEE Computer Society Press, 1993.
 - [3] Noam Chomsky. Formal properties of grammars. In *Handbook of Mathematical Psychology*, volume 2, pages 323 – 418. Wiley, New-York, 1963.
 - [4] Noam Chomsky. Three models for the description of language. *IRE Transactions on Information Theory*, IT2:113–124, 1956.
 - [5] Jean-Yves Pollock. *Langage et cognition: le programme minimaliste de la grammaire générative*. Presses Universitaires de France, Paris, 1997.
 - [6] Ray Jackendoff. *The Architecture of the Language Faculty*. Number 28 in Linguistic Inquiry Monographs. M.I.T. Press, Cambridge, Massachusetts, 1995.

lorsqu'un argument d'efficacité algorithmique est avancé, la formalisation permet de calculer la complexité algorithmique et de vérifier si l'argument est fondé.

Les principes de la grammaire universelle doivent correspondre aux règles générales des grammaires lexicalisées ou en être dérivés, tandis que la variation des paramètres d'une langue à l'autre doit pour sa part correspondre aux catégories que le lexique associe aux mots : c'est à la fois un guide pour la modélisation informatique et un moyen de préciser, voire de vérifier les hypothèses linguistiques. La traditionnelle grammaire comparée trouve là une nouvelle dimension qui tend plus à dégager des principes universels qu'à distinguer les particularités de chaque langue, et cela permet d'envisager, d'un point de vue applicatif, une boîte à outils générique.

2.3.2 Objectifs applicatifs

Réalisation d'une boîte à outils de linguistique informatique La convergence des formalismes grammaticaux et logiques permet d'envisager la programmation d'outils génériques. Certains ont été réalisés : le traitement de la morphologie par des automates d'états finis (G. Huet, OCAML), la construction et la manipulation de réseaux de démonstration qui modélisent les analyses syntaxiques et les représentations sémantiques (S. Pogodalla, OCAML) un analyseur pour les grammaires catégorielles multimodales (R. Moot, Prolog), les algorithmes d'apprentissage des grammaires catégorielles (ARC GRACQ, Java et OCAML). Il s'agit de poursuivre par la programmation d'autres modules et surtout de s'assurer de leur bonne complémentarité, afin que les données produites par un module puissent aisément être utilisées par un autre.

On notera que les langages comme OCAML sont particulièrement adaptés, puisqu'ils sont eux aussi issus de la théorie de la démonstration et du lambda-calcul. On peut donc en attendre des retombées sur la pratique de CaML; celui-ci étant enseigné un peu partout, cette thématique peut fournir des exemples et des sujets de projet.

Sanskrit Gérard Huet a entrepris un traitement de la morphologie du sanskrit, par des techniques d'automates d'états finis programmés en OCAML. Cette application à la philologie devrait contribuer à un site où les textes sanskrits seraient accessibles en ligne, et où les mots renverraient, via une analyse morphologique, à l'entrée du dictionnaire qui leur correspond — un tel site existe pour le grec et le latin : <http://www.perseus.tufts.edu/>.

Traitement de la langue des signes Olivier De Langhe et Pierre Guitteny, et, dans une moindre mesure, Henri Portine et Christian Retoré, connaissent la langue

des signes et la communauté sourde, et dans le cadre d'un partenariat entre IBM et l'université Bordeaux I sur l'intégration des handicaps se proposent de travailler sur la formalisation de la grammaire de la langue des signes. Une application réelle serait d'intégrer au CD-ROM éducatif *Les signes de Mano* d'IBM la génération automatique d'énoncés en langue des signes (sous forme textuelle dans un premier temps).

Un projet du LaBRI sur ce thème a été déposé par notre équipe auprès de la région Aquitaine: *Traitement Informatique de la langue des signes française: génération automatique d'énoncés en LSF* et une allocation de recherche régionale a été obtenue sur un sujet similaire.

2.4 Spécificité de ce projet

Dans sa définition formelle ce projet présente la particularité d'inclure des enseignants-chercheurs de la section *Sciences du Langage* du CNU, et de réunir des informaticiens d'origines différentes : lambda calcul et théorie de la démonstration, théorie des langages formels, programmation fonctionnelle. On s'est ainsi assuré d'avoir à l'intérieur même du projet les compétences requises:

- Une bonne connaissance des faits de langue, de leur organisation et des théories linguistiques qui permettent de les décrire.
- Une maîtrise des structures mathématiques par lesquelles rendre compte des structures linguistiques: grammaires formelles et automates, logiques des ressources, lambda calcul, calcul des prédicats d'ordre supérieur.
- La connaissance de l'algorithmique de ces structures de données spécifiques, que sont les démonstrations formelles et les arbres de dérivation.
- La pratique de la programmation sur ces structures dans des langages de haut niveau adaptés: *CaML*, *Prolog*.

On peut résumer ainsi le domaine des membres permanents situés à Bordeaux: ^a Christian Bassac (syntaxe générative, morphologie, sémantique lexicale); Joan Busquets (sémantique de Montague, représentation du discours); Kim Gerdes (syntaxe, grammaire d'arbres adjoints, grammaires de dépendance); Richard Moot (logique modale, grammaires catégorielle); Henri Portine (sémantique, pragmatique, énonciation); Christian Retoré logique linéaire, grammaires catégorielles, grammaire générative.

Cette équipe, commune aux deux universités bordelaises concernées, permet aussi d'ouvrir aux étudiants de l'une des deux disciplines une fenêtre sur l'autre

a. Ne sont pas mentionnés deux chercheurs souhaitant se reconverter dans le traitement automatique des langues: Renaud Marlet chercheur INRIA reconnu en évaluation partielle et spécialisation de programmes et Patrick Henry ingénieur de recherche administration système du Gréco de programmation puis du LaBRI.

discipline à l'occasion du séminaire et de cours de master ouvert aux deux populations estudiantines.

L'application envisagée à la Langue des Signes Française requiert la connaissance de cette langue mais aussi un partenariat avec les communautés locales et des partenaires industriels, qui est le fruit d'un concours de circonstances. Notre étude participe ainsi au mouvement social en faveur de la reconnaissance de la langue des signes et à l'intégration des personnes handicapées.

3 Fondements scientifiques

La linguistique informatique est un vaste domaine. On peut déjà circonscrire les phénomènes linguistiques auxquels on se consacrera en les écrivant en gras dans une classification traditionnelle, et aux frontières perméables, de la linguistique (inspirée de ^[1]).

- (1) phonétique
- (2) phonologie (dont **prosodie**)
- (3) morphologie dérivationnelle
- (4) **morphologie flexionnelle**
- (5) **syntaxe**
- (6) **sémantique**
 - (6.1) **prédicative**
 - (6.2) **lexicale**
- (7) pragmatique

L'interface entre syntaxe et sémantique constitue la ligne essentielle de notre projet, ce qui correspond à ce que Chomsky appelle l'interface entre le système linguistique au sens étroit et le système conceptuel informationnel au sein de la faculté de langage ^[2,3]. On n'omettra pas d'aussi considérer, lorsque cela est pertinent, l'interface avec la phonologie, surtout dans ses aspects prosodiques, ce que Chomsky appelle interface avec le système articulatoire/perceptuel.

On se propose pour formaliser puis traiter ces phénomènes d'utiliser et de développer les modèles mathématiques et informatiques suivants :

- (1) **grammaires formelles** (syntaxe)
 - (1.1) classes de langages faiblement contextuels
 - (1.2) algorithmes d'analyse syntaxique
 - (1.3) complexité
 - (1.4) inférence grammaticale
- (2) **logique des ressources** (interface syntaxe / sémantique)
 - (2.1) logique linéaire non commutative ou partiellement commutative
 - (2.2) logique linéaire et grammaires formelles
 - (2.3) syntaxe graphique (réseaux de démonstration)

[1] John Lyons. *Introduction to theoretical linguistics*. Cambridge university press, 1968.

[2] Noam Chomsky. *The minimalist program*. MIT Press, Cambridge, MA, 1995.

[3] Ray Jackendoff. *The Architecture of the Language Faculty*. Number 28 in Linguistic Inquiry Monographs. M.I.T. Press, Cambridge, Massachusetts, 1995.

(3) **logiques d'ordre supérieur et lambda calcul** (sémantique)

(3.1) logique intensionnelle

(3.2) unification de lambda termes

Les objectifs se situent à la confluence de trois directions de recherche :

- (1) **La modélisation de la structure et des algorithmes mis en œuvre dans le traitement des langues**, ce qui, si l'on s'inspire de ce que l'on sait du traitement humain des langues, relève aussi des sciences cognitives
- (2) **La réalisation de modules de traitement automatique des langues** dont le Graal est la traduction automatique, et où des objectifs plus modestes sont la recherche d'information, les correcteurs orthographiques, l'interrogation en langage naturel, la génération de textes à partir de données très structurées (des démonstrations formelles, par exemple)
- (3) **L'étude des modèles mathématiques sous-jacents**, car s'il est vrai que les domaines concernés offrent un vaste espace pour modéliser les phénomènes linguistiques, il est également vrai que la modélisation linguistique pose des questions pertinentes et suggère des extensions et des modifications des modèles mathématiques existants.

Dans cette action l'aspect cognitif sera traité de manière implicite, par exemple en se référant aux modèles proposés par Chomsky qui le prennent déjà en compte. Du reste les critères psycholinguistiques retenus par Chomsky coïncident le plus souvent avec la nécessité d'avoir des processus efficaces au sens de la complexité algorithmique, que ce soit en temps de calcul ou en espace mémoire. On envisage d'étendre cette étude aux modèles théoriques des grammaires de dépendance, comme le modèle Sens-Texte ^[1] ou la « Word Grammar » ^[2] dont les aspects formels et computationnels ne sont que rarement étudiés dans un contexte linguistique. Ces modèles utilisent plusieurs niveaux d'arbres non ordonnés et les liens entre ces niveaux s'effectuent à l'aide de règles de réécritures de graphes.

3.1 **Syntaxe et grammaires formelles**

Suivant la tradition chomskyenne, la morphologie flexionnelle (production ou analyse des formes fléchies) est considérée comme partie intégrante de la syntaxe. Ce choix fait partie du point de vue universaliste. En effet, il est bien connu que certaines langues utilisent la morphologie et d'autres des constructions syntaxiques pour décrire un même phénomène; par exemple le français indique le temps par une désinence verbale tandis que la langue des signes utilise un signe

[1] Igor Mel'čuk. *Dependency syntax – theory and practice*. Linguistics. State University of New York Press, 1988.

[2] Richard A. Hudson. *Word Grammar*. B. Blackwell, Oxford, UK, 1984.

générique en fin de proposition. Cependant les techniques utilisées diffèrent légèrement car le plus souvent la morphologie est décrite par des automates d'états finis tandis que les grammaires utilisées en syntaxe nécessitent au strict minimum des automates à pile, et sont souvent traitées par des grammaires d'arbres ou des systèmes déductifs.

3.1.1 Morphologie et automates

La modélisation informatique de la phonétique et de la morphologie est l'une des applications principales des méthodes d'état fini : les expressions régulières, les langages rationnels, les automates et transducteurs d'état fini ont été exploités extensivement dans les applications de reconnaissance de la parole et de traitement de la langue ^[1,2,3]. Ces méthodes combinent souvent des structures logiques (règles de réécriture notamment) avec des modèles statistiques (automates pondérés construits à partir de l'analyse sur des corpus de chaînes de Markov cachées).

Pour la morphologie, le travail pionnier de Koskenniemi ^[4] a été étendu à un cadre général de relations et de transducteurs rationnels par Kaplan et Kay ^[5], à la base du Xerox morphology toolset ^[6,7]. Dans cette approche, les bases de données lexicales et les tables de transformations phonétiques et morphologiques sont compilées (ou interprétées) en une algèbre d'opérations sur des automates. Des boîtes à outils de fonctionnalité similaire ont été développées à l'Université Paris VII, à l'Université de Groningen, aux Bell Labs, à Mitsubishi Labs, etc.

La technologie de compilation de telles machines est complexe et relativement confidentielle. L'utilisation naïve de tels outils peut mener à une explosion combinatoire difficile à maîtriser. Lorsque les règles de transformation ont de bonnes propriétés de localité, il est toutefois possible de les compiler en un transducteur inversible obtenu directement à partir de l'arbre lexical représentant le lexique. Une technologie générique de partage permet alors de représenter de tels automates de façon très compacte. Gérard Huet a ainsi mis au point une boîte à outils

[1] Richard Sproat. *Morphology and Computation*. MIT Press, 1992.

[2] Mehryar Mohri. Finite-state transducers in language and speech processing. *Computational Linguistics*, 23,2:269–311, 1997.

[3] Emmanuel Roche and Yves Schabes. *Finite-State Language Processing*. MIT Press, 1997.

[4] K. Koskenniemi. A general computational model for word-form recognition and production. In *10th International Conference on Computational Linguistics*, 1984.

[5] Ronald M. Kaplan and Martin Kay. Regular models of phonological rule systems. *Computational Linguistics*, 20,3:331–378, 1994.

[6] Lauri Karttunen. Applications of finite-state transducers in natural language processing. In *Proceedings, CIAA-2000*, 2000.

[7] Kenneth R. Beesley and Lauri Karttunen. *Finite-State Morphology: Xerox Tools and Techniques*. Cambridge University Press, 2002.

basée sur cette méthodologie, qui permet notamment de segmenter automatiquement un texte phonémique continu. Ceci a permis la réalisation d'un algorithme d'étiquetage du sanskrit, par inversion des règles d'euphonie (sandhi). L'extension de cette technologie à des règles plus générales sera l'un de nos sujets de recherche.

Ce niveau de représentation de la langue, ou *structure linéaire*, est le prérequis de tout traitement informatique de la langue, et il permet déjà d'envisager des applications telles que la recherche d'informations ou la vérification orthographique, sans mettre en œuvre les techniques plus sophistiquées d'analyse des couches syntaxiques et sémantiques.

3.1.2 Structure syntagmatique

Tandis que la structure linéaire suffit en général pour la morphologie, la structure syntagmatique (ou hiérarchique) de la phrase et de ses constituants nécessite une structure d'arbre pour associer représentations sémantiques et analyses syntaxiques.

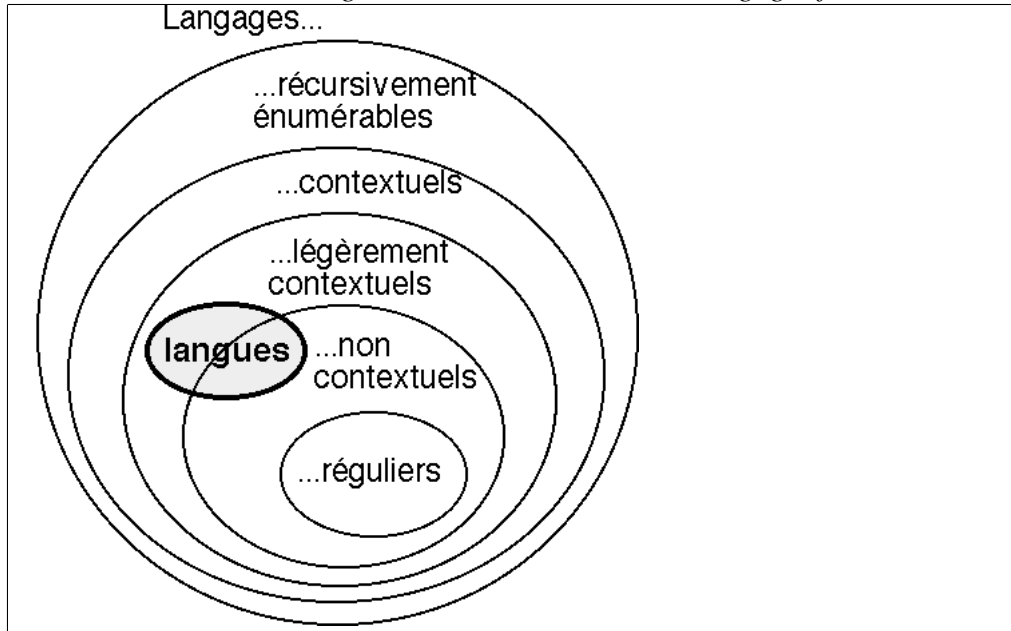
Les compétences de l'équipe concernent deux familles de modèles syntaxiques, d'une part les grammaires de réécriture, d'autres part les grammaires déductives. Par grammaires de réécriture, on désigne ici les grammaires de la célèbre hiérarchie de Chomsky ^[1] et par grammaires déductives on se réfère à la présentation des grammaires catégorielles comme des systèmes déductifs dans les logiques des ressources. Comme on le voit sur la figure 3.1.2, il est admis que la classe des langues, du point de vue des langages formels, contient des langages hors contexte (ou algébriques) et des langages faiblement contextuels (*mildly context sensitive*) ^[2], mais reste bien en deçà de la classe des langages contextuels.

Grammaires de réécriture Ces grammaires sont omniprésentes en informatique, y compris dans son enseignement. La célèbre hiérarchie de Chomsky a pourtant dû être étendue pour traiter de la syntaxe des langues. D'une part Chomsky lui-même a proposé des grammaires transformationnelles, qui autorisent des déplacements de constituants, et qui rendent compte du lien entre des énoncés associés, comme par exemple celui qui relie une affirmative à l'une des questions correspondantes. D'autre part les grammaires de réécriture ont incorporé l'unification de traits portés par les non-terminaux, en particulier pour rendre compte des

[1] Noam Chomsky. Three models for the description of language. *IRE Transactions on Information Theory*, IT2:113–124, 1956.

[2] Aravind Joshi, K. Vijay-Shanker, and David Weir. The convergence of mildly context-sensitive grammar formalisms. In P. Sells, S. Schieber, and T. Wasow, editors, *Fundational issues in natural language processing*. MIT Press, 1991.

FIG. 1 – Les langues dans la hiérarchie des langages formels



phénomènes d'accord; les DCG de Prolog sont par exemple obtenues par ajout de traits atomiques aux grammaires hors-contextes (ou algébriques).^[1]

Dans cette famille de grammaires, deux manières de décrire les langages faiblement contextuels retiennent notre attention : les TAG ^[2,3] et surtout les grammaires minimalistes de Stabler ^[5] qui formalisent le récent programme minimaliste de Chomsky^[7]. Un attrait important des TAG est leur grande efficacité algorithmique ainsi que l'existence de bases de données d'arbres, tandis que l'atout majeur des grammaires minimalistes est de reposer sur une théorie linguistique

-
- [1] Fernando C. N. Pereira and Stuart M. Shieber. *Prolog and Natural-Language Analysis*. Number 10 in CSLI Lecture Notes. University of Chicago Press, Chicago, IL, 1987.
 - [2] Aravind Joshi, Leon Levy, and Masako Takahashi. Tree adjunct grammar. *Journal of Computer and System Sciences*, 10:136–163, 1975.
 - [3] Aravind Joshi and Yves Schabes. Tree adjoining grammars. In Rozenberg and Salomaa ^[4], chapter 2.
 - [4] G. Rozenberg and A. Salomaa, editors. *Handbook of Formal Languages*. Springer Verlag, Berlin, 1997.
 - [5] Edward Stabler. Derivational minimalism. In Retoré ^[6], pages 68–95.
 - [6] Christian Retoré, editor. *Logical Aspects of Computational Linguistics, LACL'96*, volume 1328 of *LNCS/LNAI*. Springer-Verlag, 1997.
 - [7] Noam Chomsky. *The minimalist program*. MIT Press, Cambridge, MA, 1995.

conséquence qui incorpore les progrès de la grammaire générative. Ces deux formalismes ont des liens étroits avec les grammaires catégorielles, dont l'intérêt premier est la correspondance avec la sémantique.

Un deuxième attrait des TAG est le lien qu'elles fournissent entre les grammaires de réécriture et les grammaires de dépendance. En effet, chaque analyse TAG ne donne pas seulement un arbre de constituants, mais aussi une structure dite « arbre de dérivation », un arbre non ordonné qui se rapproche des représentations syntaxiques des grammaires de dépendance. Du point de vue des grammaires de dépendance, les TAG constituent un système efficace de linéarisation d'un arbre de dépendance, et beaucoup d'applications informatiques des TAG se trouvent en effet dans la génération de textes; par exemple tous les systèmes au centre allemand de recherche en intelligence artificielle, DFKI, basent leur modules de génération syntaxiques sur les TAG. En même temps, pour des phénomènes complexes d'ordre des mots, par exemple le « scrambling » en allemand, l'insuffisance du formalisme a été démontrée ^[1,2,3]. Un but important de la syntaxe formelle est de trouver un formalisme similaire, suffisamment puissant pour une analyse satisfaisante des langues naturelles mais qui garde les qualités computationnelles exceptionnelles des TAG.

Grammaires catégorielles La description des grammaires peut aussi se faire par des systèmes déductifs.^[4,6] Ces formalismes représentent une grammaire par un lexique qui à chaque mot m associe une ou plusieurs formules $\text{Lex}(m)$ qui décrivent le comportement syntaxique de m , le cas le plus simple étant celui des grammaires de Lambek^[7] (voir figure 3.1.2). Les formules de Lex sont engendrées par des connecteurs logiques à partir de propositions élémentaires et l'appartenance d'un énoncé $m_1 m_2 \dots m_n$ au langage engendré par Lex est vrai lorsqu' il

-
- [1] Tilman Becker and Owen Rambow et Michael Niv. The derivational generative power, or, scrambling is beyond lcfers. Technical report, University of Pennsylvania, 1992.
 - [2] O. Rambow and A. K. Joshi. A formal look at dependency grammars and phrase-structure grammars, with special consideration of word-order phenomena. In Leo Wanner, editor, *Current Issues in Meaning-Text Theory*. Pinter, London, UK, 1994.
 - [3] Owen Rambow. *Formal and Computational Aspects of Natural Language Syntax*. PhD thesis, University of Pennsylvania, 1994.
 - [4] Michael Moortgat. Categorical type logic. In van Benthem and ter Meulen ^[5], chapter 2, pages 93–177.
 - [5] Johan van Benthem and Alice ter Meulen, editors. *Handbook of Logic and Language*. North-Holland Elsevier, Amsterdam, 1997.
 - [6] Wojciech Buszkowski. Mathematical linguistics and proof theory. In van Benthem and ter Meulen ^[5], chapter 12, pages 683–736.
 - [7] Joachim Lambek. The mathematics of sentence structure. *American mathematical monthly*, pages 154–170, 1958.

FIG. 2 – Les grammaires de Lambek

Les formules de Lex sont engendrées par des connecteurs logiques à partir de propositions élémentaires dont S :

$$\mathcal{F} ::= \{S, sn, n, \dots\} \quad | \quad \mathcal{F} \multimap \mathcal{F} \quad | \quad \mathcal{F} \circ \mathcal{F} \quad | \quad \dots$$

et l'appartenance d'un énoncé $m_1 m_2 \dots m_n$ au langage engendré par Lex est définie par

$$\forall i \in [1, n] \exists t_i \in \text{Lex}(m_i) \quad t_1, \dots, t_n \vdash S$$

où \vdash est une relation de déduction dans une logique sensible aux ressources:

A hypothèse libre la plus à gauche

$\dots [A] \dots$

\vdots

B

$$\frac{B}{A \multimap B} \multimap_i \text{ qui lie } A$$

$$\frac{\begin{array}{c} \Delta \\ \vdots \\ A \end{array} \quad \begin{array}{c} \Gamma \\ \vdots \\ A \multimap B \end{array}}{B} \multimap_e$$

A hypothèse libre la plus à droite

$\dots [A] \dots$

\vdots

B

$$\frac{B}{B \circ A} \circ_i \text{ qui lie } A$$

$$\frac{\begin{array}{c} \Gamma \\ \vdots \\ B \circ A \end{array} \quad \begin{array}{c} \Delta \\ \vdots \\ A \end{array}}{B} \circ_e$$

Les règles d'introductions ci dessus ne s'appliquent qu'en présence d'au moins deux hypothèses libres.

est possible d'attribuer à chaque mot du lexique un de ses types, de sorte que la suite de types obtenus entraîne S dans une logique des ressources.

Du point de vue des langages décrits, les grammaires de Lambek correspondent aux langages hors-contexte (ou algébriques) ^[1] mais du point de vue des arbres d'analyse leur pouvoir d'expression est plus grand : ce sont des langages algébriques et non rationnels d'arbres ^[2].

[1] Yehoshua Bar-Hillel, Chaim Gaifman, and Eli Shamir. On categorial and phrase-structure grammars. *Bulletin of the research council of Israel*, F(9):1–16, 1963.

[2] Hans-Jörg Tiede. *Deductive Systems and Grammars: Proofs as Grammatical Structures*.

Grammaires de dépendance Contrairement aux formalismes des grammaires catégorielles ou de réécriture, les grammaires de dépendance ne sont pas, *stricto sensu*, un formalisme. Il s'agit plutôt d'un courant linguistique qui met en avant des arbres de dépendance comme représentation principale de la syntaxe. Ces arbres décrivent les liens syntaxiques, dits « dépendances », entre les mots d'une phrase et la manière dont ces arbres sont obtenus à partir de la phrase peut varier. Le plus souvent, les formalismes proposés s'apparentent aux systèmes de réécriture et on ne peut donc pas opposer directement ces deux approches syntaxiques. Les représentations en arbre de dépendance sont sûrement plus anciennes que celles utilisant des arbres de constituants et leurs origines remontent aux analyses des grammairiens classiques. Cependant, ce n'est qu'avec Tesnière ^[1] que le terme de dépendance est mis à la base d'une théorie complète de la syntaxe — voir par exemple ^[2] pour un état de l'art récent. Dans un arbre de constituants, les dépendances sont représentées implicitement lorsque les catégories non terminales permettent d'identifier une unique tête du syntagme (par exemple, la tête du syntagme verbal est le verbe) ; un arbre de dépendance les représente directement. Dans une approche de dépendance, les arbres syntagmatiques deviennent superflus, ou ne sont qu'un artefact du formalisme sans qu'on lui attache une quelconque valeur syntaxique. De cette distance entre la suite de mots constituant une phrase et sa représentation syntaxique, résulte une plus grande liberté dans le choix de la représentation et, en général, ce choix découle plus de considérations linguistiques plutôt que de principes formels. Ce n'est qu'après avoir fixé la représentation correspondant à une phrase que l'algorithmique tâche d'améliorer son calcul.

Comparaisons des différents types de formalismes Une propriété importante des grammaires est d'être ou non *lexicalisées*. Ce néologisme signifie que la grammaire d'un langage est totalement définie par un lexique qui spécifie le comportement des mots (les terminaux, d'un point de vue formel) tandis que les règles de production sont les mêmes pour tous les langages de la classe. C'est un avantage théorique puisque dans une perspective universaliste la variation entre les langues est supposée être purement lexicale, mais c'est aussi un avantage pratique qui facilite la définition d'algorithmes d'inférence grammaticale et le calcul de représentations sémantiques. Les grammaires catégorielles sont toujours lexicalisées, tandis que seules certaines grammaires de réécriture le sont.

PhD thesis, Illinois Wesleyan University, 1999. <http://www.iwu.edu/htiede/>.

[1] Lucien Tesnière. *Éléments de syntaxe structurale*. Éditions Klincksieck, 1959. 5^e édition: 1988.

[2] S. Kahane, editor. *Grammaires de dépendances*, volume 41(1) of *Traitement Automatique des Langues*. Hermès, 2000.

FIG. 3 – Comparaisons entre différents types de grammaires pour le TAL

	<i>grammaires déductives</i>	<i>TAGs ou MGs</i>	<i>grammaires de réécriture du type CFG</i>
<i>lexicalisées</i>	<i>oui</i>	<i>oui</i>	<i>non</i>
<i>interface syntaxe / sémantique</i>	<i>oui</i>	<i>non</i>	<i>non</i>
<i>complexité de l'analyse</i>	<i>NP</i>	<i>P</i>	<i>P</i>
<i>algorithmes d'inférence grammaticale</i>	<i>oui</i>	<i>non</i>	<i>non</i>

Si l'on souhaite, pour les classes pertinentes du point de vue de la syntaxe des langues, comparer ces deux familles de formalismes, ainsi que les TAGs^[1] ou les MGs de Stabler^[3] qui présentent des particularités intéressantes, on obtient le tableau de la figure 3.1.2.

Ce tableau montre qu'on a tout intérêt à approfondir le lien entre ces deux types de formalismes dont les avantages sont complémentaires. En particulier pour la modélisation de la langue, ou pour les algorithmes d'analyse, les grammaires de réécriture sont préférables, tandis que pour le calcul de représentations sémantiques ou la production d'énoncés à partir de représentations sémantiques les grammaires catégorielles sont préférables.

3.2 Morphologie dérivationnelle et sémantique lexicale

Le Lexique Génératif est un mode de représentation riche de la sémantique interne des mots comme des morphèmes ce qui le rend particulièrement pertinent

-
- [1] Aravind Joshi and Yves Schabes. Tree adjoining grammars. In Rozenberg and Salomaa^[2], chapter 2.
 - [2] G. Rozenberg and A. Salomaa, editors. *Handbook of Formal Languages*. Springer Verlag, Berlin, 1997.
 - [3] Edward Stabler. Derivational minimalism. In Retoré^[4], pages 68–95.
 - [4] Christian Retoré, editor. *Logical Aspects of Computational Linguistics, LACL'96*, volume 1328 of *LNCS/LNAI*. Springer-Verlag, 1997.

dans le traitement de la morphologie. Les informations qui définissent le sens sont représentées dans la structure des qualia via les prédicats qui le caractérisent. Les trois niveaux de représentation du sens sont la structure argumentale, la structure événementielle et la structure des qualia. La structure argumentale détaille et type les arguments qui sont encodés dans la structure des qualia. Elle précise non seulement le nature des arguments réalisés syntaxiquement mais également la nature et le type des arguments optionnels (D_Arguments ou arguments par défaut), et la nature et le type des arguments cachés (S_Arguments pour arguments cachés, shadow_arguments en anglais). Soit pour un élément lexical α :

$$\left[\begin{array}{c} \alpha \\ \text{Argst} = \left[\begin{array}{c} \text{Arg}_1 = x : \tau \\ \text{Arg}_2 = x' : \tau' \\ \dots \\ \text{D_Arg}_1 = y : \pi \\ \text{D_Arg}_2 = y' : \pi' \\ \dots \\ \text{S_Arg}_1 = z : \rho \\ \text{S_Arg}_2 = z' : \rho' \end{array} \right] \end{array} \right]$$

La structure événementielle adaptée de ^[1] est définie par le triplet $(E, <, *)$, dans lequel E est l'ensemble des sous événements constituant d'un élément e , ordonnés par $<$, et $*$ désigne la tête du sous-événement défini comme le sous-événement le plus saillant et qui contribue le plus à l'interprétation, la relation $*$ entre deux événements e_1 et e_2 étant définie par : $*(e_1, e_2) = [e_1, \dots, e_2 * \dots]$. Chacun des sous-événements e_i est typé selon la typologie aspectuelle de Vendler ^[2] état, procès ou transition, ce dernier type subsumant les types achèvement et accomplissement. D'où la représentation de la structure événementielle de l'élément lexical α :

$$\left[\begin{array}{c} \alpha \\ \text{Eventst} = \left[\begin{array}{c} E_1 = e_1 : \epsilon \\ E_2 = e_2 : \epsilon' \\ \text{Ordre} = < \\ \text{Tête} = e_1 * \end{array} \right] \end{array} \right]$$

La structure des qualia lie les éléments des structures événementielle et argumentale et précise leur rôle dans la force prédictive du mot. Chaque rôle ou quale Q_i formel, constitutif (qui établit les relations de méronymie), télique (qui étalite la fonction du mot) ou agentif (son mode de création) est donc encodé dans la

[1] H. Kamp and U. Reyle. *From Discourse to Logic*. D. Reidel, Dordrecht, 1993.

[2] Zeno Vendler. *Linguistics in philosophy*. Cornell University Press, 1967.

structure des qualia comme ci-dessous, R étant un prédicat qui participe au sens de l'élément lexical:

$$[\alpha \dots \text{Qualia} = [\dots [Q_i = R(e_i : \epsilon_i, Arg_i : \tau_i, Arg_j, : \tau_j \dots)] \dots] \dots]$$

L'ensemble des informations qui participent au sens et sont ainsi structurées permet d'expliciter les comportements polysémiques des items lexicaux, les propriétés de leur projection en syntaxe telles que les alternances de transitivité ^[1], et leurs propriétés compositionnelles mises en évidence par exemple dans la morphologie des composés nominaux ^[2].

3.3 Sémantique prédicative

Précisons tout de suite qu'en ce qui concerne la sémantique, nous nous limitons à des modèles formels issus de la logique et de la grammaire générative: DRT, sémantique de Montague, lexique génératif.

Avant d'éventuellement nous engager dans la représentation des connaissances d'un domaine particulier, nous préférons nous concentrer sur les problèmes que soulèvent les relations entre prédicats, entre arguments et prédicat ou entre catégories et modificateurs. Le lexique génératif permet d'exprimer les relations entre prédicats et de faire le lien avec la sémantique prédicative (Montague, DRT) De ce point de vue la sémantique de Montague est incontournable, et son grand attrait est d'être compositionnelle et de suivre la structure syntaxique des énoncés.

3.3.1 Sémantique de Montague

La sémantique de Montague ^[3,4] est tout simplement un moyen d'associer des formules du calcul des prédicats à un énoncé, à partir de son analyse syntaxique et de formules partielles associées aux mots. Pour cela on exprime le calcul des prédicats dans le lambda calcul simplement typé à deux types primitifs (e pour les entités, et t pour les valeurs de vérité) et le lexique associe à chaque mot un lambda terme décrivant son comportement.

La formalisation de la syntaxe dans les grammaires catégorielles simplifie considérablement cette correspondance, comme on le voit en figure 3.3.1.

La sémantique de Montague a immédiatement incorporé un traitement de la logique intensionnelle, pour traiter des lectures *de re* et *de dicto* de certaines expressions, ce qui ne complique pas outre mesure le lambda calcul.

(1) James Bond croit que l'un des chercheurs est un espion.

[1] Christian Bassac and Pierrette Bouillon. Middle transitive alternations in english: a generative lexicon approach. In Paul Boucher, editor, *Many Morphologies*. Cascadilla Press, 2002.

[2] Christian Bassac. *Principes de morphologie anglaise*. Linguistica. Presses Universitaires de

FIG. 4 – La correspondance syntaxe sémantique dans les grammaires de Lambek

(1) à chaque type syntaxique on associe un type sémantique:

$$\begin{aligned}
 S^\circ &= t \\
 sn^\circ &= e \\
 n^\circ &= (e \rightarrow t) \\
 (a \multimap b)^\circ &= (a^\circ \rightarrow b^\circ) \\
 (b \multimap a)^\circ &= (a^\circ \rightarrow b^\circ)
 \end{aligned}$$

(2) à chaque mot m de type syntaxique a le lexique associe un lambda terme clos $S\text{Lex}(m,a)$ de type a° , qui peut contenir des constantes exprimant les opérations logiques:

$$\begin{aligned}
 \forall, \exists &: (e \rightarrow t) \rightarrow t \\
 \wedge, \vee, \Rightarrow &: t \rightarrow (t \rightarrow t)
 \end{aligned}$$

ainsi que d'autres constantes exprimant les prédicats utilisés dans les énoncés — un prédicat à n arguments est une constante de type

$$e \rightarrow (e \rightarrow (\dots (e \rightarrow t) \dots)) \quad [n \text{ fois } e]$$

(3) l'analyse syntaxique d'un énoncé $m_1 \dots m_n$ est une démonstration en logique linéaire non commutative de $t_1, \dots, t_n \vdash S$, et en appliquant l'homomorphisme des types syntaxiques vers les types sémantiques on obtient une démonstration de $(t_1)^\circ, \dots, (t_n)^\circ \vdash S^\circ = t$ valide en logique intuitionniste, et donc un lambda terme V de type t avec des variables de type $x_1 : (t_1)^\circ, \dots, x_n : (t_n)^\circ$

(4) la représentation sémantique de l'énoncé est la forme normale de

$$(\dots ((\lambda x_1 \lambda x_2 \dots \lambda x_n V) \text{ SLex}(m_1, t_1)) \dots) \text{ SLex}(m_n, t_n)$$

qui est bien de type t , c'est-à-dire une formule du calcul des prédicats.

FIG. 5 – Pronoms, référence et quantification

Dans les exemples ci-dessous, le pronom il est référentiel dans 3, lié par un quantificateur dans 4, et ni l'un ni l'autre dans 5 :

- (3) Pierre entra dans la pièce puis **il** referma la porte.
- (4) Tout homme sait qu'**il** doit mourir.
- (5) Si un étudiant travaille, **il** a des chances de réussir.

En effet, dans ce dernier exemple il n'est pas référentiel puisque l'antécédent est un indéfini (« un étudiant ») et si d'autre part, nous représentons cet indéfini, comme il est courant, par un quantificateur existentiel, alors le pronom, ne figurant pas dans son champ, ne peut être lié par lui.

- (a) (il sait de quel chercheur il s'agit)
- (b) (il ne sait pas de quel chercheur il peut s'agir)

Par contre elle échoue à traiter des énoncés où la compositionnalité est prise en défaut, comme dans le célèbre exemple de Geach

- (2) Tout paysan qui possède un âne le bat.

3.3.2 DRT – Théorie de la Représentation du Discours

C'est entre autres pour ce genre de raisons qu'a été introduite la **théorie de la représentation du discours** (DRT) ^[1,2,4]. La DRT donne une approche dynamique du discours: chaque énoncé introduit est interprété dans le contexte fourni par les énoncés précédents et en même temps, enrichit le contexte pour les énoncés suivants. Cela permet notamment de résoudre les problèmes de liens anaphoriques, mais aussi de concordance des temps et même de présupposition.

Bordeaux, 2004.

- [3] Richmond Thomason, editor. *The collected papers of Richard Montague*. Yale University Press, 1974.
- [4] L. T. F. Gamut. *Logic, Language and Meaning*, volume 2. The University of Chicago Press, 1991.
- [1] H. Kamp and U. Reyle. *From Discourse to Logic*. D. Reidel, Dordrecht, 1993.
- [2] Jan van Eijck and Hans Kamp. Representing discourse in context. In van Benthem and ter Meulen ^[3], chapter 3, pages 179–237.
- [3] Johan van Benthem and Alice ter Meulen, editors. *Handbook of Logic and Language*. North-Holland Elsevier, Amsterdam, 1997.
- [4] Francis Corblin. *Représentation du discours et sémantique formelle. Introduction et applications au français*. Linguistique Nouvelle. Presses Universitaires de France (PUF), 2002.

La DRT se situe ainsi dans une tradition philosophique qui intègre le principe selon lequel l'interprétation d'un énoncé s'identifie avec ses conditions de vérité (sémantique vériconditionnelle) et un principe dynamique qui voit chaque énoncé comme un potentiel de changement de contexte (*context change potential*)^[1,2]

La DRT a notamment été inventée pour pouvoir étendre la sémantique compositionnelle lorsque figurent des pronoms qui n'appartiennent pas au champ de leur antécédent et ne sont pas pour autant des expressions référentielles (voir figure 3.3.2). Le problème du liage des pronoms est particulièrement patent dans les fameux exemples de Geach comme (2) où, manifestement, *il* et *le* ne réfèrent à rien de défini et où il semble difficile a priori d'établir un lien (au sens du liage des variables) entre respectivement *un fermier* et *il* et *un âne* et *le*. De fait, il semble que les indéfinis reçoivent là une interprétation universelle : pour tout fermier x et pour tout âne y , si x possède y , alors x bat y .

La DRT propose quatre principes pour résoudre ce type de problème:

- les pronoms et les indéfinis se traduisent comme de simples variables libres (et donc les indéfinis ne se traduisent pas directement par des expressions quantifiées existentiellement),
- les vrais quantificateurs peuvent lier toutes variables dans leur champ (qui coïncide avec le domaine de c-commande),
- par défaut, les variables restantes sont liées par des quantificateurs existentiels,
- les phrases conditionnelles sont traitées comme des quantifiées universelles (autrement dit « si un fermier a un âne, il le bat » comme « tout fermier qui a un âne le bat »)

Grâce à ces principes, comme on l'explique rapidement en figure 3.3.2 la phrase 2 reçoit effectivement, et de manière compositionnelle et donc calculable, l'interprétation correcte suivante :

$$(6) \quad \forall x, y (\text{fermier}(x), \text{âne}(y), \text{possède}(x,y)) \Rightarrow \text{bat}(x,y)$$

On mentionnera deux extensions actuelles de la DRT. La DRT sous-spécifiée (UDRT) introduite par Uwe Reyle^[3,4,5], qui s'appuie sur une syntaxe de type

-
- [1] Tim Fernando. Ambiguity under changing context. *Linguistics and Philosophy*, 7(20):575–606, 1997.
- [2] Tim Fernando. Ambiguous discourse in a compositional context. *Journal of Logic, Language and Information*, 1(10):63–86, 2001.
- [3] Uwe Reyle. Underspecified discourse representation structures and their logic. *Bulletin of the IGPL*, 3(2–3), 1995.
- [4] Uwe Reyle. Co-indexing labeled drss to represent and reason with ambiguities. In Kees van Deemter and Stanley Peters, editors, *Semantic Ambiguity and Underspecification*, volume 55 of *Lecture Notes*, pages 239–268. CSLI, 1996.
- [5] Annette Frank and Uwe Reyle. Principled based semantics for HPSG. In *Proceedings of EACL-95*. Association for Computational Linguistics, 1995.

grammaire d'unification et la DRT segmentée (SDRT) introduite par Nicolas Asher^[1] qui s'appuie sur une syntaxe chomskienne^[2].

La UDRT est une extension de la DRT dont le but principal est de ne pas construire simultanément toutes les interprétations possibles, qui peuvent être en grand nombre. Par exemple, en présence de n quantificateurs généralisés,

- (7) tout-le-monde(Q1) croit que la-plupart-des-politiciens(Q2) ont lu un-livre-d'économie(Q3)

le nombre de lectures possibles est de $n!$ (6 pour l'exemple proposé). L'idée est alors d'autoriser et de gérer des structures de représentations avec des informations non spécifiées comme la portée relative des quantificateurs (comme une relation acyclique partiellement définie), et de préciser cette information ultérieurement lorsque le contexte permet d'exclure certaines interprétations. Cette démarche a été mise en relation avec un mécanisme grammatical d'unification (HPSG).

LA SDRT donne une méta-structure au discours en mettant explicitement en relation des unités de discours appelés segments. Le type de relation entre segments peut être de spécialisation, de mise en parallèle, de mise en opposition, etc. A la différence de la UDRT, on choisit en l'absence de tous les paramètres une lecture donnée, quitte à revenir ensuite sur les choix effectués lorsqu'ils entrent en conflit avec des informations nouvelles.

3.3.3 Sémantique des jeux

Une autre direction qui nous permet particulièrement prometteuse est la **sémantique des jeux**^[3], qui permet aussi de traiter de tels phénomènes, et même d'autres plus subtils comme:

- (10) Un représentant de chaque université connaît un représentant de chaque EPST.

Une modélisation correcte de la sémantique de cette phrase doit faire appel aux quantificateurs de Henkin puisqu'il s'agit d'un énoncé de type $\forall u \forall e \exists r \exists r' P(u, e, r, r')$ où le r ne dépend que de u et le r' ne dépend que de e , ce que les quantificateurs usuels ne savent exprimer.

[1] Nicholas Asher. *Abstract objects in Discourse*. Kluwer Academic Publisher, 1993.

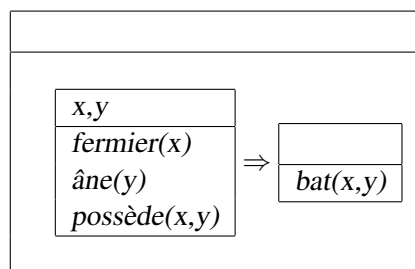
[2] Joan Busquets, Laure Vieu, and Nicolas Asher. La SDRT: Une approche de la cohérence du discours dans la tradition de la sémantique dynamique. *Verbum*, XXIII(1):73–101, 2001.

[3] Jaakko Hintikka and Gabriel Sandu. Game-theoretical semantics. In van Benthem and ter Meulen^[4], chapter 6, pages 361–410.

[4] Johan van Benthem and Alice ter Meulen, editors. *Handbook of Logic and Language*. North-Holland Elsevier, Amsterdam, 1997.

FIG. 6 – Calcul de la sémantique de 2 par la DRT

Pour calculer la sémantique de 2 l'idée est de considérer des étapes intermédiaires où les indéfinis créent des variables libres (appelées « référents de discours ») au sein de la représentation sémantique (DRS) associée à la proposition élémentaire qui les contient — ici l'antécédent de l'implication. Les variables x et y associées à un fermier et à un âne sont ensuite reprises via les pronoms au sein d'une deuxième DRS, les deux se trouvant maintenant enchâssées au sein d'une DRS plus grande. Cette imbrication associée à la forme conditionnelle est représentée par une notation en termes de « boîtes » comme sur le schéma suivant :



Les emboîtements de DRS permettent de plus d'exprimer le fait que certains référents puissent être accessibles ou non, à certains moments de l'analyse, à partir d'autres boîtes. On ne pourra par exemple pas avoir :

(8) Pierre n'a pas de voiture. *Elle est garée devant la porte.

(le référent introduit par « voiture » n'est pas accessible depuis la deuxième phrase à cause de la négation dans la première), ni (7) suivant :

(9) *Si un fermier a tout âne, il le bat.

Ce dernier type de sémantique semble bien adapté à la modélisation du dialogue, et donc du dialogue homme-machine. Le lien avec la syntaxe devrait être calculable puisqu'on sait que les logiques des ressources utilisées pour la syntaxe admettent de telles sémantiques.

3.3.4 Prise en compte de la présupposition

On ne peut traiter le discours, comme le dialogue, sans inclure une problématique de la présupposition et, plus généralement, des implicites. Nous distinguerons, comme il est d'usage désormais, les présuppositions et les implicatures. Les premières sont soumises à des critères stricts (test de la négation, de l'enchâssement dans un contexte de croyance. . .) et sont déclenchées par des items lexicaux ou des constructions grammaticales – types (phrases clivées par exemple). Les présupposés ne sont en général pas annulables. Les implicatures se divisent en implicatures conventionnelles (elles aussi attachées à l'usage de mots ou de tournures grammaticales) et implicatures conversationnelles (résultant de l'application des maximes de Grice), elles sont annulables. Si on dit par exemple :

(11) Pierre s'est risqué à chercher la solution de ce problème

le sous-entendu est que le problème était difficile. Néanmoins il est possible que la phrase se poursuive par :

(12) Le problème était en réalité facile

On a donc une implicature conventionnelle. On peut opposer cela à :

(13) Pierre est allé chercher sa femme à l'aéroport

qui présuppose que Pierre est marié et qui ne peut se poursuivre par :

(14) pourtant Pierre n'est pas marié.

L'utilisation de la logique de l'information partielle est pertinente pour étudier ce genre de question, d'abord parce que, comme l'ont fait remarquer beaucoup d'auteurs dans le passé ^[1] le phénomène de la présupposition fait intervenir la notion de phrase sans valeur de vérité définie – cas d'échec de la présupposition -, ensuite parce que cette logique permet de traiter efficacement les problèmes de défaisabilité.

[1] David Ian Beaver. Presupposition. In van Benthem and ter Meulen ^[2], chapter 17, pages 939–1008.

[2] Johan van Benthem and Alice ter Meulen, editors. *Handbook of Logic and Language*. North-Holland Elsevier, Amsterdam, 1997.

4 Objectifs

4.1 Théorie des langages et algorithmique

Ces objectifs partent du constat suivant : si les grammaires catégorielles sont très satisfaisantes parce qu’elles fournissent une interface aisée avec la sémantique, ou parce qu’elles sont apprenables à partir d’exemples, leur complexité d’analyse est un inconvénient qui nuit à toute réalisation pratique utilisant la recherche de preuve: la prouvabilité est déjà un problème NP-complet pour le calcul de Lambek ^[1], et devient un problème PSPACE-complet pour les grammaires multimodales avec un ensemble de postulats linguistiquement raisonnables ^[2]. On sait cependant qu’elles correspondent aux langages hors contexte, ou, pour certaines extensions des langages faiblement contextuels, classes de langages qui sont analysables en temps polynomial dans d’autres formalismes — qui ne présentent pas les avantages précités, sans quoi on se passerait des grammaires catégorielles! C’est pourquoi il est essentiel de mieux comprendre la correspondance avec les grammaires usuelles pour proposer des algorithmes d’analyse performants. La compréhension de ce lien peut également être utilisée dans l’autre sens, par exemple en offrant des algorithmes d’acquisition simples (basés sur l’unification) à des classes de langages contenant des langages hors contexte et même faiblement contextuels, ou en associant des représentations sémantiques aux analyses syntaxiques d’une grammaire de réécriture.

4.1.1 Algorithmes d’analyse et complexité

MAXIME AMBLARD, KIM GERDES, YANNICK LE NIR, RICHARD MOOT, CHRISTIAN RETORÉ

Tandis que les grammaires hors contexte ou faiblement contextuelles admettent des algorithmes d’analyse polynomiaux, les grammaires catégorielles, pourtant faiblement équivalentes,^[3,4] utilisent des algorithmes NP. Une meilleure compréhension des liens entre ces deux familles de formalismes devrait permettre de définir des formes normales pour les grammaires catégorielles dont l’analyse soit un problème polynomial.

-
- [1] Mati Pentus. Lambek calculus is NP-complete. Technical Report TR-2003005, CUNY - City University of New York, 2003. <http://www.cs.gc.cuny.edu/tr/>.
 - [2] Richard Moot. *Proof nets for linguistic analysis*. PhD thesis, UIL-OTS, Universiteit Utrecht, 2002.
 - [3] Yehoshua Bar-Hillel, Chaim Gaifman, and Eli Shamir. On categorial and phrase-structure grammars. *Bulletin of the research council of Israel*, F(9):1–16, 1963.
 - [4] Mati Pentus. Lambek grammars are context-free. In *Logic in Computer Science*. IEEE Computer Society Press, 1993.

Toujours par des méthodes syntaxiques on peut préférer la construction de réseaux de démonstration. Il faut pour cela étendre la définition des réseaux comme les graphes bicolores (un couplage parfait et un cographe) à la logique linéaire non commutative, et définir une axiomatisation de la logique linéaire non commutative par réécriture, comme cela a été fait dans le cas commutatif ^[1].

Une autre technique pour établir la complexité de la prouvabilité de certaines sous-classes, et donc de l'analyse syntaxique, est d'utiliser les modèles pour les logiques considérées, qui peuvent être des groupes (par exemple le groupe libre) ou l'ensemble des parties du monoïde libre.^[2]

4.1.2 Liens entre familles de grammaires

MAXIME AMBLARD, KIM GERDES, YANNICK LE NIR, ALAIN LECOMTE, RICHARD MOOT, CHRISTIAN RETORÉ

les grammaires considérées en syntaxe, sont d'une grande variété: grammaires de réécriture et transformationnelles (par exemple, grammaires minimalistes), grammaires de dépendance et grammaires catégorielles Néanmoins le lien entre ces formalismes est loin d'être maîtrisé.

Pour prendre un exemple simple, considérons la transformation d'une grammaire hors contexte en une grammaire de Lambek. Elle se fait via la forme normale de Greibach ^[4], et on obtient alors des grammaires de Lambek très particulières, dont les types sont tous du premier ordre, et les arbres d'analyse sont totalement différents.^[5,2] Réciproquement une grammaire de Lambek se transforme en une grammaire hors contexte en forme normale de Chomsky,^[6] dont on ignore la taille et le temps qu'il faut pour la construire. On remarquera que les grammaires de Lambek admettent des langages d'arbres d'analyse hors contexte tandis que les grammaires hors contexte ont toujours des langages réguliers d'arbres d'analyse.^[7]

-
- [1] Christian Retoré. Handsome proof-nets: perfect matchings and cographs. *Theoretical Computer Science*, 294(3):473–488, 2003. Complete version RR-3652 <http://www.inria.fr>.
 - [2] Wojciech Buszkowski. Mathematical linguistics and proof theory. In van Benthem and ter Meulen ^[3], chapter 12, pages 683–736.
 - [3] Johan van Benthem and Alice ter Meulen, editors. *Handbook of Logic and Language*. North-Holland Elsevier, Amsterdam, 1997.
 - [4] Sheila A. Greibach. A new normal-form theorem for context-free phrase structure grammars. *Journal of the ACM*, 12(1):42–52, 1965.
 - [5] Yehoshua Bar-Hillel, Chaim Gaifman, and Eli Shamir. On categorial and phrase-structure grammars. *Bulletin of the research council of Israel*, F(9):1–16, 1963.
 - [6] Mati Pentus. Lambek grammars are context-free. In *Logic in Computer Science*. IEEE Computer Society Press, 1993.
 - [7] Hans-Jörg Tiede. *Deductive Systems and Grammars: Proofs as Grammatical Structures*.

On aimerait donc obtenir des transformations directes, qui, si cela est possible, préservent les arbres d'analyse et qui, dans le cas contraire, expliquent comment ils se transforment. cela permettrait de compiler les grammaires catégorielles (qui ont l'avantage de mettre en rapport l'analyse syntaxique et le sens des énoncés) en des grammaires de réécritures ou transformationnelles qui admettent des algorithmes d'analyse très performants.

Le lien avec les TAGs et la grammaires de dépendance devrait aussi nous aider à clarifier les liens et les avantages et inconvénients des divers formalismes.

4.2 Modélisation linguistique

4.2.1 Morphologie et sémantique lexicale

CHRISTIAN BASSAC, PATRICK HENRY, GÉRARD HUET, RENAUD MARLET

Le Lexique Génératif est également un outil puissant et unificateur qui grâce à une riche représentation des affixes et des mots permet une modélisation à la fois de la morphologie et de la syntaxe. Ici l'objectif est de poursuivre les travaux sur la morphologie du turc et de l'anglais (noms composés, noms déverbaux, morphologie de la prédication verbale et non verbale en turc) et également d'utiliser le L.G pour la résolution d'anaphore.

4.2.2 Syntaxe et morphologie

GÉRARD HUET, ALAIN LECOMTE, RENAUD MARLET, HENRI PORTINE, CHRISTIAN RETORÉ

Si théoriquement, morphologie flexionnelle et syntaxe participent d'un même phénomène, leur traitement informatique diffère: automates d'états finis pour l'une, et grammaires au moins hors-contexte pour l'autre. Il faut cependant savoir intégrer ces deux aspects.

Les grammaires d'unification restent la solution la plus répandue, mais, à ce jour il n'y a pas d'extension satisfaisante des grammaires catégorielles qui incorpore l'unification de traits complexes.

Une autre solution est d'avoir des catégories d'accord, comme dans la grammaire générative, et cette piste n'a pas été explorée.

Quelle que soit la solution retenue, il conviendra d'articuler ces deux modules, que ce soit pour analyser ou pour engendrer des énoncés.

PhD thesis, Illinois Wesleyan University, 1999. <http://www.iwu.edu/htiede/>.

4.2.3 Grammaire de la langue des signes

PIERRE GUITTENY, OLIVIER DE LANGHE, KIM GERDES, HENRI PORTINE, CHRISTIAN RETORÉ, EMILIE VOISIN

En vue de l'application à la langue des signes, un travail de modélisation de la grammaire de cette langue est nécessaire. Bien peu d'ouvrages proposent une description synthétique de la grammaire de cette langue, et on peut espérer des retombées linguistiques importantes de nos tentatives de formalisation, notamment dans une perspective de grammaire comparée.

Une première comparaison s'impose, celle entre la langue des signes française (LSF) et la langue des signes américaine (ASL). Seule cette dernière bénéficie d'ouvrages conséquents, par exemple ^[1,2] et cela nous aidera sans doute dans notre travail de formalisation. Cette première comparaison incite à comparer les langues des signes entre elles. Ce sera notamment l'occasion de vérifier s'il est bien vrai que les langues des signes partagent une même grammaire et que seuls leurs vocabulaires diffèrent — ce qui expliquerait que les sourds de pays distincts puissent converser au bout de quelques jours.

Si l'on se fie à des travaux existants, une certaine similitude entre la grammaire des langues des signes et celles de langues non indo-européennes comme le chinois ^[3] ou le japonais est pressentie: présence de classificateurs, sujet implicite, temps implicite, aspect explicite. Dans la classification rudimentaire, mais désormais traditionnelle, des langues par la structure des phrases simples par l'ordre naturel du sujet (S), du verbe (V) et de l'objet (O), la langue des signes utilise souvent des constructions OSV (l'exemple « standard » étant les langues amazoniennes telles le Hixkaryana), qui sont rares parmi les langues humaines. On notera que dans ces constructions le verbe s'accorde souvent avec l'objet, l'objet étant repris dans le verbe par fusion avec un classificateur. Cela rapproche la langue des signes de langues méconnues d'Indonésie, selon une remarque d'Edward Keenan (UCLA).

(15) <journal> <moi> <lire+classificateur-objet-plan> <passé>
(j'ai lu le journal)

Dans une perspective universaliste qui affirme que seul le médium diffère entre les langues des signes et les langues parlées, l'un étant gestuel/visuel et l'autre

[1] Carol Neidle, Judy Kegl, Dawn MacLaughlin, Benjamin Bahan, and Robert G. Lee. *The Syntax of American Sign Language – Functional Categories and Hierarchical Structure*. MIT Press, 2000.

[2] Diane Lillo-Martin. *Universal Grammar and American Sign Language: Setting the Null Argument Parameters*. Kluwer, 1991.

[3] François-Xavier Nève. *Essai de grammaire de la langue des signes française*. Droz, 1996.

phonatoire/auditif, notre étude sera bien sûr un moyen de confirmer ou d'infirmier cette hypothèse. On pourra par la même occasion vérifier si la théorie générative universaliste s'applique à ces langues, notamment par les catégories syntaxiques qu'elle propose.

Plus précisément, les arborescences des formalismes syntaxiques que nous utilisons suggèrent une catégorisation de la langue mise en œuvre dans le regroupement syntagmatique. En particulier tous les formalismes s'appuient sur une opposition entre syntagme verbal et syntagme nominal, et on est en droit de se demander si, pour la langue des signes, cette catégorisation et cette opposition sont fondées sur des faits linguistiques, et non produites artificiellement par les formalismes, plutôt issus de l'étude des langues indo-européennes. En effet la langue des signes, comme le turc et le mongol, recourt fréquemment à des phrases nominales s'il n'y a aucune action, tandis que celles-ci sont exceptionnelles en français ou en anglais.

(16) <vélo> <sien> <là-non-déictique + classificateur>
(il a un vélo)

Tandis que les comparaisons susmentionnées devraient nous aider dans notre travail de formalisation, la langue des signes pose problème par l'absence de système d'écriture utilisés — les systèmes notationnels apparus récemment sont peu utilisés. Nous sommes donc confrontés à un corpus « oral » constitué de contacts avec des locuteurs compétents et de séquences vidéo. Or chacun sait, malgré l'accent placé depuis de Saussure sur la langue parlée, que cette dernière est bien plus irrégulière que la langue écrite, et échappe encore davantage aux modèles grammaticaux proposés.

La constitution d'un corpus pertinent pour extraire la structure grammaticale est donc problématique. On pourrait souhaiter l'émergence des structures à partir de corpus pour ne pas plaquer des structures grammaticales de langues connues sur la langue étudiée, mais surtout en l'absence de langue écrite et normée, le risque est de ne pas réussir à identifier les structures recherchées. Réciproquement, on peut partir de modèles connus les développer et voir s'ils rendent compte des corpus connus, ou sont validés par des locuteurs compétents, mais on risque alors d'obtenir des représentations peu naturelles. La bonne solution est assurément d'avancer en parallèle dans ces deux directions et de confronter les résultats. On peut alors espérer constituer un recueil d'exemples types, à partir desquels définir la grammaire puisque c'est ainsi que fonctionne habituellement la recherche en syntaxe.

Les premiers exemples que nous avons recueillis nous ont conduit, dans deux publications, à nous interroger sur le statut des structures Objet Sujet Verbe en langues des signes française. Ces structures sont rarissimes dans les langues humaines, et, par conséquent, ne sont pas prévues dans les formalismes courants.

L'ouvrage ^[1] consacré à la langue des signes américaine présente ces constructions comme des topicalisations à partir de structures Sujet Verbe Objet, mais nous ne pouvons adopter ce point de vue pour la langue des signes française, puisque certains énoncés de base n'admettent pas d'autre formulation. C'est pourquoi nous nous documentons sur les langues où ces constructions existent et sur la modélisation qu'en ont donnée les linguistes.

4.2.4 Résolution d'anaphores

ROBERTO BONATO, JOAN BUSQUETS, GÉRARD HUET, CHRISTIAN RETORÉ

L'un des sujets classiques de l'interface entre syntaxe et sémantique est la résolution des anaphores et notamment le calcul de la référence des pronoms. Cette question est assez bien modélisée par la relation d'accessibilité dans le cadre de la DRT, mais c'est alors le passage de l'analyse syntaxique aux DRS qui est moins évident. Les travaux de Reinhard Muskens sur la DRT lexicalisée et les grammaires catégorielles ^[2] semblent adaptés à la définition d'une telle interface, même si les questions qu'il a abordées concernent plutôt le temps.

Cette question du calcul de références ne se pose pas que pour les groupes nominaux, mais aussi pour d'autres constituants, notamment en présence d'ellipses verbales, un exemple classique étant le suivant (où figure aussi un pronom):

- (17) Pierre aime sa femme et Paul aussi.
(a) lecture stricte (*strict*): Paul aime la femme de Pierre
(b) lecture souple (*sloppy*): Paul aime sa femme

Pour traiter de cette question dans le cadre de la sémantique de Montague, l'algorithme d'unification d'ordre supérieur de Gérard Huet ^[3] est fréquemment

-
- [1] Carol Neidle, Judy Kegl, Dawn MacLaughlin, Benjamin Bahan, and Robert G. Lee. *The Syntax of American Sign Language – Functional Categories and Hierarchical Structure*. MIT Press, 2000.
- [2] Reinhard Muskens. Combining Montague Semantics and Discourse Representation. *Linguistics and Philosophy*, 19:143–186, 1996.
- [3] G. Huet. A unification algorithm for typed lambda calculus. *Theoretical Computer Science*, 1(1):27–57, 1973.

utilisé.^[1,2,3,4,6,7].

Un phénomène en rapport est celui de contrôle (*control*):

- (18) Je te promets de venir.
(a) (c'est moi qui viens)
- (19) Je te permets de venir.
(a) (c'est toi qui viens)

4.2.5 L'interface entre sémantique et discours

AGNÈS BRACKE, JOAN BUSQUETS, GÉRARD HUET, HENRI PORTINE

On peut également voir les phénomènes ci-dessus comme des phénomènes de discours, et traiter ainsi d'autres phénomènes en rapport, par exemple :

- l'ellipse en présence d'auxiliaires modaux :
 - (20) Il n'a pas vu ce film, mais il aurait dû.
 - (21) Il voulait résoudre la conjecture de Chomsky, mais il n'a pas pu.
- le *stripping* qui supprime tous les constituants d'une clause sauf un :
 - (22) Il aime étudier la syntaxe et aussi la sémantique.
 - (23) Il aime étudier la syntaxe mais pas la phonétique.
- les ellipses contenant des pronoms réfléchis :
 - (24) Il parle de lui-même et Marie aussi.
 - (25) Jacques a voté pour lui-même et Bernadette aussi.

-
- [1] Dale Miller and Gopalan Nadathur. Some uses of higher-order logic in computational linguistics. In *24th Annual Meeting of the Association for Computational Linguistics*, pages 247–255, 1986.
- [2] Stuart M. Shieber Mary Dalrymple and Fernando C.N. Pereira. Ellipsis and higher-order unification. *Linguistic and Philosophy*, 14:399–452, 1991.
- [3] S. Pulman. Higher-order unification and the interpretation of focus. *Linguistics and Philosophy*, 20:73–115, 1997.
- [4] Claire Gardent. Sloppy identity. In Retoré^[5], pages 188–207.
- [5] Christian Retoré, editor. *Logical Aspects of Computational Linguistics, LACL'96*, volume 1328 of *LNCS/LNAI*. Springer-Verlag, 1997.
- [6] C. Gardent. Deaccenting and higher-order unification. *J. of Logic, Language and Information*, 9:313–338, 2000.
- [7] M. Kohlhase C. Gardent and K. Konrad. Higher-order colored unification: A linguistic application. *J. of Logic, Language and Information*, 9:313–338, 2001.

Sur ces sujets la SDRT de Nicolas Asher permet d’adresser ces questions dans trois directions, comme en témoignent les travaux de Joan Busquets ^[1,2,3]

- contrôle et constructions parallèles
- résolution d’anaphores
- ambiguïtés entre lecture stricte et souple

4.2.6 Modélisation du discours et du dialogue

JOAN BUSQUETS, GÉRARD HUET, ALAIN LECOMTE

Dans les systèmes informatiques l’interaction homme-machine donne naturellement lieu à des dialogues où la bonne marche du système requiert une modélisation de la sémantique du dialogue. Plus encore que dans les questions précédemment mentionnées, le dialogue construit peu à peu un univers du discours qui fait intervenir les croyances de chacun des participants.

Le dialogue est souvent traité comme une forme particulière de discours, à laquelle on applique le même genre d’outils (par exemple la SDRT, comme dans les travaux de N. Asher, ^[4,5] ou bien la sémantique des situations de J. Barwise et R. Cooper ^[6]). Or, il se trouve que des cadres formels basés sur la notion de dialogue existent (GTS de Hintikka ^[8,9]), et permettent de rendre compte des phénomènes traités par la DRT et la théorie des situations. On peut donc se demander si ce n’est pas faire un détour inutile que ne pas utiliser ces derniers travaux dans l’étude du

-
- [1] Nicolas Asher, Daniel Hardt, and Joan Busquets. Discourse parallelism, ellipsis, and ambiguity. *Journal of Semantics*, 18(1):1–25, 2001.
 - [2] Joan Busquets. The polarity parameter for ellipsis coherence. *Grammars*, 2(2):107–125, 1999.
 - [3] Joan Busquets and Pierre Denis. L’ellipse modale en français: le cas de devoir et pouvoir. *Cahiers de grammaire*, 26, 2002.
 - [4] Nicolas Asher and Alex Lascarides. Questions in dialogue. *Linguistics and philosophy*, 21:237–309, 1998.
 - [5] Nicolas Asher, Joan Busquets, and Anne Le Draoulec. Cooperativity in dialogue. In Myriam Bras and Laure Vieu, editors, *Semantic and Pragmatic Issues in Discourse and Dialogue Experimenting with Current Dynamic Theories*, number 9 in Current Research in the Semantics/Pragmatics Interface. Elsevier, 2001.
 - [6] Jerry Sligman and Lawrence Moss. Situation theory. In van Benthem and ter Meulen ^[7], chapter 4, pages 239–309.
 - [7] Johan van Benthem and Alice ter Meulen, editors. *Handbook of Logic and Language*. North-Holland Elsevier, Amsterdam, 1997.
 - [8] Jaakko Hintikka and Gabriel Sandu. Game-theoretical semantics. In van Benthem and ter Meulen ^[7], chapter 6, pages 361–410.
 - [9] G. Sandu and A. Pietarinen. Informationally independent connectives. *Logic, Language and Computation*, 9:1–15, 2001.

dialogue. Il existe en outre des travaux appliquant les idées de la logique dialogique (Lorenzen, Lorenz, Rahman) aux dialogues argumentatifs^[1]. Nous nous proposons donc de développer dans ce projet une approche « dialogique » des phénomènes discursifs, inspirée de ces différents auteurs.

4.2.7 Présupposition et modèle partiel

JOAN BUSQUETS, ALAIN LECOMTE

En collaboration avec Areski Nait Abdallah, Alain Lecomte développe un modèle de l'implicite englobant les implicatures et les présuppositions, basé sur la notion de modèle partiel^[2]. Les connaissances d'arrière-fonds sont supposées se répartir en un noyau central (vérités « dures ») et un ensemble périphérique (vérités « softs »). Un modèle partiel est défini par un triplet (i_0, J, i_1) où i_0 et i_1 sont deux assignations partielles de valeurs de vérité, compatibles entre elles et où J est un ensemble de « justifications ». Les vérités dures sont les propositions qui reçoivent la valeur « vrai » par i_0 , les vérités « softs » étant celles qui reçoivent la valeur « vrai » par l'union de i_0 et de i_1 . L'ensemble des implicites conventionnels est défini comme l'ensemble des formules qui sont vraies au sein des classes de modèles partiels minimaux associées aux phrases énoncées. Le modèle permet de rendre compte des cas d'accommodation et de résolution par le contexte. Il reste à compléter ce modèle, notamment en le rendant dynamique et en lui adjoignant la possibilité de traiter les implicatures conversationnelles au moyen d'une modélisation appropriée des maximes de Grice.

4.2.8 Syntaxe et sémantique des pronoms

ROBERTO BONATO, JOAN BUSQUETS, CHRISTIAN RETORÉ

Du point de vue de l'interface syntaxe sémantique, les pronoms sont un véritable défi, tant au niveau sémantique, qu'on vient d'évoquer, qu'au niveau syntaxique où ils ont souvent un comportement compliqué (par exemple en français, voir figure 4.2.8). Leur modélisation est pourtant nécessaire à toute application raisonnable de traitement des langues, car ils sont omniprésents.

La résolution des anaphores évoquée ci-dessus omet généralement de prendre en compte l'un des principes universels dégagés par la grammaire générative, la *c-*

[1] D. Walton and E. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. SUNY Press, 1995.

[2] Areski Nait Abdallah. *The logic of partial information*. Monographs in Theoretical Computer Science. Springer-Verlag, 1995.

FIG. 7 – *La syntaxe des pronoms clitiques du français*

La description formelle des pronoms clitiques des langues romanes reste une tâche ardue. En effet, aucun outil, même sophistiqué de traitement du français n'intègre une modélisation correcte des pronoms clitiques et notamment de leur montée en présence d'auxiliaires modaux. En effet, ce n'est ni le clitique ni son verbe d'accueil qui détermine la montée ou non du clitique mais l'auxiliaire modal, ce qui pose problème dans une grammaire lexicalisée:

(26) *Je répare ma voiture.*

(27) *Je la répare.*

(28) *Je sais la réparer.*

(29) *Je la fais réparer.*

Une des possibilités pour modéliser proprement le comportement des clitiques consiste à utiliser les travaux de la grammaire générative sur ce sujet, puisqu'on sait représenter les grammaires minimalistes dans les grammaires catégorielles.

commande qui veut par exemple que le pronom *il* et *Jean* puissent être coréférents dans le premier énoncé et ne le puissent pas dans le second:^[1,2]

(30) Combien de livres que Jean_i aime a-t-il_i vendu?

(31) * Il_i a vendu trois livres que Jean_i aime.

C'est la position respective de *Jean* et de *il* dans l'arbre d'analyse qui permet ou interdit la coréférence. Nous souhaitons inclure dans le lambda calcul une notion d'égalité qui permette de rendre compte de ce phénomène dans la sémantique de Montague.

4.3 Objectifs de développement

4.3.1 Implantation des modèles proposés en OCaml

MAXIME AMBLARD, ROBERTO BONATO, GÉRARD HUET, YANNICK LE NIR, RICHARD MOOT

Les réalisations académiques en linguistique informatique sont le plus souvent isolées et confidentielles. Néanmoins on constate une certaine convergence et complémentarité entre les modules développés ici et là. La réalisation la plus

[1] Noam Chomsky. *Some concepts and consequences of the theory of Government and Binding*. MIT Press, Cambridge, MA, 1982.

[2] Jean-Yves Pollock. *Langage et cognition: le programme minimaliste de la grammaire générative*. Presses Universitaires de France, Paris, 1997.

importante dans notre domaine est assurément l'analyseur des grammaires multi-modales développé par Richard Moot à Utrecht, en Sicstus Prolog, mais d'autres outils ont aussi été développés: la boîte à outils "Zen" de Gérard Huet <http://pauillac.inria.fr/~huet/ZEN/> qui analyse les formes fléchies (en OCaml), construction et vérification de réseaux de démonstration ou analyses syntaxiques par Sylvain Pogodalla en OCaml, implantation des algorithmes d'inférence grammaticale par certains membres de l'ARC INRIA GRACQ (qui viennent compléter automatiquement les grammaires) en OCaml et en Java, analyseur syntaxique des grammaires d'interaction de Guy Perrier en Oz, ... Il convient de recenser ces outils, de s'assurer qu'ils puissent communiquer, d'en programmer d'autres et d'ainsi constituer une boîte à outils de linguistique informatique.

On peut s'étonner que le langage CaML ne soit pas plus souvent utilisé plus souvent lors de la réalisations d'outils de traitement des langues. En effet, CaML étant basé sur la lambda calcul il est aisé d'implanter dans ces langages les modèles linguistiques fondés sur le lambda calcul que ce soit pour la sémantique ou la syntaxe catégorielle. On explore ainsi un champ d'applications naturelles pour CaML, qui devrait avoir des retombées sur sa pratique et son enseignement : ce domaine fournit nombre d'exemples et de sujets de projet pour les modules de programmation fonctionnelle.

Mentionnons en particulier la boîte à outils « Zen » développée par Gérard Huet pour la morphologie du Sanskrit (voir ci-après); tout à fait générique, elle propose d'ores et déjà de nombreuses fonctionnalités. La compléter, en particulier par un traitement de la syntaxe, est une priorité.

4.3.2 Génération d'énoncés en langue des signes

OLIVIER DE LANGHE, PIERRE GUITTENY, HENRI PORTINE, CHRISTIAN RETORÉ, EMILIE VOISIN

À partir d'une description de la syntaxe de la langue des signes, et d'une sémantique à la Montague on peut synthétiser un énoncé ayant un sens donné par une formule du calcul des prédicats — ce qui est assez restrictif mais constitue une première étape vers la traduction automatique d'énoncés en langue des signes. Ce travail de génération est comparable à celui réalisé par Sylvain Pogodalla, mais sur une langue dont la grammaire est moins bien connue.

Cela présuppose d'avoir une description écrite d'une phrase en langue des signes, et donc de faire abstraction de l'intonation; fort heureusement, des systèmes notationnels existent, même s'ils sont peu utilisés, par exemple SignWriting <http://www.signwriting.org/> qui est issu des notations chorégraphiques et qui commence à être informatisé. On notera qu'il s'agit plus d'une transcription que d'une écriture, car elle est peu pratiquée et n'a pas ou pas en-

core donné lieu à une langue écrite plus normée et qui serait plus facile à traiter. En effet, même s'il s'agit de génération, notre travail préalable de formalisation, évoqué au paragraphe 4.2.3, serait grandement facilité par l'existence d'un corpus dans une langue normée.

À terme la collaboration avec des chercheurs ou une société spécialisée dans la synthèse d'images devrait permettre de rendre les énoncés plus « parlants » (peu de sourds savent lire cette notation pour la langue des signes).

5 Domaines d'application

Parmi les objectifs de développement mentionnés les deux derniers, qui s'appuient sur les autres, nous semblent atteindre un intérêt qui dépasse la communauté académique concernée.

5.1 Philologie du Sanskrit

GÉRARD HUET

Le développement de l'accès Internet permet aujourd'hui à beaucoup, du simple curieux au spécialiste, d'accéder au patrimoine culturel et en particulier littéraire de l'humanité. S'il existe un site très complet pour les textes latins et grecs, Perseus <http://www.perseus.tufts.edu/>, rien de tel n'existe pour le sanskrit. Perseus propose l'accès en ligne aux textes classiques: par un simple clic sur un mot, celui-ci est analysé et le système renvoie l'entrée lexicale correspondante, ses différents sens possibles, les autres textes où elle figure, etc. En poursuivant les travaux de Gérard Huet, il est envisageable de réaliser pour les philologues un site analogue pour le sanskrit, voire d'étendre les fonctionnalités de ce genre de sites.

Ce projet s'appuiera sur une collaboration qui vient de démarrer entre G. Huet et Peter Scharf, professeur de sanskrit au Department of Classics de l'Université Brown. P. Scharf a réalisé, en collaboration avec plusieurs collègues informaticiens, un lecteur de sanskrit permettant à ses étudiants de lire des textes simples à plusieurs niveaux: en écriture davanagari continue, puis en mot à mot, puis en traduction anglaise mot à mot, puis en traduction phrase à phrase. <http://cgi-user.brown.edu/Departments/Classics/Faculty/Scharf/> Ce lecteur s'appuie sur une représentation étiquetée à la main. Une première étape consistera en l'étalonnage de l'algorithme d'analyse automatique de sandhi de G. Huet sur ce corpus de référence. Une deuxième étape prévoit la mise au point d'une version robuste de cet algorithme de segmentation, apte à l'acquisition de vocabulaire lors de la lecture d'un texte pour lequel le lexique initial n'est pas

complet. Cette étape nécessitera la mise au point d'un algorithme d'inversion de formes fléchies. On disposera alors d'un assistant semi-automatique à la segmentation/étiquetage de textes sanskrits continus, permettant à un philologue de réaliser rapidement des lecteurs de corpus - et donc ultimement la mise au point d'un outil pour le sanskrit similaire au site « Perseus » pour le grec.

Un autre aspect du traitement automatique du Sanskrit est la construction d'un tree bank d'exemples du sanskrit à partir de l'analyse par Brendan Gillon, linguiste de McGill, du corpus de phrases du manuel d'Apte. ^[1,2]

5.2 Traitement Automatique de la Langue des Signes Française

HENRI PORTINE, CHRISTIAN RETORÉ, EMILIE VOISIN

Un partenariat a été annoncé entre IBM et l'Université Bordeaux 1 dont le thème est le handicap. Ce partenariat a pour premier objectif d'aider dans leurs études les étudiants handicapés en leur fournissant des aides issues des nouvelles technologies, et en leur proposant au terme de leurs études un emploi chez IBM. Il propose aussi de soutenir les recherches qui contribuent à l'intégration des handicapés.

Suite à l'appel d'offre de Bordeaux I, nous avons proposé de travailler sur le traitement automatique de la langue des signes et notre proposition est la seule à avoir passé le premier tour de sélection. À l'occasion de la 23^e école d'été de la *communication* nous avons pu rencontrer les responsables d'IBM et découvrir les produits que propose IBM pour l'aide aux sourds. Le logiciel *Les signes de Mano*, CD-ROM éducatif d'apprentissage de la langue des signes destiné aux enfants a particulièrement retenu notre attention. Il s'agit d'un dictionnaire de langue des signes mis en scène, où les phrases associées aux situations sont des séquences vidéo, commentés par une succession de signes statiques (dessinés) lesquels renvoient aux entrées du dictionnaire. Notre projet de synthèse de phrases en langue des signes trouverait assurément sa place dans ce didacticiel, en remplaçant les séquences vidéos par des phrases synthétisées à partir des signes contenus dans le dictionnaires.

Cette application semble réalisable car il s'agit de phrases simples isolées et non de textes entiers, et certaines sociétés^b savent synthétiser des images lisibles

b. Par exemple les sociétés *Nuances du Sud* (Toulouse) <http://www.nuances-du-sud.fr/> et *InSite* (Lille) <http://www.insite.fr/> qui participent à la réalisation du site Web-

[1] Brendan S. Gillon. Word order in classical Sanskrit. *Indian Linguistics*, 57(1):1-35, 1996.

[2] V.S. Apte. *The Student's Guide to Sanskrit Composition – A Treatise on Sanskrit Syntax*. Poona, India, 1885.

par les locuteurs de la langue des signes, pour peu que la description de la phrase signée soit exprimée dans un langage intermédiaire accessible et à la grammaire et à la synthèse d'images.

Cette application fait l'objet d'un projet du LaBRI déposé par notre équipe auprès de la région Aquitaine: *Traitement Informatique de la langue des signes française: génération automatique d'énoncés en LSF*.

6 Principales collaborations

6.1 Nationales

6.1.1 Collaborations Bordelaises

Les membres de SIGNES appartiennent à trois équipes de recherche bordelaises: le Laboratoire Bordelais de Recherche en Informatique UMR 5800 CNRS (Patrick Henry, Richard Moot, Christian Retoré, Maxime Amblard), l'Équipe de Recherche en Syntaxe et Sémantique UMR 5610 CNRS (Christian Bassac, Joan Busquets, Kim Gerdes), et l'équipe Textualisation, Langage et Cognition, JE 2385 (Henri Portine, Pierre Guitteny, Emilie Voisin).

Outre les personnes du projet, nous sommes amenés à collaborer plus ou moins étroitement avec d'autres chercheurs de ces équipes sur certains sujets. Certaines thématiques du LaBRI retiennent plus particulièrement notre attention: la théorie des langages formels, et en particulier les automates à piles de piles (Séverine Frantani, Géraud Sénizergues), la synthèse d'images pour nos applications à la langue des signes (Jean-Baptiste Leproux) ainsi que la vérification en Coq des propriétés de certains modèles linguistiques (Pierre Castéran, Paul Gloess, Houda Anoun). Certains membres de l'ERSS et de TELANCO ont des intérêts communs avec SIGNES: syntaxe (Claude Muller), sémantique (Frédéric Lambert, Claude Muller), énonciation et pragmatique (Bettina Derville, Jeanine Gerbault, Catherine Moreau).

6.1.2 INRIA

Calligramme (Lorraine) Philippe de Groote, François Lamarche, Jean-Yves Marion, Guy Perrier, Sylvain Pogodalla — Pour des raisons historiques évidentes (Christian Retoré a participé à la création du projet Calligramme, Alain Lecomte en a été membre, Sylvain Pogodalla a fait sa thèse sous la direction d'Alain Lecomte et Christian Retoré) le projet que nous proposons partage certains centres d'intérêts avec Calligramme: l'utilisation des grammaires catégorielles et des techniques logiques pour le traitement des langues. Il y a avec Calligramme une collaboration de longue date qui est profitable.

Il y a cependant des différences:

- Calligramme est essentiellement consacré, en ce qui concerne le traitement des langues, à leur syntaxe, tandis que le projet Signes propose un traitement de la morphologie et se concentre sur l'interface syntaxe sémantique (qui fera jouer à la sémantique un rôle central).

- Parmi les formalismes utilisées, les grammaires génératives sont présentes dans Signes et sont un des points forts du Labri.
- Signes inclut des linguistes qui apporteront leur connaissance des langues et des théories linguistiques, en particulier de la grammaire générative et de la sémantique générative.
- Les deux applications que nous proposons ne recoupent pas celles proposées par Calligramme : traitement de la langues des signes, et réalisation d'un corpus navigable du Sanskrit.

Langue et Dialogue (Lorraine) Patrick Blackburn, Claire Gardent — Les sémantiques logiques sont aussi à l'honneur dans ce projet, mais leur travail est purement sémantique et n'inclut pas une automatisation de l'interface syntaxe/sémantique.

Atoll (Rocquencourt) Pierre Boullier — Les travaux de Pierre Boullier sur l'analyse syntaxique pour les langages faiblement contextuels sont tout à fait pertinents pour notre action, mais là encore c'est l'accent placé sur l'interface syntaxe/ sémantique qui distingue notre action de ce projet.

Logical (Futurs) et Lemme (Sophia) Ces projets développent *Coq* et des outils associés (notamment de puissantes interfaces graphiques), ce qui devra être l'occasion d'une collaboration fructueuse en ouvrant un nouveau champ d'expérimentation à *Coq* et ses outils associés.

6.1.3 Autres

CLIPS-IMAG Cette collaboration concerne notamment l'aspect « dialogue » : analyse de corpus de dialogues basée sur la DRT et la SDRT. Alain Lecomte participe au séminaire et à l'encadrement d'étudiants dans cette équipe.

IRIT et ERSS, CNRS, Toulouse Myriam Bras, Andreas Herzig, Philippe Muller — Représentation du discours et du dialogue.

IRISA, CNRS & Université Rennes 1 Annie Foret — Ses travaux sur l'acquisition des grammaires catégorielles s'inscrivent dans la même ligne que celle que nous proposons sur ce sujet.

IRIN, Université de Nantes Alexandre Dikovsky — Le travail effectué par Dikovsky sur les grammaires de dépendances montre clairement leur lien avec les grammaires catégorielles et aussi avec les grammaires minimalistes, notamment par le biais des grammaires de dépendance aux dépendances non bornées et des grammaires catégorielles avec itérateurs qu'il a proposées.

LIFL, CNRS & Université Lille 3 Isabelle Tellier, Alain Terlutte, Marc Tommasi —
L'équipe Grappa partage tout à fait notre approche de l'inférence grammaticale pour la construction automatique de grammaires.

Institut Jean Nicod, CNRS & EHSS & ENS Claire Beyssade — Nous partageons un intérêt commun pour la formalisation logique de la sémantique du discours, que ce soit dans le cadre de la DRT ou de la sémantique des jeux, et en particulier un intérêt pour un traitement syntaxique et sémantique des pronoms.

LATTICE-TALANA, CNRS & ENS & Université Paris 7 Pascal Amsili, Sylvain Kahane, Laurent Roussarie — grammaires de dépendance; DRT et génération de texte

MSH de Lille Fabienne Blaise, Shadi Rahman — sémantique des jeux

6.2 Internationales

OTS, Utrecht, Pays-Bas Michael Moortgat, Raffaella Bernardi, Christophe Costa-Forêncio, Richard Moot, Willemijn Vermaat — Cette équipe est l'équipe phare sur les grammaires catégorielles depuis des années, et c'est pourquoi nous entretenons avec elle des rapports étroits. Notons que nous accueillons un membre de cette équipe, Richard Moot comme ATER, et qu'une action intégrée sur *Grammaire générative et système déductifs pour le traitement de la syntaxe et de la sémantique des langues* a été acceptée pour l'année 2004.

UPC, Barcelone, Espagne Glyn Morrill — Glyn Morrill est l'un des acteurs principaux des grammaires de types logiques avec lequel nous sommes régulièrement en contact.

Göteborg, Suède Aarne Ranta — Le "Grammatical Framework" d'Aarne Ranta, lequel adapte à la linguistique les systèmes de types dépendants tels que la théorie constructive des types de Martin-Löf est un formalisme prometteur intermédiaire entre les structures de déduction logique et les arbres d'analyse syntaxique. Aarne Ranta connaît Gérard Huet et Christian Retoré depuis des années, et a adapté à Haskell des algorithmes du premier.

UTA, Austin, États-Unis Nicolas Asher — Joan Busquets a fait un Post-Doc à Austin et a écrit des articles avec Nicolas Asher sur la DRT et ses raffinements; cette collaboration se poursuit.

UCLA, Los Angeles, États-Unis Edward Stabler, Edward Keenan — C'est le modèle formel du programme minimaliste formalisé par Stabler en termes de grammaires d'arbres qui a permis de faire le lien entre les théories chomskyennes de la syntaxe et les grammaires catégorielles, et depuis plusieurs années nous entretenons une collaboration scientifique suivie. Les travaux qu'il mène avec Edward Keenan sur les universaux linguistiques nous concernent également.

Harvard University, États-Unis Pr Stuart Shieber (Computer Science) Pr Michael Witzel (Chair, Dept of Sanskrit) — avec lesquels Gérard Huet entretient une correspondance suivie.

Brown University, États-Unis Pr Peter Scharf (Dept of Humanities) — est un spécialiste de la grammaire de Pāṇini qui développe des outils informatiques d'analyse de corpus étiqueté du sanskrit.

Université d'Hyderabad, Andhra Pradesh, Inde Pr Narayan Murthy — Ce professeur est l'un des principaux spécialistes de l'informatisation des langues indiennes. Il est le correspondant indien de l'équipe associée franco-indienne en Traitement Informatique de la Langue dont Gérard Huet est le correspondant français.

Technion, Haifa, Israël Nissim Francez — Nissim Francez travaille sur les grammaires catégorielles, et notamment sur leurs aspects sémantiques, ce qui nous rapproche, et de plus un livre de cours sur les grammaires catégorielles est en cours de rédaction avec C. Retoré.

7 Contrats

Notre équipe a proposé une collaboration avec IBM France sur le traitement de la langue des signes dans le cadre du partenariat entre IBM et l'Université Bordeaux I sur l'intégration du handicap. Celle-ci a passé une première sélection.

Un projet du LaBRI a été déposé par notre équipe auprès de la région Aquitaine: *Traitement Informatique de la langue des signes française: génération automatique d'énoncés en LSF*.

8 Actions nationales et internationales (2000-2004)

8.1 Actions nationales

- L'équipe Signes fait partie du GDR *Sémantique et Modélisation* (Resp. Francis Corblin, Université Paris 4), et cela concerne plus particulièrement Joan Busquets, Alain Lecomte et Christian Retoré.
- Joan Busquets est membre du projet Cognitique *Planification d'actions : de la validation psychologique et linguistique à la programmation d'agents autonomes en intelligence artificielle* (resp. Andreas Herzig et Jérôme Lang, IRIT, Toulouse)
- Christian Retoré est coordinateur de l'Action de Recherche Coopérative de l'INRIA GRACQ *Acquisition de grammaires catégorielles*. 2001 & 2002

8.2 Actions internationales

- L'équipe Signes a obtenu une action intégrée avec OTS (Utrecht) intitulée *Grammaire générative et système déductifs pour le traitement de la syntaxe et de la sémantique des langues* à compter je janvier 2004.
- L'équipe Signes est un *active node* de l'*Excellence Network on Computational Logic, section Computational Logic for Natural Language Processing*.
- Gérard Huet est le correspondant français de l'équipe associée franco-indienne en Traitement Informatique de la Langue <http://pauillac.inria.fr/~huet/FIRNCL/>.

9 Diffusion de résultats

9.1 Animation de la Communauté scientifique (2000-2002)

9.1.1 Distinctions et nominations

- Gérard Huet est membre de l'Académie des Sciences depuis novembre 2002.
- Gérard Huet préside le comité d'évaluation de l'UMR 7126 du C.N.R.S *Preuves, programmes et systèmes*.
- Gérard Huet est l'une des personnalités extérieures du Conseil d'Administration de l'Université Paris 7.
- Henri Portine est membre élu du Conseil d'Administration de l'Université Bordeaux 3.

9.1.2 Participation à des comités de rédaction

- Alain Lecomte fait partie du comité de rédaction de la revue *TAL – Traitement Automatique des Langues*, Editions Hermès, Paris depuis 2001.
- Alain Lecomte et Christian Retoré font partie du comité de rédaction de la collection *Research in Logic and Formal Linguistics*, Edizione Bulzoni, Roma, depuis 1999.
- Henri Portine fait partie du comité de rédaction de la revue *ALSIC – Apprentissage des Langues et Systèmes d'Information et de Communication*
- Christian Retoré est rapporteur pour les *Mathematical Reviews* depuis 2003.
- Christian Retoré est rédacteur en chef de la revue *TAL – Traitement Automatique des Langues*, Editions Hermès, Paris depuis 2004 — il est membre du comité de rédaction depuis 2001.

9.1.3 Participation à des comités de lecture de colloques

- Christian Bassac a été membre du comité de lecture du 2^e colloque international *Generative approaches to the lexicon, GL'2003* (Genève).
- Christian Bassac a été membre du comité de lecture du 3^e colloque international *Generative approaches to the lexicon, GL'2005* (Genève).
- Roberto Bonato a été membre du comité de programme de la session étudiante *Logic and Language* d'ESSLLI 2003 (Vienne).
- Roberto Bonato a fait partie du comité de programme de la session étudiante d'EACL 2003 (Bucarest).

- Gérard Huet fait partie du Comité de programme de la conférence *LEC-2002 (Language Engineering Conference, Hyderabad, India, December 2002)* <http://www.LanguageTechnologies.ac.in/lec2002.html>
- Alain Lecomte et Christian Retoré ont fait partie du comité de programme de la 8^e conférence *Mathematics of Language*, Bloomington, juin 2003.
- Christian Retoré a été *area chair* de la session étudiante *Logic and Language* d'ESSLLI 2003.
- Christian Retoré a fait partie du comité de programme du workshop de LREC 2004 *Sign Language Processing* Lisbonne, mai 2004.
- Christian Retoré a fait partie du comité de programme de *Categorial Grammar 2004*, Montpellier, juin 2004.
- Christian Retoré a été membre du comité de programme de ESSLLI 2005 (Edinburgh) section *Logic and Language*.
- Christian Retoré a fait partie du comité de programme de la conférence *Tree Adjoining Grammars and related formalisms, TAG+5* en 2000.

9.1.4 Implication dans la *Fundation of Logic, Language and Information*

- Pierre Castéran et Richard Moot ont assuré le cours *Proof Automation for Type-Logical Grammars* à la *16th European Summer School in Logic, Language and Information* (Nancy, 2004)
- Philippe de Groote et Christian Retoré ont assuré les cours (15h) *Proof-theoretical methods in computational linguistics I and II* à la *15th European Summer School in Logic, Language and Information* (Wien, 2003)
- Gérard Huet et Alain Lecomte ont assuré le cours (7h30) *Survey of a few Fundamental Representation Structures for Computational Linguistics* à la *14th European Summer School in Logic, Language and Information* (Trento, 2002)
- Christian Retoré a assuré le cours (7h30) *The Logic of Categorial Grammar* à la *12th European Summer School in Logic, Language and Information* (Birmingham, 2000)
- Christian Retoré fait partie du comité de sélection du *E. W. Beth dissertation prize* anciennement *Best PhD award of the Fundation of Logic Language and Information* depuis 1999.

9.1.5 Autres interventions

- Christian Retoré a fait un cours (6h) *Predicative and syntactic structures of sentences: towards a computational correspondence* à l'École d'Automne de Linguistique EALING (École Normale Supérieure, Paris, septembre 2003).

9.2 Participation à des jurys de thèse et d'habilitation (2000-2003)

9.2.1 Habilitations à Diriger des Recherches

- Gérard Huet a été rapporteur de l'habilitation à diriger des recherches de Claire Gardent *Inférence et traitement automatique des langues*, soutenue à Nancy en septembre 2004.
- Gérard Huet a été rapporteur de l'habilitation à diriger des recherches de Guy Perrier *Grammaires d'interactions*, soutenue à Nancy en novembre 2003.
- Gérard Huet a été rapporteur de l'habilitation à diriger des recherches de Christian Retoré *Logique linéaire et syntaxe des langues*, soutenue à Nantes en janvier 2002.

9.2.2 Thèses

- Gérard Huet a été rapporteur et membre du jury de la thèse de Sylvain Pogodalla *Réseaux de preuve et génération pour les grammaires de types logiques* soutenue le 27 septembre 2001 à l'INPL, Nancy.
- Christian Retoré a été membre du jury de cette même thèse.
- Alain Lecomte est rapporteur de la thèse de Jeanne Villaneau *Contribution au traitement syntaxico-pragmatique de la langue naturelle parlée : approche logique pour la compréhension de la parole* soutenance prévue le 4 décembre 2003 à l'Université de Bretagne-Sud (Vannes).
- Christian Retoré a été rapporteur et membre du jury de la thèse de Quintijn Puite *Sequents and link graphs – contraction criteria for refinements of multiplicative linear logic* (Mathematisch Instituut, Universiteit Utrecht, 26 januari 2001)
- Christian Retoré a été rapporteur de la thèse de Christophe Costa-Florêncio *Learning categorial grammars* soutenance prévue à Utrecht le 14 novembre 2003.
- Christian Retoré a été rapporteur et membre du jury de la thèse de Daniela Dudau-Sofronie, TITRE Université Lille 1, avril 2004.
- Christian Retoré a été rapporteur de la thèse Fanch Lejeune, TITRE, Université Paris-Sud, septembre 2004.

9.3 Encadrement (2000-2003)

9.3.1 Maîtrise

- Christian Bassac a encadré le TER de maîtrise de Florence Fariello *Sur l'opposition Argument vs Circonstant* Université de Bordeaux 3 (2002).
- Christian Bassac a encadré le TER de maîtrise de Simon Mille *Les propriétés du gérondif* Université de Bordeaux 3 (2003).
- Jean Caelen et Alain Lecomte ont encadré le mémoire de maîtrise de Anne Gerard-Xuereb *Analyse et représentation du discours dans le cadre du dialogue homme-machine*, soutenu en juin 2003 à l'université Grenoble 3.
- Christian Retoré a encadré, avec Alexandre Dikovsky, deux TER de maîtrise sur l'acquisition de grammaires catégorielles (2002). Université de Nantes.
- Joan Busquets, Henri Portine et Christian Retoré ont encadré le mémoire de maîtrise d'Émilie Voisin *Dialogue, planification, logique: un état de l'art suivi de quelques propositions*. (2003) Université Bordeaux 3.
- Christian Retoré a encadré le TER de maîtrise de Thomas Poussevin, Jean-François Deverge, Fahd Haiti et Anthony Herbe *Traitement automatique des langues: analyse syntaxique pour les grammaires catégorielles* Université Bordeaux 1.

9.3.2 DEA

- Pierre Castéran a encadré le mémoire de DEA d'Houda Anoun *Atelier logique sur le calcul de Lambek*
- Henri Portine encadre le DEA d'Olivier De Langhe, *Grammaire comparée de la langue des signes française et du japonais* 2002 / 2004, Université Bordeaux 3.
- Henri Portine a encadré le DEA de Pierre Guitteny *Enseignement de la grammaire du français aux locuteurs de la langue des signes* Université Bordeaux 3, juin 2003.
- Henri Portine a encadré le DEA d'Émilie Voisin *Modélisation et Analyse des dialogues de théâtre: la problématique de la révision des croyances*, Université Bordeaux 3, juin 2004.
- Alain Lecomte et Christian Retoré ont encadré le mémoire de DEA de Maxime Amblard *Implémentation de représentations sémantiques dans les grammaires minimalistes* Université Bordeaux 1.
- Christian Retoré a encadré le mémoire de DEA de Bart Georges, *Réseaux de Petri, logique linéaire et grammaires formelles*, Université Bordeaux 1, juin 2004.

- Christian Retoré a encadré le mémoire de DEA d’Emmanuel Daviaud, *Logique d’ordre supérieur avec égalité dans un cadre compositionnel et intentionnel*, Université Bordeaux 1, septembre 2004.

9.3.3 Thèses

- Joan Busquets co-encadre avec Andrée Borillo la thèse de Laurent Prévot *Une approche sémantique et pragmatique de dialogue, le rôle des topiques dans les dialogues d’explication d’itinéraire* IRIT et ERSS, Université Paul Sabatier et Le Mirail, Toulouse.
- Alain Lecomte et Christian Retoré ont encadré la thèse de Sylvain Pogodalla *Réseaux de preuve et génération pour les grammaires de types logiques* soutenue le 27 septembre 2001.
- Alain Lecomte et Christian Retoré co-encadrent la thèse de Maxime Amblard *calcul de représentations sémantiques dans les grammaires minimalistes*, bourse MENRT, Bordeaux 1, inscription en octobre 2003.
- Henri Portine a encadré la thèse de Agnès Bracke *Deixis et modes de structuration des concepts*, soutenue le 2 décembre 2002.
- Henri Portine encadre la thèse de Yoann Debuys *La métaphore : approches linguistique et cognitive* (inscription en octobre 2002).
- Henri Portine a encadré la thèse de Malika Kaheraoui *Représentation de la cause en français : Connecteurs de causalité et verbes causatifs*, soutenue le 20 décembre 2002.
- Henri Portine encadre la thèse de Pierre Guitteny *Le passif en langue des signes française* (inscription en octobre 2003).
- Henri Portine encadre la thèse de Xavier Lavry *La constitution d’un environnement hypermédia en didactique des langues* (inscription en octobre 2001).
- Henri Portine a co-encadré la thèse d’état de Abderrahmane Amsidder (Agadir) sur l’anaphore (soutenue en 2002), et la thèse de Sonia Berbinschi (Bucarest) sur l’antonymie discursive (soutenue en 2003).
- Christian Retoré co-encadre avec Annie Foret (Université de Rennes 1) la thèse de Yannick Le Nir *Logique et grammaires formelles appliquées au traitement des langues*. Bourse MENRT, IRISA, Rennes puis ATER Université Bordeaux 1 (inscription en octobre 1999). Soutenance prévue en décembre 2003.
- Christian Retoré co-encadre avec Alexandre Dikovsky (Université de Nantes) la thèse d’Erwan Moreau *Acquisition de grammaires de dépendance et de grammaires catégorielles*. Bourse MENRT, IRIN, Nantes. (inscription en octobre 2001)

- Christian Retoré co-encadre avec Denis Delfitto (Università di Verona) la thèse de Roberto Bonato *Algorithmes de calcul de représentations sémantiques à partir d'analyses syntaxiques de type générativiste et algorithmes inverses* (inscription en octobre 2002)

9.4 Enseignements en rapport en 2002/2003 et 2003/2004 à Bordeaux

9.4.1 2002/2003

5^e année

- Gérard Huet et Christian Retoré ont assuré 8 heures sur *Introduction au lambda calcul et aux systèmes déductifs* dans le module *Informatique Fondamentale* du DEA d'Informatique de Bordeaux I.
- Gérard Huet et Christian Retoré ont assuré une option de 24 heures intitulée *Traitement symbolique des langues* dans le DEA d'Informatique de Bordeaux I.
- Henri Portine assure le cours *Énonciation et sémantique* du DEA de Lettres option Sciences du Langage de l'université Bordeaux III.
- Henri Portine a fait une intervention dans le DEA de Linguistique théorique, descriptive et automatique, Université Paris 7: *Sur quel mode penser le temps dans sa représentation linguistique? Phénoménisme, géométrie et cognition*

4^e année

- Joan Busquets a assuré le cours de *Pragmatique* en Licence de Sciences du Langage celui d'*Analyse du discours et du dialogue* en Maîtrise de Sciences du Langage à Bordeaux III.
- Henri Portine et Christian Retoré ont assuré 26 heures de cours sur la *Syntaxe formelle* en Licence de Sciences du Langage à Bordeaux III.

9.4.2 2003/2004

doctorat

- Christian Retoré assure le cours de 10h pour doctorants de Bordeaux 1: *Traitement Automatique du Langage Naturel: un panorama*

5^e année

- Christian Bassac a assuré le cours d'agrégation *The syntax of wh clauses and extraction*.
- Christian Retoré assure une option de 24 heures intitulée *Traitement symbolique des langues* dans le master Sciences et Technologie spécialité Informatique de l'Université Bordeaux I.
- Henri Portine assure le cours *Énonciation et sémantique* du DEA de Lettres option Sciences du Langage de l'université Bordeaux III.

4^e année

- Christian Retoré assure le cours (24h) et les TD/TP (24h) du master d'informatique première année *Introduction au Traitement Automatique des Langues*
- Joan Busquets assure le cours *Des conventions aux maximes: actes de langage et maximes conversationnelles* de la maîtrise de Sciences du langage de l'Université Bordeaux III.
- Kim Gerdes assure le cours *L'ordre libre des mots et sa formalisation* de la maîtrise de Sciences du langage de l'Université Bordeaux III.

3^e année

- Christian Bassac a assuré le cours de linguistique anglaise *Le gérondif: syntaxe et morphologie*
- Joan Busquets assure les cours *Grammaires et automates* et *Pragmatique* de la licence de Sciences du langage de l'Université Bordeaux III.
- Kim Gerdes assure les cours *Informatique appliquée à la linguistique* de la licence de Sciences du langage de l'Université Bordeaux III.

9.5 Publications des membres du projet (2000-2003)

Les publications de Renaud Marlet ne sont pas mentionnées car jusqu'ici il travaillait dans un autre domaine, l'évaluation partielle et spécialisation de programmes. Même chose pour Patrick Henry qui, ces dernières années, était administrateur du Gréco de programmation puis du LaBRI, et dont les recherches auparavant portaient sur les types abstraits et le traitement automatique des langues.

9.5.1 DEA, thèses et habilitations à diriger des recherches

- DEA** – Maxime Amblard. Représentations sémantiques pour les grammaires minimalistes. Mémoire de D.E.A, Université Bordeaux 1, juin 2003.
<http://www.labri.fr/Recherche/LLA/signes>

- Houda Anoun. Réalisation d'un atelier logique sur le calcul de Lambek. Mémoire de dea, Université Bordeaux 1 et ENSEIRB, 2003. <http://www.labri.fr/Recherche/LLA/signes>
- Roberto Bonato. Uno studio sull'apprendibilità delle grammatiche di Lambek rigide — a study on learnability for rigid Lambek grammars. Tesi di Laurea & Mémoire de D.E.A, Università di Verona & Université Rennes 1, 2000.
- Pierre Guitteny. L'enseignement de la grammaire du français en langue des signes. Mémoire de DEA, Université Bordeaux 3, juin 2003.
- Thèses**
 - Agnès Bracke. *Deixis et modes de structuration des concepts*. Thèse de doctorat, Université Bordeaux 3, décembre 2002.
 - Kim Gerdes. *Topologie et grammaires formelles de l'allemand*. Thèse de doctorat, Université Paris 7, 2002.
 - Yannick Le Nir. *Structure des analyses catégorielles – application à l'inférence grammaticale*. Thèse de doctorat, Université Rennes 1, décembre 2003.
 - Richard Moot. *Proof nets for linguistic analysis*. PhD thesis, UILOTS, Universiteit Utrecht, 2002.
- HDR**
 - Christian Retoré. *Logique linéaire et syntaxe des langues*. Mémoire d'habilitation à diriger des recherches, Université de Nantes, Janvier 2002.

9.5.2 Volumes édités

- Philippe de Groote, Glyn Morrill, and Christian Retoré, editors. *Logical Aspects of Computational Linguistics, LACL'2001*, number 2099 in LNCS/LNAI. Springer-Verlag, 2001.
- C. Retoré and E. Stabler, editors. *Special Issue on Resource Logics and Minimalist Grammars*, volume 2(1) of *Journal of Research on Language and Computation*. Kluwer, 2004.

9.5.3 Livres

- Christian Bassac. *Principes de morphologie anglaise*. Linguistica. Presses Universitaires de Bordeaux, 2004.
- Christian Retoré. The logic of categorial grammars - ESSLLI Lecture Notes 100pp <http://www.labri.fr/Recherche/LLA/signes/>

9.5.4 Articles et chapitres de livre

- Jean Albrespit and Henri Portine. Décrire les termes grammaticaux dans les dictionnaires monolingues et bilingues, 2002. À paraître.
- Nicolas Asher, Daniel Hardt, and Joan Busquets. Discourse parallelism, ellipsis, and ambiguity. *Journal of Semantics*, 18(1):1–25, 2001.
- Christian Bassac. Sur quelques aspects de la composition nominale en turc: le point de vue du lexique génératif. *Cahiers de grammaire*, 28:197–215, 2003. Numéro spécial Morphologie et Lexique.
- Raffaella Bernardi and Richard Moot. Generalized quantifiers in declarative and interrogative sentences. In Johan Bos and Michael Kohlhase, editors, *Proceedings of Inference in Computational Semantics 2*, pages 1–15, Dagstuhl, Germany, 2000. To appear in *Language and Computation*.
- Joan Busquets and Pierre Denis. L’ellipse modale en français: le cas de *devoir* et *pouvoir*. *Cahiers de grammaire*, 26, 2002.
- Joan Busquets, D. A. Koike, and Vann R. E. Spanish *no*, *sí*: Reactives moves to face-threatening acts. discourse relations and cognitive states part I. *Journal of Pragmatics*, 33(5):701–723, 2001.
- Joan Busquets, Laure Vieu, and Nicolas Asher. La SDRT: Une approche de la cohérence du discours dans la tradition de la sémantique dynamique. *Verbum*, XXIII(1):73–101, 2001.
- D. A. Koike, R. E. Vann, and Joan Busquets. Spanish *no*, *sí*: Reactives moves to face-threatening acts. discourse relations and cognitive states. *Journal of Pragmatics*, 33(6):879–899, 2001.
- Yannick Le Nir. From proof trees in Lambek calculus to Ajdukiewicz Bar-Hillel elimination binary trees. *Research on Language and Computation*, 1(3-4):181–201, 2003. special issue in *Trees in Linguistics, Logic, and Computer Science* (V. Gorenko, ed).
- Alain Lecomte. Partial proof-nets, hybrid logics and minimalist representations. *Grammars*, 3(2-3):243–257, 2001.
- Alain Lecomte. Rebuilding the minimalist program on a logical ground. In C. Retoré and E. Stabler, editors, *Special Issue on Resource Logics and Minimalist Grammars*, volume 2(1) of *Journal of Research on Language and Computation*, pages 27–55. Kluwer, 2004.
- Richard Moot and Mario Piazza. Linguistic applications of first order multiplicative linear logic. *Journal of Logic, Language and Information*, 10(2):211–232, 2001.
- Richard Moot and Quintijn Puite. Proof nets for the multimodal Lambek calculus. In Wojciech Buszkowski and Michael Moortgat, editors, *Studia Logica*. Kluwer Academic Publishers, 2002. To appear.

- Henri Portine. Le(s) conditionnel(s) et le schizo : De la mise en texte. *Travaux de Linguistique*, pages 85–106, 2000.
- Henri Portine. Notice sur *The Meaning of Meaning* de C. K. Ogden et I. A. Richards (1923). *Langage*, 2000. Hors Série n° 3: Corpus représentatif des grammaires et des traditions linguistiques, tome 2, Histoire - Epistémologie - Langage.
- Henri Portine. Grammaticalisation de l’altérité spatiale : D’ailleurs et l’assertion allocentrée. *Travaux linguistiques du CERLICO*, 14:247–266, 2001.
- Henri Portine and André Rousseau. Tesnière et la syntaxe structurale. *Modèles linguistiques*, XXIII(2):99–121, 2002.
- Henri Portine. La syntaxe de Damourette et Pichon comme outil de représentation du sens. *Modèles linguistiques*, XXIII(2):21–46, 2002.
- Henri Portine. Les modalités: entre logique et langue. *Revue belge de philologie et d’histoire*, 2004. À paraître.
- Christian Retoré. Systèmes déductifs et traitement des langues: un panorama des grammaires catégorielles. *Technique et Science Informatiques*, 20(3):301–336, 2000. Numéro spécial *Traitement Automatique du Langage Naturel* sous la direction de D. Kayser et B. Levrat. Version préliminaire RR-3917 <http://www.inria.fr/>.
- Christian Retoré. Handsome proof-nets: perfect matchings and cographs. *Theoretical Computer Science*, 294(3):473–488, 2003. Complete version RR-3652 <http://www.inria.fr/>.

9.5.5 Communications à des congrès, colloques, etc.

- Maxime Amblard, Alain Lecomte, and Christian Retoré. Syntax and semantics interacting in a minimalist theory. In Denys Duchier, editor, *Perspects and Advances in the Syntax/Semantics Interface*, pages 17–22, Nancy, 2003.
- Nicolas Asher, Joan Busquets, and Anne Le Draoulec. Cooperativity in dialogue. In Myriam Bras and Laure Vieu, editors, *Semantic and Pragmatic Issues in Discourse and Dialogue Experimenting with Current Dynamic Theories*, number 9 in Current Research in the Semantics/Pragmatics Interface. Elsevier, 2001.
- Christian Bassac. Construction des composés endocentriques. In *Journées de linguistique anglaise à Bordeaux*, 2003.
- Christian Bassac. Compositionality and nominal compounds. In *Journées de linguistique anglaise à Toulouse*, 2004.
- Christian Bassac and Pierrette Bouillon. The telic relationship in french and turkish compounds. In Pierrette Bouillon and Kyoko Kanzaki, editors, *First international conference on generative approaches to the lexicon*, 2001.

- Christian Bassac and Pierrette Bouillon. Middle transitive alternations in english: a generative lexicon approach. In Paul Boucher, editor, *Many Morphologies*. Cascadilla Press, 2002.
- Christian Bassac and Pierrette Bouillon. The telic relationship in compounds. In James Pustejovsky, editor, *Generative approaches to the lexicon — selected papers*. Kluwer Academic Publisher, 2005. To appear.
- Christian Bassac and Mehmet Çiçek. Les noms déverbaux dérivés par affixation de -(y)iş et me. In *Noms déverbaux*, Lille, 2004.
- Christian Bassac and Mehmet Çiçek. Morphologie de la prédication verbale et non verbale en turc. In *Colloque sur la prédication*, 2004.
- Roberto Bonato and Christian Retoré. Learning rigid lambek grammars and minimalist grammars from structured sentences. In Luboš Popelínský and Miloslev Nepil, editors, *Proceedings of the third workshop on Learning Language in Logic, LLL 01*, number FI-MU-RS-2001-08 in FI MU Report series, pages 23–34, Strasbourg, September 2001. Faculty of Informatics – Masaryk University.
- Lionel Clément, Kim Gerdes, and Sylvain Kahane. An LFG-type grammar for german based on the topological model. In *LFG 2002*, Athens, 2002.
- Philippe de Groote, Glyn Morrill, and Christian Retoré, editors. *Logical Aspects of Computational Linguistics, LACL'2001*, number 2099 in LNCS/LNAI. Springer-Verlag, 2001.
- Olivier de Langhe, Pierre Guitteny, Henri Portine, and Christian Retoré. A propos des structures OSV en Langue des Signes Française. In Anne-Marie Berthonneau and Georgette Dal, editors, *Journées d'études internationales – La linguistique de la LSF: recherches actuelles*, pages 20–21. SILEX-C.N.R.S., 2003. <http://www.labri.fr/Recherche/LLA/signes/>.
- Olivier de Langhe, Pierre Guitteny, Henri Portine, and Christian Retoré. Vers un traitement informatique de la syntaxe et de la sémantique de la langue des signes française. In ATALA, editor, *Workshop on Minority Languages – TALN 2003*, Batz-sur-mer, 2003.
- Kim Gerdes et Hi-Yon Yoo. The fields on the way to prosody - alternatives to phrase structure based approaches to prosody. In *Actes du 15e Congrès international des sciences phonétiques*, Barcelona, 2003.
- Annie Foret and Yannick Le Nir. Lambek rigid grammars are not learnable from strings. In *COLING'2002, 19th International Conference on Computational Linguistics*, volume 1, pages 274–279, Taipei, Taiwan, August 2002.
- Annie Foret and Yannick Le Nir. Les grammaires de Lambek rigides ne

- sont pas apprenables à partir des chaînes. In *CAP'2002, Conférence Francophone d'Apprentissage*, pages 71–86, Orléans, France, June 2002. PUG.
- Annie Foret and Yannick Le Nir. On limit points for some variants of rigid lambek grammars. In M. van Zaanen P. Adriaans, H. Fernau, editor, *ICGI'2002, the 6th International Colloquium on Grammatical Inference*, volume 2484 of *LNCS/LNAI*, pages 106–119, Amsterdam, the Netherlands, September 2002. Springer Verlag.
 - Kim Gerdes. Tag and topology - problems and proposals for German. In *Joint conference Formal Grammar / Mathematics of Language FGMOL'01*, Helsinki, 2001.
 - Kim Gerdes. Dtag? - attempt to generate a useful TAG for German using a metagrammar. In *6th TAG+ workshop*, Venezia, 2002.
 - Kim Gerdes and Sylvain Kahane. Description of German syntax based on a topological hierarchy. In *Seventh Germanic Linguistics Annual Conference, GLAC7*, Banff, 2001.
 - Kim Gerdes and Sylvain Kahane. Pas de syntaxe sans prosodie : illustration par l'allemand. In *Actes des Journées Prosodie*, Grenoble, 2001.
 - Kim Gerdes and Sylvain Kahane. Word order in German: a formal dependency grammar using a topological hierarchy. In *Proceedings of the 39th annual meeting of the ACL, ACL 2001*, Toulouse, 2001.
 - Kim Gerdes and Sylvain Kahane. Phrasing it differently. In *Recent Works in Meaning-Text Theory in Honour of Igor Mel'cuk*, 2003.
 - Kim Gerdes and Sylvain Kahane. A polynomial parser for the topological model - illustration by a topological grammar for German, 2003.
 - Kim Gerdes and Patrice Lopez. Shared semantic representations for LTAG. In *Workshop on Representation and Treatment of Syntactic Ambiguity in Natural Language Processing*, Paris, 2000.
 - Kim Gerdes and Hi-Yon Yoo. Greek topology: A prosodic generation system for modern greek. In *Actes de la conférence Interfaces Prosodiques*, Nantes, 2003.
 - Gérard Huet. Structure of a sanskrit dictionary. Technical report, INRIA, 2000.
 - Gérard Huet. From an informal textual lexicon to a well-structured lexical database: an experiment in data reverse engineering. In *IEEE working conference on reverse engineering (WCRE'2001)*, 2001.
 - Gérard Huet. Transducers as lexicon morphisms phonemic segmentation by euphony analysis, application to a sanskrit tagger. Technical report, INRIA, 2002.

- Gérard Huet. The zen computational linguistics toolkit. Technical report, ESSLLI, 2002.
- Gérard Huet. Automata mista. In Nachum Dershowitz, editor, *Festschrift in Honor of Zohar Manna for his 64th anniversary*, volume 2772 of LNCS, Taormina, 2003. <http://pauillac.inria.fr/~huet/PUBLIC/zohar.pdf>.
- Gérard Huet. Lexicon-directed segmentation and tagging of sanskrit. In *XIIIth World Sanskrit Conference*, Helsinki, 2003.
- Gérard Huet. Linear contexts and the sharing functor: Techniques for symbolic computation. In Fairouz Kamareddine, editor, *Thirty Five Years of Automating Mathematics*, page ?? Kluwer, 2003. <http://pauillac.inria.fr/~huet/PUBLIC/DB.pdf>.
- Gérard Huet. Towards computational processing of sanskrit. In *International Conference On Natural Language Processing*, Mysore, 2003.
- Gérard Huet. Zen and the art of symbolic computing: Light and fast applicative algorithms for computational linguistics. In *Practical Aspects of Declarative Languages (PADL) symposium*, New Orleans, 2003. <http://pauillac.inria.fr/~huet/PUBLIC/padl.pdf>. Invited lecture.
- Alain Lecomte. Semantic representations in logic for minimalism. In *Workshop on Logic and Language*, pages 155–162, Seville, 2000.
- Alain Lecomte. A computational approach to minimalism. In *International Conference On Natural Language Processing*, Mysore, 2003.
- Alain Lecomte. Y a-t-il une logique de la classification ? In *L'organisation des connaissances : approches conceptuelles*, conférence invitée, Grenoble, juillet 2003. International Society for Knowledge Organization (ISKO).
- Alain Lecomte and Areski Nait Abdallah. Un modèle de raisonnement avec propositions implicites. In *Journées Nationales sur les Modèles de Raisonnement*, 2003.
- Alain Lecomte and Christian Retoré. Extending Lambek grammars: a logical account of minimalist grammars. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics, ACL 2001*, pages 354–361, Toulouse, July 2001. ACL.
- Alain Lecomte and Christian Retoré. Bi-grammars for the syntax-semantics interface. In Gehrard Jäger, Paola Monachesi, Gerald Penn, and Shuly Winter, editors, *Proceedings Formal Grammar 2002*, pages 77–90. FoLLI, 2002.
- Alain Lecomte and Christian Retoré. Sémantique dans les grammaires minimalistes. In *Journées Sémantique et Modélisation*. C.N.R.S., 2003.

- Michael Moortgat and Richard Moot. CGN to Grail: Extracting a type-logical lexicon from the CGN annotation. In Walter Daelemans, editor, *Proceedings of CLIN 2000*, 2001.
- Michael Moortgat and Richard Moot. Using the Spoken Dutch Corpus for type-logical grammar induction. In *Proceedings of the International Conference on Language Resources and Evaluation*, 2002.
- Michael Moortgat, Richard Moot, and Richard T. Oehrle. T_EX in teaching. In Simon Pepping, editor, *Proceedings EuroT_EX 2001*, pages 130–140, 2001.
- Richard Moot. A short introduction to Grail. In Carlos Areces and Maarten de Rijke, editors, *Proceedings of Methods for Modalities 2*, 2001.
- Henri Portine. Connecteurs et marqueurs : Quels termes métalinguistiques ? In B. Colombat and M. Savelli, editors, *Métalangage et terminologie linguistique (Grenoble, 1998)*, pages 239–249. Peters, 2001.
- Henri Portine. Repérages énonciatifs et pronoms personnels. In Jean-Paul Bernier, editor, *Apprentissage, développement et significations : Hommage à Michel Brossard*, pages 111–128. Presses Universitaires de Bordeaux, 2001.
- Henri Portine. La subordination relative en français et en anglais. In *2003 Association for French Language Study annual conference*, Tours, 2003.
- Christian Retoré. A description of the non-sequential execution of Petri nets in partially commutative linear logic. In Jan van Eijck, Vincent van Oostrom, and Albert Visser, editors, *Logic Colloquium '99 (Utrecht)*, Lecture Notes in Logic. Association for Symbolic Logic & A. K. Peters, Ltd, 2002. Complete version: RR-INRIA 4288 <http://www.inria.fr/>.
- Christian Retoré. A description of the non-sequential execution of petri nets in partially commutative linear logic. In Jan van Eijck, Vincent van Oostrom, and Albert Visser, editors, *Logic Colloquium 99*, Lecture Notes in Logic. ASL and A. K. Peters, 2003. Complete version INRIA RR-4288 (2001) <http://www.inria.fr>.
- Christian Retoré. Semantic aspects of minimalist grammars. In F. Spoto, G. Scollo, and A. Nijholt, editors, *Algebraic Methods in Language Processing – AMiLP 2003*, volume 21 of *TWLT*. University of Twente, 2003. Invited lecture