



# Présentation à Melodi (IRIT)

Christian Retoré

prof. d'info université de Bordeaux

en délégation C.N.R.S. à l'IRIT en 2012–2013

mardi 18 septembre 2012

# 1. Quand j'étais jeune : logique (preuves)

Thèse de maths (Paris 7, 1993) en logique, théorie de la démonstration

**logique linéaire en particulier non commutative**

.

Techniquement :

syntaxe logique, structure des preuves, graphes de preuves  
sémantique catégorique (dénotationnelle)  
logique intuitionniste, *proofs-as-programs*

Applications en vue :

programmation fonctionnelle, partage de calculs, calcul de processus

16 mois à Imperial College, Londres (service national ;-)

## 2. INRIA (PostDoc, CR) Sophia, Nancy, Rennes : grammaires catégorielles

Révélation en 1994 (merci à Alain Lecomte, Grenoble),

logique linéaire intuitionniste non commutative

= calcul de Lambek, "moteur" des grammaires catégorielles

analyser une preuve c'est construire une preuve dans cette logique de

$$t_1, \dots, t_n \vdash S$$

où  $t_i$  est une formule logique associée au  $i$ -ième mot par le lexique

**parsing-as-deduction**

Lexicalisées (règles universelles, seul le lexique varie)

Apprenables à la limite (Gold)

Interface syntaxe sémantique

Nancy (94-97) : Equipe Projet INRIA Calligramme (co-fondateur)

Rennes (1998-2002) : Equipe Projet INRIA Aïda & Paragraphe  
acquisition de grammaires catégorielles

### 3. Prof. Bordeaux (2002→) : syntaxe et sémantique

Principalement dans le cadre des grammaires catégorielles  
mais pas seulement (grammaires minimalistes)  
sémantique de Montague mais aussi Lambda DRT

développement de la linguistique informatique à Bordeaux  
montage de l'équipe projet INRIA Signes (2002–2011)  
(analyse syntaxique et sémantique, Néerlandais, Français,  
Langues des Signes avec des gens de l'ERSS)  
rédacteur en chef de TAL (2004–2009)  
fort investissement dans le GDR sémantique et modélisation  
(2006 Bordeaux, 2008 Toulouse)  
organisation d'ESLLI 2009 à Bordeaux  
(des volontaires pour ESLLI Toulouse 2014 ?)

## 4. Situation à Bordeaux

Beaucoup moins de soutien de la direction depuis 2007  
par exemple plus de recrutement depuis 2006,  
équipe INRIA arrêtée fin 2011.

Quelques départs (mutations, promotions). Qui reste-t-il ?

Lionel Clément (syntaxe des verbes en LFG)

Sylvain Salvati (grammaire d'arbres, lambda calcul)

Richard Moot

(logique, grammaires catégorielles, syntaxe et sémantique)

soit **1 interlocuteur...**

Deux thèses en 3e et 4e année :

acquisition statistique et symbolique de grammaires catégorielles (Noémie-Fleur Sandillon-Rezer),

modélisation du temps dans les récits de voyage (Anaïs Le-feuvre, à Blois en 2012-2013),

Pas fâché de venir un an dans Melodi !



## 5. Liens officiels avec Toulouse

ITIPY (région & INRIA)

Bordeaux Pau Toulouse (2009–2012)

Reconstruction d'itinéraires à partir de récits de voyage du  
XIXe dans les Pyrénées

Une thèse en cours, Anaïs Lefeuvre

POLYMNIE ANR CONTINT

Nancy Paris 7 Bordeaux Toulouse (2012–2015)

discours dans les grammaires catégorielles abstraites

PRELUDE/LOCI ANR Blanc (parfois avec Nicholas)

(2008–2014) Bordeaux Nancy Paris 8

Utilisation de la théorie de la démonstration et de jeux ("ludique") pour modéliser discours et dialogue.



## 6. Thématique : logique et linguistique

### Thèmes linguistiques

- Syntaxe et sémantique
- Syntaxe, convergence entre :
  - grammaires de Lambek,
  - grammaires minimalistes
- Sémantique compositionnelle
- Interface syntaxe sémantique
- Interface sémantique discours
- Pragmatique lexicale

### Méthodes

- Logique, preuves formelles, théorie des types, catégories
- Arbres et graphes de preuves = analyses syntaxiques
- Syntaxe → sémantique (transformations logiques)
- Logique d'ordre supérieur en théorie des types

---

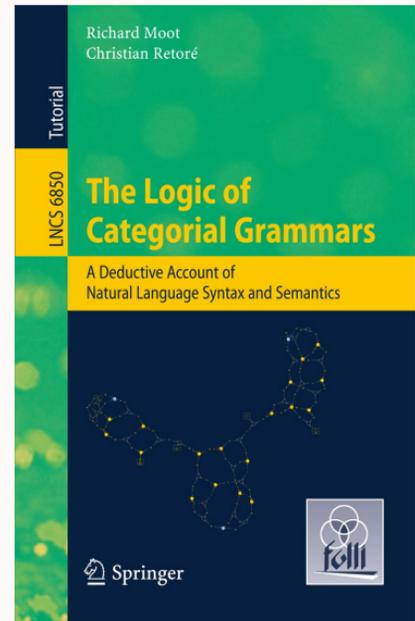
### Données, expériences (à développer !)

- Initialement : exemples livresques
- Corpus Itipy (Pau Bordeaux Toulouse) récits de voyage
- Corpus CID dialogues (Aix) discussion sur "insolite"
- Futur : expériences (phénomènes rares, énoncés incorrects)

## 7. Syntaxe et logique non commutative : une page est tournée

R. Moot & Ch. Retoré, The logic of categorial grammars, Springer, Juillet 2012.

Même si nous continuons à étudier quelques points de syntaxe catégorielle, nous sommes moins actifs sur ces questions et davantage tournés vers la sémantique.



## 8. Syntaxe $\Leftrightarrow$ sémantique

Analyse catégorielle (= preuve =  $\lambda$ -terme) +  $\lambda$ -termes lexicaux  
→ formule logique exprimant « qui fait quoi »

<b>word</b>	<b><i>type sémantique</i></b> $u^*$ <b><i>semantics</i></b> : $\lambda$ -terme de type $u^*$ $x^v$ la variable ou constante $x$ est de type $v$
<b>certain</b>	$(e \rightarrow t) \rightarrow ((e \rightarrow t) \rightarrow t)$ $\lambda P^{e \rightarrow t} \lambda Q^{e \rightarrow t} ((e \rightarrow t) \rightarrow t (\lambda x^e (t \rightarrow (t \rightarrow t) (P x)(Q x))))$
<b>énoncés</b>	$e \rightarrow t$ $\lambda x^e (\text{statement}^{e \rightarrow t} x)$
<b>parlent_de</b>	$e \rightarrow (e \rightarrow t)$ $\lambda y^e \lambda x^e ((\text{speak\_about}^{e \rightarrow (e \rightarrow t)} x)y)$
<b>eux-mêmes</b>	$(e \rightarrow (e \rightarrow t)) \rightarrow (e \rightarrow t)$ $\lambda P^{e \rightarrow (e \rightarrow t)} \lambda x^e ((P x)x)$



Pour la phrase :

Some statements speak about themselves.

La syntaxe fournit un  $\lambda$ -terme de ce genre :

*((some statements) (themselves speak\_about))*

de type  $S^* = t$

On remplace les mots par les lambda termes correspondant, et on réduit....

$$\begin{aligned} & \left( (\lambda P^{e \rightarrow t} \lambda Q^{e \rightarrow t} ((e \rightarrow t) \rightarrow t (\lambda x^e ((P x)(Q x)))) \right. \\ & \quad \left. (\lambda x^e (\text{statement}^{e \rightarrow t} x)) \right) \\ & \left( (\lambda P^{e \rightarrow (e \rightarrow t)} \lambda x^e ((P x)x) \right. \\ & \quad \left. (\lambda y^e \lambda x^e ((\text{speak\_about}^{e \rightarrow (e \rightarrow t)} x)y)) \right) \end{aligned}$$

$$\begin{aligned} & \downarrow \beta \\ & (\lambda Q^{e \rightarrow t} ((e \rightarrow t) \rightarrow t (\lambda x^e (t \rightarrow (t \rightarrow t) (\text{statement}^{e \rightarrow t} x)(Q x)))) \\ & \quad (\lambda x^e ((\text{speak\_about}^{e \rightarrow (e \rightarrow t)} x)x)) \end{aligned}$$

$$\begin{aligned} & \downarrow \beta \\ & ((e \rightarrow t) \rightarrow t (\lambda x^e ((\text{statement}^{e \rightarrow t} x)((\text{speak\_about}^{e \rightarrow (e \rightarrow t)} x)x)))) \end{aligned}$$

le résultat est y plus lisible ainsi :

$$\exists x : e (\text{statement}(x) \wedge \text{speak\_about}(x, x))$$

## 9. Programme à l'IRIT : sémantique en théorie des types

Dans le cadre compositionnel ci-dessus comment rendre compte :  
de la sémantique lexicale :

filtrer les compositions impossibles

*\* la chaise crie*

mais ajouter de la flexibilité quand nécessaire

*Liverpool a battu Milan*

de certaines questions classiques de sémantique

quantification généralisée

*un tiers, la plupart,*

pluriels

*Les étudiants ont moins de 30 ans.*

*Ils se sont rencontrés.*

Implémentation dans la plateforme d'analyse syntaxique et sémantique  
Grail de Richard Moot (grammaires catégorielles, lambda-DRT).

## 10. Sémantique et pragmatique lexicale

avec Bruno Mery et Christian Bassac

Modélisation dans le lexique

( $\lambda$ -calcul polymorphe du 2nd ordre, avec quantification sur les types  $\Lambda$ ) :

- Structure argumentale :  $\lambda$ -terme usuel à utiliser
- Glissement de sens :  $\lambda$ -termes optionnels, transformations.

Langue  $\neq$  monde : *Ma voiture est crevée/\*bouchée.*



## 11. Sémantique et pragmatique lexicale – suite

Contraintes

- Rigide : la même transformation pour chaque occurrence.
- Flexible : des transformations différentes pour chaque occurrence.

Flexible :

*Paris est une capitale cosmopolite. (gens, administration)*

*Ce livre est intéressant mais encombrant.*

Rigide :

*Montpellier est mal classé mais doit pourtant affronter Arsenal. (équipe)*

*\* Washington est au bord du Potomak et a attaqué l'Iraq par deux fois.*



## 12. Sémantique et pragmatique lexicale – fin

Dans la version implémentée (en  $\lambda$ -DRT),  
les contraintes sont spécifiées dans le lexique  
mais gérées par un mécanisme externe au  $\lambda$ -calcul.

Une amélioration récente : en utilisant un  $\lambda$ -calcul linéairement typé,  
on peut exprimer les contraintes seulement avec les types, sans mé-  
canisme externe.



## 13. Phénomènes traités

Le cas classique :

*une petite pierre (pierre : phys)*

*\*la chaise crie (chaise : phys crie : animal → t)*

Les facettes :

*Barcelone a battu Séville.*

*Barcelone refuse un statut transfrontalier à la Cerdagne.*

*\* Barcelone a battu Séville et refuse un statut transfrontalier  
à la Cerdagne.*



## 14. À quoi sert le second ordre ?

Il y a beaucoup de types avec beaucoup de liens entre eux (une ontologie langagière ?).

Le second ordre permet de définir des opérations uniformes sur des familles de types :

Étant donnés

deux prédicats  $P^{\alpha \rightarrow t}$  (par ex. lourd) et  $Q^{\beta \rightarrow t}$  (par. intéressant)  
sur des sortes respectives  $\alpha$  (par ex. objets) et  $\beta$  (par ex. info)

deux transformations de  $\xi$  (livre) vers  $\alpha$  et vers  $\beta$

on peut coordonner des objets de type  $\xi$  :

$$\Lambda \xi \lambda x^{\xi} \lambda f^{\xi \rightarrow \alpha} \lambda g^{\xi \rightarrow \beta} . (and (P (f x))(Q (g x)))$$

pour analyser ce genre de phrase :

*Ce livre est lourd mais intéressant.*

## 15. Extension au voyageur fictif avec Richard Moot et Laurent Prévot

Corpus Itipy :

- (1) (...) cette route monte jusqu'à Lux où l'on arrive par une jolie avenue de peupliers.
- (2) (...) cette route qui monte sans cesse pendant deux heures le complément de temps nécessite un voyageur fictif empruntant la route.

Montée de type  $\lambda$ -terme optionnel transformation sur le type élevé.

Cas particulier du précédent : conflit de type entre *animé* et *route*, et *route* possède un  $\lambda$ -terme optionnel qui introduit le *voyageur* et l'*événement*  
— ce qui permet aussi de traiter aussi *la route agréable*.



## 16. Lien avec Itipy

L'extension précédente à été conçue pour le projet Itipy (récits du 19<sup>e</sup> de voyages dans les Pyrénées), reconstruction d'itinéraire :

1. (autres outils) Sélection des passages concernés.
2. Analyse syntaxique (grammaire catégorielle acquise), sémantique en lambda-DRT
3. Sémantique / pragmatique lexicale.
4. (à faire) S-DRT pour exprimer les relations discursives entre DRS



## 17. Quantification - projet Loci, avec Abrusci

*Chacun / tous les* : collection, ensemble } Th. de complétude :  
*Tout, un (générique)* : preuves } les 2 coïncident...

La langue

– Quantifie sur des ordres supérieurs :

(3) *Tout le monde ment, mais lui croit tout.*

– Regorge de quantifications généralisées (souvent implicites)

(4) *Les français grelottent pour faire des économies*

– **Plus de théorème de complétude !**

Modèles (mesures et non cardinaux) :

(5) — *La plupart des nombres sont premiers.* — *Non.*

(6) — *La plupart des nombres ne sont pas premiers.* — *Oui.*

Les modèles requièrent de considérer tous les mondes possibles, et sont donc peu réalistes.

## 18. Quantification

Qu'est-ce qu'un quantificateur ? Modèles et RÈGLES.

Règles ? Réfutations ?

Jeux entre modèles (ensembles) et preuves (générique) :

- Générique : — *Un chien, ça peut toujours mordre.*
- (élément du modèle) — *Mais non, pas Rex.*
- (autre générique) — *Mais non, pas un basset.*

Mais quelles données pour valider notre modélisation ?

dialogue pour observer l'incompréhension,

mais données maigres et la quantification est rare

dialogues ciblés

mais expériences spécifiques (EEG, temps de réaction).

## 19. Les pluriels

avec Richard Moot

Introduire des opérateurs permettant de constituer des ensembles, de forcer la lecture collective ou distributive.

*Les deux déménageurs ont porté le piano.*

*Les deux déménageurs avaient une quarantaine d'années.*



## 20. Quelques références

Structure du lexique pour la sémantique compositionnelle et la pragmatique lexicale : [article dans Journal of Logic, Language and Information, 2010](#) Un article plus récent est en préparation pour : [Language and linguistic compass](#) (un peu de pub pour cette nouvelle revue)

Voyageur virtuel : [en français à la Montague \(TALN 2011\)](#) ou [en anglais avec la  \$\lambda\$ -DRT \(CID 2011\)](#)

Quantification : ["most" dans ce cadre \(article dans RLV\)](#) ou bien [notre communication sur les interprétations en termes de preuves au 14th CLMPS \(article à venir\)](#)

Pluriels : [notre communication Coconat workshop](#)



## 21. Quelques questions que j'aimerais aborder avec MELODI

(discussions avec R. Moot, N. Asher, Z. Luo, L. Prévot, S. Soloviev)

Quels sont les types de base ? les noms communs ? les sortes d'une ontologie ? toute propriété ? Peut-on les acquérir ?

Quel est le lien entre la formulation logique  $livre(x)$  (qui peut être nié) et le typage  $x : livre$  (qui ne peut pas être nié, comme une présupposition).

Quel est la bonne notion de sous-typage qui fonctionne avec le typage polymorphe ?

Une question toute récente : les modèles dénotationnels de la théorie des types (domaine de Scott, espace cohérent) peuvent-ils être utilisés comme structure pour rendre compte d'une ontologie ? quelques travaux en ce sens (article récent de Abrusci, Fouqueré, Romano).

Tests en corpus, expériences pour valider notre modélisation.  
Implantation dans Grail des avancées.

## 22. En pratique

Exposé plus détaillé le vendredi 5 octobre à 14h à l'IRIT.

Workshop sémantique lexicale dans le cadre de l'ANR blanc Loci (invités : Robin Cooper, Zhaohui Luo) octobre ou novembre 2012

Mise en place de l'ANR ContInt Polymnie  
réunion d'ouverture à Nancy en octobre 2012  
réunion à Toulouse au printemps 2012 ?

**Sémagramme, Nancy ; IRIT, Toulouse ; Labri, Bordeaux ; Alpage, Paris**

