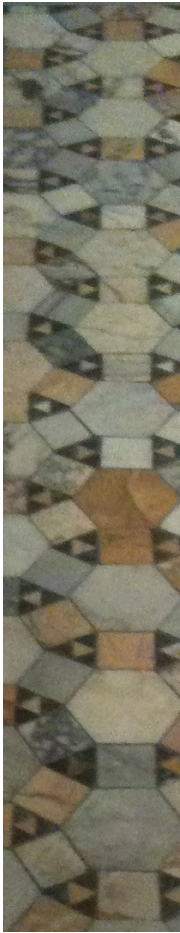




# COMPRÉHENSION AUTOMATIQUE DU LANGAGE NATUREL

Équipe **TEXTE** du LIRMM  
Christian Retoré

Visite ÉNS Cachan — 6 janvier 2016



# **A Traitement automatique des langues**



## A.1. Famille de problèmes à traiter

Deux tâches classiques — traduction (depuis 1949 !) = A puis B

A analyse automatique (de texte, de parole, annoté ou non)

→ forme du résultat ?

B génération automatique (de texte, de parole)

→ à partir de quoi ?

Autre domaine utile pour A et B et pour la recherche d'information :

fouille, acquisition des données



## A.2. Applications

**Traduction automatique**, aide à la traduction

The flesh is weak but the spirit is willing.

(russian)

The meat is rotten but the vodka is strong.

**Aide à la traduction** : domaine spécifique, bonne représentation des connaissances, reformation, interaction avec l'utilisateur.

**Correcteurs orthographiques** (Word(Synapse) : vert : grammaires rouge : lexique)

(1) Quels livreS crois-tu qu'il sait que je pense que tu as luS ?

**Dialogue homme machine**

(2) Quels sont les films des années cinquante qui passent actuellement à Bordeaux ?

(3) Les enfants prendront une pizza.



## Résumé automatique

représentation formelle et élagage

utiliser la structure de la rédaction

**Réponse à des questions** (question answering cf. text entailment)

Geach était-il l'élève de Wittgenstein ? (cf. ci-après)

Domaine relié : **recherche d'information**

Différence : Big Data, beaucoup de données, analyse superficielle, méthodes statistiques, machine learning

CD ? "ici, on ne vend pas de CD seulement des vinyles"  
"production laitière / production de lait" mais "production minière / production de mine(s)"



**B Niveaux d'analyse de la langue  
& méthodes informatiques**



## **B.1. Les sons : phonétique/phonologie**

**B.1.a Phonétique** Acoustique / système phonatoire/auditif

**Traitement du signal / médecine**

**B.1.b Phonologie**

Les sons abstraits : système discret (dans un continu)

(4) Bali / Paris indistincts pour un japonais

**Automates et transducteurs**



## B.2. Les mots : morphologie

### B.2.a Morphologie dérivationnelle

changement de catégorie possible

barrer/barrage maquiller/maquillage garer/garage

maison /maisonnette camion /camionnette carpe/carpette

**syntaxe : automates et transducteurs / sémantique : difficile**

### B.2.b Morphologie flexionnelle (même catégorie)

cheval → chevaux

aller → allons aller → va aller → irons

**automates et transducteurs**

### B.2.c Étiquetage grammatical

nom, verbe, déterminant etc.

**règles ou probas (modèles de Markov cachés)**





### B.3. Analyse de la phrase (arbre) : syntaxe

- (5) Réparer fais les la le.
- (6) \*Je fais la réparer
- (7) Je la fais réparer
- (8) \* [[Pierre [mange une]] pomme]
- (9) Pierre [mange [une pomme]]

**Grammaires formelles** (grammaires algébriques, d'arbres, TAGs,..)

**Déduction formelle** (cf. ci-après)

**Ensembles de contraintes** (e.g. "groupe nominal sujet avant verbe")



## B.4. Le sens d'un mot en contexte : sémantique lexicale

(10) J'ai fini mon livre.

de quelle action s'agit-il ?

LIVRE :  $\left[ \begin{array}{l} \text{Rôle téléique : être lu, informer, cultiver,} \\ \text{Rôle constitutif : pages, couvertures} \\ \text{Rôle agentif : imprimeur, auteur,} \end{array} \right.$

Ambiguïtés, restrictions de sélection, relation privilégiée entre sens

**acquisition : machine learning, (grands) graphes, jeux**  
**utilisation : logique**

**(juste après : atelier JeuxDeMots par Mathieu Lafourcade)**



## B.5. Le sens d'une phrase : sémantique compositionnelle, formelle

- (11) a. Les enfants prendront une pizza.  
b.  $\exists x. \text{pizza}(x) \wedge \forall w. (\text{enfant}(w) \Rightarrow F(\text{prendre}(w, x)))$   
c.  $\forall u. \text{enfants}(u) \Rightarrow \exists x. \text{pizza}(x) \wedge F(\text{prendre}(u, x))$

Phrase  $\rightarrow$  formule logique (quelques détails ci-après)

Ensuite : ensemble des mondes possibles dans lequel la phrase est vraie (modèles de Kripke)

**lambda calcul typé, logique**



## **B.6. L'interprétation en en contexte : pragmatique (discours, dialogue)**

Prise en compte du contexte linguistique et extra linguistique

- (12) a. Allons plutôt dans ce restaurant.  
b. "nous" ? "ce" ?
- (13) a. Il est tombé. Quelqu'un l'a poussé.  
b. "l"="il" + causalité

**le contexte est difficile à modéliser**  
**tentatives : logique, structures attribut valeurs récursives,**



## B.7. Deux notions de "sémantique"

Geach était-il l'élève de Wittgenstein ?

Wikipédia : En 1941, IL épousa la philosophe Elizabeth Anscombe, grâce à LAQUELLE IL entra en contact avec Ludwig Wittgenstein. Bien qu'IL N'ait JAMAIS suivi l'enseignement académique de CE DERNIER, cependant IL EN éprouva fortement l'influence.

1. De quoi ça parle ?

**réseaux lexicaux, vecteurs de mots, analyse de texte, par des méthodes statistiques**

Atelier JeuxDeMots : relations entre les sens d'un mot, construction de ce graphe des relations sémantiques entre mots par des jeux collaboratifs en ligne.

2. Qui fait quoi, qu'est-ce qui est affirmé, supposé, réfuté, **analyse logique**. Cf. ci-après.

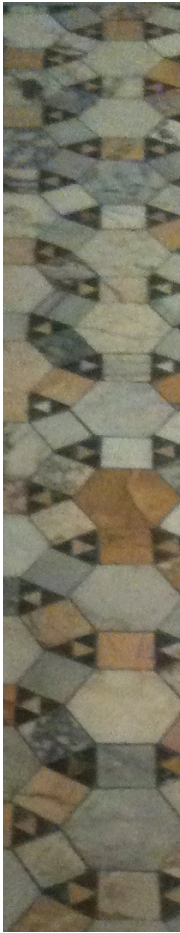
3. combinaison des deux ? (en cours et très intéressant : stages)



**C Analyse logique  
du sens d'une phrase**

## C.1. L'ingrédient clé : un lexique syntaxico/sémantique

mot	<i>catégorie syntaxique</i> $u$ <i>type sémantique</i> $u^*$ <i>sémantique</i> : $\lambda$ -term of type $u^*$ $x^v$ signifie $x$ (variable, constante) de type $v$
les	$(S/(np \setminus S))/n$ (subject) $((S/np) \setminus S)/n$ (object) $(e \rightarrow t) \rightarrow ((e \rightarrow t) \rightarrow t)$ $\lambda P^{e \rightarrow t} \lambda Q^{e \rightarrow t} (\forall^{(e \rightarrow t) \rightarrow t} (\lambda x^e (\Rightarrow^{t \rightarrow (t \rightarrow t)} (P x)(Q x))))$
une	$((S/np) \setminus S)/n$ (object) $(S/(np \setminus S))/n$ (subject) $(e \rightarrow t) \rightarrow ((e \rightarrow t) \rightarrow t)$ $\lambda P^{e \rightarrow t} \lambda Q^{e \rightarrow t} (\exists^{(e \rightarrow t) \rightarrow t} (\lambda x^e (\wedge^{t \rightarrow (t \rightarrow t)} (P x)(Q x))))$
enfant(s)	$n$ $e \rightarrow t$ $\lambda x^e (\text{enfant}^{e \rightarrow t} x)$
pizza	$n$ $e \rightarrow t$ $\lambda x^e (\text{pizza}^{e \rightarrow t} x)$
prendront	$(np \setminus S)/np$ $e \rightarrow (e \rightarrow t)$ $\lambda y^e \lambda x^e ((\text{prendront}^{e \rightarrow (e \rightarrow t)} x)y)$



## C.2. Syntaxe catégorielle : principe

catégories S : phrase, np : noun phrase (groupe nominal), n : nom commun

si  $w : B/A$  alors  $w$  suivi de  $u : A$  donne  $wu : B$

réciroquement, si  $w$  suivi de n'importe quoi (une variable) de type  $A$  est de type  $B$  alors  $w$  est de type  $B/A$

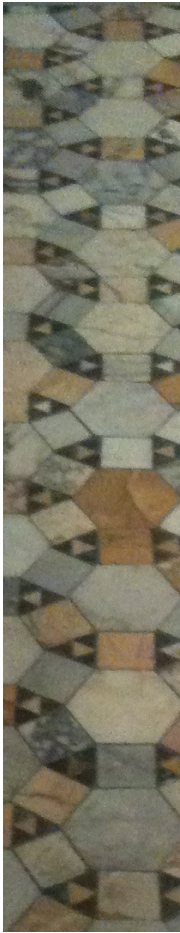
$A$  hypothèse la plus à gauche

$$\frac{\begin{array}{c} \dots [A] \dots \\ \vdots \\ B \end{array}}{A \backslash B} \backslash_i \text{ — l'hyp. } A \text{ est annulée}$$

$$\frac{\begin{array}{c} \Delta \quad \Gamma \\ \vdots \quad \vdots \\ A \quad A \backslash B \end{array}}{B} \backslash_e$$

$A \backslash B$  même chose, mais  $A$  est à gauche.





### C.3. Analyse syntaxique = déduction (parsing as deduction)

**Une phrase est correcte si on peut assigner à chaque mot une catégorie de sorte que la suite des catégories entraîne  $S$  (dans une logique particulière).**

$m_1 \dots m_n$  est une phrase ssi :

$$\forall i \exists c_i \in Lex(m_i) \quad c_1 \dots c_n \vdash S$$

## C.4. Analyse syntaxique $\exists \forall$

Il y a deux analyse syntaxique possibles. Une :

$\exists \forall$

$$\frac{\frac{(S/(np \setminus S))/n \quad n}{(S/(np \setminus S))} /_e \quad \frac{(np \setminus S)/np \quad [np]^1}{(np \setminus S)} /_e}{\frac{S}{S/np} /_i(1)} /_e \quad \frac{((S/np) \setminus S)/n \quad n}{(S/np) \setminus S} \setminus_e$$

$S$

## C.5. Syntaxe $\rightarrow$ $\lambda$ -terme sémantique de la phrase

$\exists \forall$

$$\begin{array}{c}
 \begin{array}{cccc}
 \textit{les} & \textit{enfants} & \textit{prendront} & \textit{o} \\
 (e \rightarrow t) \rightarrow (e \rightarrow t) \rightarrow t & (e \rightarrow t) & e \rightarrow e \rightarrow t & [e]^1
 \end{array} \\
 \hline
 (e \rightarrow t) \rightarrow t \quad \rightarrow_e \quad e \rightarrow t \quad \rightarrow_e
 \end{array}$$
  

$$\begin{array}{c}
 \begin{array}{cc}
 \frac{t}{e \rightarrow t} \rightarrow_i(1) & \frac{\textit{une} \quad \textit{pizza}}{(e \rightarrow t) \rightarrow (e \rightarrow t) \rightarrow t \quad (e \rightarrow t)} \rightarrow_e
 \end{array} \\
 \hline
 t \quad \rightarrow_e
 \end{array}$$

Le  $\lambda$ -terme correspondant est :

$$\exists \forall = (\textit{une pizza})(\lambda o^e(\textit{les enfants})(\textit{prendront } o))$$

Il faut encore :

1. insérer les lambda terme lexicaux et
2. réduire/calculer

## C.6. Calculs, par étapes 1/2

(une pizza)

$$\begin{aligned} &= (\lambda P^{e \rightarrow t} \lambda Q^{e \rightarrow t} (\exists^{(e \rightarrow t) \rightarrow t} (\lambda x^e (\wedge^{t \rightarrow (t \rightarrow t)} (P x)(Q x)))))(\lambda z^e (\text{pizza}^{e \rightarrow t} z)) \\ &= (\lambda Q^{e \rightarrow t} (\exists^{(e \rightarrow t) \rightarrow t} (\lambda x^e (\wedge^{t \rightarrow (t \rightarrow t)} ((\lambda z^e (\text{pizza}^{e \rightarrow t} z)) x)(Q x)))))) \\ &= (\lambda Q^{e \rightarrow t} (\exists^{(e \rightarrow t) \rightarrow t} (\lambda x^e (\wedge^{t \rightarrow (t \rightarrow t)} ((\text{pizza}^{e \rightarrow t} x))(Q x)))))) \end{aligned}$$

(les enfants)

$$\begin{aligned} &= (\lambda P^{e \rightarrow t} \lambda Q^{e \rightarrow t} (\forall^{(e \rightarrow t) \rightarrow t} (\lambda x^e (\Rightarrow^{t \rightarrow (t \rightarrow t)} (P x)(Q x)))))(\lambda u^e (\text{enfant}^{e \rightarrow t} u)) \\ &= (\lambda Q^{e \rightarrow t} (\forall^{(e \rightarrow t) \rightarrow t} (\lambda x^e (\Rightarrow^{t \rightarrow (t \rightarrow t)} ((\lambda u^e (\text{enfant}^{e \rightarrow t} u)) x)(Q x)))))) \\ &= (\lambda Q^{e \rightarrow t} (\forall^{(e \rightarrow t) \rightarrow t} (\lambda x^e (\Rightarrow^{t \rightarrow (t \rightarrow t)} (\text{enfant}^{e \rightarrow t} x)(Q x)))))) \end{aligned}$$

(les enfants)(prendront o) =

$$\begin{aligned} &(\lambda Q^{e \rightarrow t} (\forall^{(e \rightarrow t) \rightarrow t} (\lambda w^e (\Rightarrow^{t \rightarrow (t \rightarrow t)} (\text{enfant}^{e \rightarrow t} w)(Q w)))))(\lambda y^e \lambda x^e ((\text{prendront}^{e \rightarrow (e \rightarrow t)} y x) o)) \\ &= (\lambda Q^{e \rightarrow t} (\forall^{(e \rightarrow t) \rightarrow t} (\lambda w^e (\Rightarrow^{t \rightarrow (t \rightarrow t)} (\text{enfant}^{e \rightarrow t} w)(Q w)))))(\lambda x^e ((\text{prendront}^{e \rightarrow (e \rightarrow t)} x) o)) \\ &= \forall^{(e \rightarrow t) \rightarrow t} (\lambda w^e (\Rightarrow^{t \rightarrow (t \rightarrow t)} (\text{enfant}^{e \rightarrow t} w)((\lambda x^e ((\text{prendront}^{e \rightarrow (e \rightarrow t)} x) o)) w)) \\ &= \forall^{(e \rightarrow t) \rightarrow t} (\lambda w^e (\Rightarrow^{t \rightarrow (t \rightarrow t)} (\text{enfant}^{e \rightarrow t} w)((\text{prendront}^{e \rightarrow (e \rightarrow t)} w) o)))) \end{aligned}$$

## C.7. Calculs, par étapes 2/2

$$\begin{aligned}
 & (\text{une pizza})(\lambda o \text{ (les enfants)})(\text{prendront } o)) \\
 &= (\lambda Q^{e \rightarrow t} (\exists^{(e \rightarrow t) \rightarrow t} (\lambda x^e (\wedge^{t \rightarrow (t \rightarrow t)} ((\text{pizza}^{e \rightarrow t} x)))) (Q x)))) \\
 & \quad (\lambda o \forall^{(e \rightarrow t) \rightarrow t} (\lambda w^e (\Rightarrow^{t \rightarrow (t \rightarrow t)} (\text{enfant}^{e \rightarrow t} w) (((\text{prendront}^{e \rightarrow (e \rightarrow t)} w) o)))))) \\
 &= (\exists^{(e \rightarrow t) \rightarrow t} (\lambda x^e (\wedge^{t \rightarrow (t \rightarrow t)} ((\text{pizza}^{e \rightarrow t} x)))) \\
 & \quad ((\lambda o \forall^{(e \rightarrow t) \rightarrow t} (\lambda w^e (\Rightarrow^{t \rightarrow (t \rightarrow t)} (\text{enfant}^{e \rightarrow t} w) (((\text{prendront}^{e \rightarrow (e \rightarrow t)} w) o)))))) \\
 &= (\exists^{(e \rightarrow t) \rightarrow t} (\lambda x^e (\wedge^{t \rightarrow (t \rightarrow t)} ((\text{pizza}^{e \rightarrow t} x)))) \\
 & \quad (\forall^{(e \rightarrow t) \rightarrow t} (\lambda w^e (\Rightarrow^{t \rightarrow (t \rightarrow t)} (\text{enfant}^{e \rightarrow t} w) (((\text{prendront}^{e \rightarrow (e \rightarrow t)} w) x))))))
 \end{aligned}$$

ce qui s'écrit communément :

$$\exists x. \text{pizza}(x) \wedge \forall w. (\text{enfant}(w) \Rightarrow \text{prendront}(w, x))$$

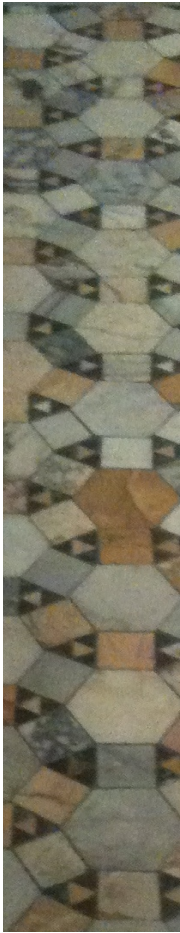
## C.8. Avec l'autre analyse syntaxique...

∃

$$\frac{\frac{\frac{(np \setminus S) / np \quad [np]^2}{(np \setminus S)} /_e}{[np]^1} \setminus_e \quad \frac{\frac{S}{S / np} /_{i(2)} \quad \frac{((S / np) \setminus S) / n \quad n}{(S / np) \setminus S} \setminus_e}{S} \setminus_e}{\frac{(S / (np \setminus S)) / n \quad n}{(S / (np \setminus S))} /_e} \setminus_e \quad \frac{S}{np \setminus S} \setminus_{i(1)} /_e} S /_e$$

On trouve l'autre interprétation :

$$\forall u. \text{enfants}(u) \Rightarrow \exists x. \text{pizza}(x) \wedge \text{prendront}(u, x)$$





## C.9. Sortes et réseaux lexicaux

- (14) ??? La chaise a gagné.
- (15) a. J'ai porté au grenier tous les livres de la bibliothèque.  
b. J'ai lu tous les livres de la bibliothèque.  
c. J'ai porté au grenier tous les livres de la bibliothèque, je les avais tous lus.
- (16) a. Ma fille de six ans a fini son premier livre.  
b. Mathias Enard a fini son dernier livre.
- (17) a. Liverpool a gagné dimanche.  
b. Liverpool a voté dimanche.  
c. Liverpool est un grand port.  
d. Liverpool, qui est un grand port, a voté dimanche.  
e. ??? Liverpool, qui est un grand port, a gagné dimanche.



**D Du pain sur la planche :  
stage, thèse, voire toute une vie**





## Défi actuel : lien entre **réseau lexical** et **sémantique logique**

j'ai fini mon livre  $\rightarrow \exists x. \text{fini}(\text{moi}, e(x)) \& \text{livre}(x) \& R(x, l)$

comment inférer de JeuxDeMots

$(e=\text{lire} \& R=\text{appartient}) \vee (e=\text{ecrire} \& R=\text{auteur})$  (+préférences ?)

### **Formalisation de phénomènes linguistiques spécifiques :**

- types de base (sortes) pour les restrictions de sélection
- déverbaux : allumage maquillage pâturage
- voyageur virtuel : "le sentier descend pendant 20 minutes"

### **Questions logiques et linguistiques :**

- sémantique des pluriels
- lexiques et ontologies

### **Questions logiques :**

- sous typage
- lien théorie des types / calcul des prédicats
- quantification (fonction de choix, epsilon de Hilbert)