



## **Compréhension automatique du langage naturel**

quelques applications récentes  
des liens entre logique et langage naturel

**Christian Retoré**

Équipe Texte, LIRMM, Université de Montpellier

ÉNS Cachan

7 septembre 2016



## I.1. Avant-propos

Merci à Jean Goubault de m'avoir invité.

Présentation partielle et partielle d'un domaine :

la linguistique informatique ou traitement automatique des langues

avec un point de vue **logique** (théorie des types, lambda calcul, système déductifs — j'aurais pu inclure la théorie des langages) qui ne sont pas les seules méthodes (approches quantitatives, statistiques)

*Comment en suis-je arrivé là ? maîtrise de maths pures à Paris 6, DEA et thèse de maths à Paris 7 en logique, avec intérêt initial pour la littérature, le langage, la linguistique. Où ai-je développé cela ? Chercheur INRIA (Nancy, Rennes) puis prof à l'université de Bordeaux, puis à celle de Montpellier.*



## I.2. Plan

1. Présentation du traitement automatique des langues de la linguistique et des liens avec la logique.
2. De la phrase à des formules logiques représentant son ou ses sens. (Grammaires catégorielles, sémantique de Montague)
3. Extension :
  - sens des mots (sémantique lexicale, lexique génératif montagovien)
  - quantification en langage naturel et opérateurs de Hilbert  $\varepsilon$  et  $\tau$ .



## **II Traitement automatique des langues**



## II.1. Famille de problèmes à traiter

Deux tâches classiques (traduction= A puis B)

- A analyse automatique (de texte, de parole, annoté ou non)  
résultat ?
- B génération automatique (de texte, de parole)  
à partir de quoi ?

Domaine propre (recherche d'information) mais aussi utile pour A  
et B ci-dessus

— fouille, acquisition (des données pour les autre tâches ou RI)



## II.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

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

Geach était-il l'élève de Wittgenstein ?

Domaine relié : **recherche d'information** différence : Big Data, beaucoup de données, analyse superficielle, méthodes statistiques, machine learning

"ici, on ne vend pas de CD seulement des vinyles"

"production laitière / production de lait"  
mais "production minière / production de mine(s)"



### **III Les niveaux d'analyse de la langue & leurs méthodes**





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

#### **III.1.a Phonétique**

Acoustique / système phonatoire/auditif

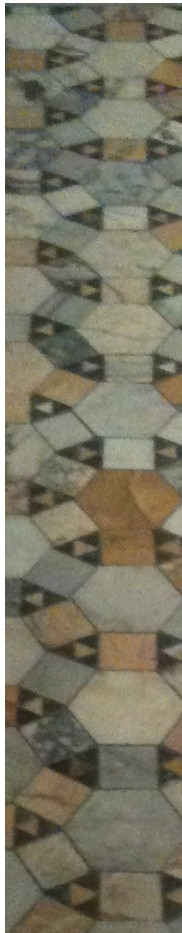
Traitement du signal / médecine

#### **III.1.b Phonologie**

Les sons abstraits : système discret (dans un continu)

(4) Bali / Paris indistincts pour un japonais

Automates et transducteurs



## III.2. Les mots : morphologie

### III.2.a Morphologie dérivationnelle

changement de catégorie possible

—

allumer/allumage

témoigner/témoignage

maquiller/maquillage

garer/garage

égarer/égarement

aménager/aménagement

—

maison /maisonnette

camion /camionnette

carpe/carpette



### III.2.b Morphologie flexionnelle

pas de changement de catégorie (automates et transducteurs)

cheval → chevaux

aller → allons

aller → va

aller → irons

### III.2.c Étiquetage grammatical

nom, verbe, déterminant etc.

Automates ou probas (modèles de Markov cachés)



### III.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 (CFG, TAG etc)

Déduction formelle (cf. ci-après)

Arbres satisfaisant des contraintes  
(par ex. "le sujet précède le verbe")



### **III.4. Le sens : sémantique**

#### **III.4.a Le sens d'un mot : sémantique lexicale**

Sens, restrictions de sélection, mots associés (logique, probabilités, graphes (petit monde), jeux)

Rôle téléique : être lu, informer, cultiver,

Rôle constitutif : pages, couvertures

Rôle agentif : imprimeur, auteur,

#### **III.4.b Le sens d'une phrase : sémantique compositionnelle, formelle**

Phrase -> formule logique

Sens : ensemble des mondes possibles dans lequel la phrase est vraie

(ces 2 approches sont compatibles mais ne coïncident pas forcément)



### III.5. L'interprétation en en contexte (discours, dialogue) : pragmatique

Prise en compte du contexte linguistique et extra linguistique

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



### III.6. Deux notions de sémantique

De quoi ça parle (analyse de texte, par des méthodes statistiques).

Qui fait quoi, ce qui est affirmé et réfuté (analyse de quelques phrases par des méthodes logiques).

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.

Dans cet exposé : syntaxe (déductive) et sémantique (logique) de la phrase et des mots.



## **IV Logique et langage : historique**





## IV.1. Syntaxe/grammaire et sémantique/logique 1

phrase = proposition qui peut être vrai ou fausse

les syllogismes sont formulés en langage naturel.

Baroco

tout conducteur montpelliérain est pressé A  
certains conducteurs ne sont pas pressés O

— donc —

certains conducteurs ne sont pas des conducteurs montpelliérains O



## IV.2. Syntaxe/grammaire et sémantique/logique 2

verbe intransitif = prédicat à un argument (=sujet)

verbe transitif = prédicat à deux arguments

nom commun = prédicat

adjectif = propriété ou propriété de propriété ? ?

Les étudiants sont des musiciens.

→ Les étudiants parisiens sont de(s) musiciens parisiens

↗ Les bons étudiants sont des bons musiciens.



### IV.3. De re / de dicto (Thomas d'Aquin, Ockham)

- (12)
- a. James Bond croit qu'un prof de l'ENS est un espion.
  - b. Il a vu Jean refermer son tiroir rapidement. (de re)
  - c. Il a trouvé un code sur un papier qui traînait. (de dicto)



#### IV.4. Sens/dénotation (Abélard, Frege)

- (13) a. L'étoile du matin (Vénus).  
b. L'étoile du berger. (Vénus)



## IV.5. Compositionnalité du sens

Le sens du tout est fonction du sens des parties (Frege) et de la structure syntaxique (Montague) :

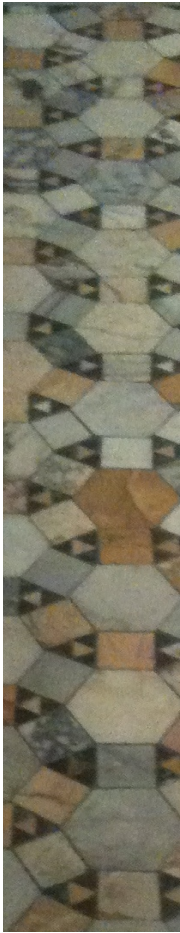
- (14) a. Marie aime Pierre.  
b. Pierre aime Marie.
- (15) a. Mon voisin que je n'apprécie guère n'aime pas la musique.  
b. Je n'apprécie guère la musique que mon voisin n'aime pas.
- (16) Si [un paysan<sub>i</sub> possède un âne<sub>k</sub>] alors [il<sub>i</sub> le<sub>k</sub> bat].



## IV.6. Différences

La langue diffère de l'interprétation logique :

- (17) a. J'avais trois clés dans ma poche, je les ai toutes perdues sauf une.  
b. Je la range dans un tiroir.
- (18) a. J'avais trois clés dans ma poche, j'en ai perdu deux.  
b. \* Je la range dans un tiroir.



## **V Quelques applications spécifiques**



## V.1. Exemple d'utilisation des méthodes formelles, logiques

Inférence textuelle (text entailment) :  
une phrase est elle conséquence d'un (petit) texte ?

extraire un système de de préférence (pour l'aide à la décision).

Re-construire a structure logique d'un dialogue, d'un discours,..

Analyse = satisfaction de contraintes  
(permet de traités d'énoncés dégradés)





## V.2. Reconstruction d'itinéraires

Projet Itipy (INRIA et ex. régions Aquitaine et Midi Pyrénées)

Corpus XVII<sup>e</sup> XX<sup>e</sup> 576 334 mots (surtout XIX<sup>e</sup>)

Récits de voyages dans les Pyrénées.

Peut-on reconstruire les itinéraires ?



### V.3. Exemples ITIPY

*Nous coupons ici un sentier qui vient du port de Barroude (...)*

*Plus loin, de nobles hêtres montent sur le versant (...) cette route qui monte sans cesse pendant deux lieues*

*Le chemin pavé de calcaire et de pierres luisantes (...) serpente à travers fourrés de buis et de noisetiers*

*Puis, cinq minutes nous conduisent à un petit pont (...) qui nous porte sur la rive droite.*

Problème : parenthèses ouvertes (comparaison avec d'autres voyages) mais dont la fermeture est peu claire (même pour un être humain).



## V.4. **Projet AREN Argumentation et numérique 1**

Université (info, pédagogie) + Académie de Montpellier + 3 start up + 2 associations

Plateforme de débats en ligne écrit par des classes de lycées sur des sujets interdisciplinaires à partir de synthèses scientifique (ControverSciences asso ENS Lyon).

ex : neuroéthique (SVT / philo)

1. surlignage d'un extrait du texte ou d'une intervention précédente
2. re formulation
3. choix (d'accord — pas d'accord — je ne comprends pas)
4. justification argumentée de quelque lignes



## V.5. **Projet AREN** Argumentation et numérique 2

Analyse automatique :

correction de la reformulation,

trouver les arguments (par opposition à des émotions),

nature des arguments,

comment se répondent ils ?

Structure du débat (S DRT graphe étiqueté avec une structure récursive) → Outil de navigation dans le débat ;

Restitution en classe, explications etc.



## V.6. Spéculatif : analyse de la créativité poétique

Pas la génération de poèmes (Oulipo)

Pas de statistiques sur les mots, les phonèmes etc mais plutôt une analyse logique standard qui va bloquer souvent en précisant la nature du blocage.

Poésie : presque pas de contexte extra linguistique

Image poétique réside dans ce qui résiste à l'analyse (par ex. métaphore : composition inattendue)

Par exemple :

*"La secrète noirceur du lait..."*

Jacques Audiberti cité par J.-Y. Girard dans "linear logic"



## V.7. Ensuite (= après la pause)

Phrase → formule logique

1. Phrase
  2. Analyse syntaxique dans une grammaire catégorielle (preuve dans une logique particulière)
  3. Conversion en lambda terme linéaire de type proposition
  4. Insertion du sens des mots
  5. Beta réduction
  6. Formule logique = sens de la phrase
- (19)
- a. Les enfants prendront une pizza.
  - b.  $\exists x. \text{pizza}(x) \wedge \forall w. (\text{enfant}(w) \Rightarrow \text{prendront}(w, x))$
  - c.  $\forall u. \text{enfants}(u) \Rightarrow \exists x. \text{pizza}(x) \wedge \text{prendront}(u, x)$

Puis : comment prendre en compte le sens lexical.

Et finalement : conclusion.