

# Programmation de Modèles Linguistiques 1, **L5SOPRG L3**

Crédits : Karën Fort (pépites pour le TAL)

---

Caroline Koudoro-Parfait  
[caroline.parfait@sorbonne-universite.fr](mailto:caroline.parfait@sorbonne-universite.fr)

Observatoire des Textes des Idées et des Corpus - Obtic,  
Sorbonne Center for Artificial Intelligence - SCAI,  
Sens Textes Informatiques Histoire - STIH EA 4509, Sorbonne Université

# Plan de la présentation

1. Introduction
2. Modélisation de la Langue et TAL
3. Analyser les caractères

## Organisation du semestre

- CM + TD le Vendredi de 8h00 à 12h00 (C. Koudoro-Parfait)
- Contrôle continu + projet + contrôle terminal

## Ressources

- TAL et Linguistique Informatique 1, ISTE Ed. (Mohamed Z. Kurdi)
- Speech and Language Processing (Dan Jurafsky),  
<https://web.stanford.edu/~jurafsky/slp3/>
- Helpdesk : mail ou bureau 206/211 à Serpente (sur RDV)

Introduction

Modélisation de la Langue et TAL

- Le TAL en 2 minutes

- Le TAL vu par ses applications

- Le TAL : définitions

- Ambiguïté à tous les étages

- Étapes de traitement

Analyser les caractères

# Introduction

---

# Contexte : l'accès aux données langagières



# Difficulté : des données conçues pour des humains



## Solution : des données sous-traitées à des machines





## Modéliser la langue

- Observables : le caractère, le phonème, le token
- Eco-système : n-grammes, sons, proposition/phrased
- Applications : identification de langue, d'auteur, de style ; indexation et génération de texte ; extraction d'information

# Modélisation de la Langue et TAL

---

- Traitement Automatique des Langues
- Traitement Automatique du Langage Naturel

- Traitement Automatique des Langues
- Traitement Automatique du Langage Naturel
- *Natural Language Processing*
- Technologies du Langage (Humain)
- Sciences de l'information
- Linguistique Computationnelle

Reoiajr oj earoij reoa o eo ao aeoi oj aroij aoeir eoaj.

Langue inconnue

# La langue vue par la machine

Reoiajr oj earoij reoa o eo ao aeoi oj aroij aeir eoaj.

Découper

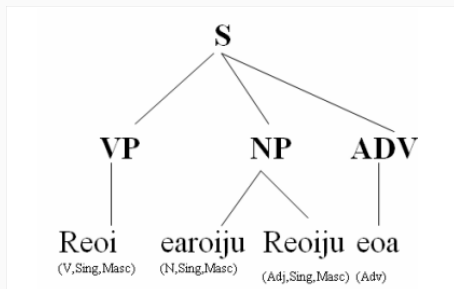
# La langue vue par la machine

Reoi	earoiju	Reoiju	eo
(V,Sing,Masc)	(N,Sing,Masc)	(Adj,Sing,Masc)	(Adv)

Reoiajr	oj	earoij	reoa	o	eo	ao	aeoi	oj	aeroij	aoeir	eoaj.
---------	----	--------	------	---	----	----	------	----	--------	-------	-------

Etiqueter

# La langue vue par la machine

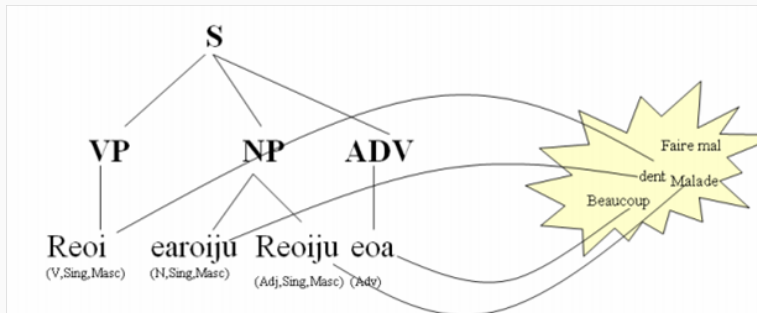


Reoiajr oj earoij reoa o eo ao aeoi oj aeroij aeir eoaj.

Reconnaître la structure



# La langue vue par la machine



Reoiajr oj earoij reoa o eo ao aeoi oj aeroij aoeir eoaj.

Évaluer le sens et mettre en contexte

Nécessité d'explicitation (réductionnisme)

# Pour quoi faire ?

- Traduction automatique
- Recherche d'information
- Indexation
- Traitement d'Emails
- Veille
- E-reputation
- Big Data
- ...

# TAL : Modéliser l'activité langagière

Pour les chercheurs :

- comprendre comment fonctionne la langue
- exemplifier/vérifier des théories

Pour les industriels :

- reproduire l'activité langagière
- optimiser les performances des systèmes
- offrir de nouveaux services

## Introduction

### Modélisation de la Langue et TAL

- Le TAL en 2 minutes

- Le TAL vu par ses applications

- Le TAL : définitions

- Ambiguïté à tous les étages

- Étapes de traitement

### Analyser les caractères

- *La mousse aux fraises est sur la table de l'avocat.* – polysémie

## Le TAL : pas si simple. . .

- *La mousse aux fraises est sur la table de l'avocat.* – polysémie
- *L'omelette au lard est parti sans payer!* – métonymie

## Le TAL : pas si simple. . .

- *La mousse aux fraises est sur la table de l'avocat.* – polysémie
- *L'omelette au lard est parti sans payer!* – métonymie
- *Le bus a renversé un passant. . .* – métonymie

## Le TAL : pas si simple. . .

- *La mousse aux fraises est sur la table de l'avocat.* – polysémie
- *L'omelette au lard est parti sans payer!* – métonymie
- *Le bus a renversé un passant. . .* – métonymie
  - . . . je l'ai entendu freiner. –Contexte



## Le TAL : pas si simple...

- *La mousse aux fraises est sur la table de l'avocat.* – polysémie
- *L'omelette au lard est parti sans payer!* – métonymie
- *Le bus a renversé un passant...* – métonymie
  - ... je l'ai entendu freiner. – Contexte
  - ... je l'ai entendu crier.

## Le TAL : pas si simple...

- *La mousse aux fraises est sur la table de l'avocat.* – polysémie
- *L'omelette au lard est parti sans payer!* – métonymie
- *Le bus a renversé un passant...* – métonymie
  - ... je l'ai entendu freiner. – Contexte
  - ... je l'ai entendu crier.
- *Le professeur a envoyé l'élève chez le proviseur...*

## Le TAL : pas si simple...

- *La mousse aux fraises est sur la table de l'avocat.* – polysémie
- *L'omelette au lard est parti sans payer!* – métonymie
- *Le bus a renversé un passant...* – métonymie
  - ... je l'ai entendu freiner. – Contexte
  - ... je l'ai entendu crier.
- *Le professeur a envoyé l'élève chez le proviseur...*
  - ... il faisait trop de bruit.

# Le TAL : pas si simple...

- *La mousse aux fraises est sur la table de l'avocat.* – polysémie
- *L'omelette au lard est parti sans payer!* – métonymie
- *Le bus a renversé un passant...* – métonymie
  - ... je l'ai entendu freiner. –Contexte
  - ... je l'ai entendu crier.
- *Le professeur a envoyé l'élève chez le proviseur...*
  - ... il faisait trop de bruit.
  - ... il était excédé.

# Le TAL : pas si simple...

- *La mousse aux fraises est sur la table de l'avocat.* – polysémie
- *L'omelette au lard est parti sans payer!* – métonymie
- *Le bus a renversé un passant...* – métonymie
  - ... *je l'ai entendu freiner.* – Contexte
  - ... *je l'ai entendu crier.*
- *Le professeur a envoyé l'élève chez le proviseur...*
  - ... *il faisait trop de bruit.*
  - ... *il était excédé.*
  - ... *il l'avait convoqué.*

# Le TAL : pas si simple...

- *La mousse aux fraises est sur la table de l'avocat.* – polysémie
- *L'omelette au lard est parti sans payer!* – métonymie
- *Le bus a renversé un passant...* – métonymie
  - ... *je l'ai entendu freiner.* – Contexte
  - ... *je l'ai entendu crier.*
- *Le professeur a envoyé l'élève chez le proviseur...*
  - ... *il faisait trop de bruit.*
  - ... *il était excédé.*
  - ... *il l'avait convoqué.*

→ **ambiguïtés** pour les systèmes et/ou pour les humains

→ **inférences** difficiles à recenser exhaustivement

# Le découpage en "mots" ou tokenization

- *l'arbre*

# Le découpage en "mots" ou tokenization

- *l'arbre*
- *aujourd'hui*



# Le découpage en "mots" ou tokenization

- *l'arbre*
- *aujourd'hui*
- *plateforme*

# Le découpage en "mots" ou tokenization

- *l'arbre*
- *aujourd'hui*
- *plateforme*
- *plate-forme*

# Le découpage en "mots" ou tokenization

- *l'arbre*
- *aujourd'hui*
- *plateforme*
- *plate-forme*
- *dit-il*

# Le découpage en "mots" ou tokenization

- *l'arbre*
- *aujourd'hui*
- *plateforme*
- *plate-forme*
- *dit-il*
- *cé-*  
*sure*

# Le découpage en "mots" ou tokenization

- *l'arbre*
- *aujourd'hui*
- *plateforme*
- *plate-forme*
- *dit-il*
- *cé-*  
*sure*

# Le découpage en "mots" ou tokenization

- *l'arbre*
- *aujourd'hui*
- *plateforme*
- *plate-forme*
- *dit-il*
- *cé-*  
*sure*
- Vablaheibárvégavinnuverkfærageymsluskúraútidyralklakkippuhringur  
("Porte-clés de la chaîne de clé pour la porte extérieure du hangar à  
outils des agents de la route sur le plateau Vablaheibi")

la *porte*

- *porte* + Nom + Fem + Sg

la *porte*

- *porte* + Nom + Fem + Sg
- *porte* + VT + 1/3P + Sg



*Jean regarde un homme sur la colline avec un télescope.*

- **Qui** est sur la colline ?
- **Qui** a un télescope ?
- Jean regarde [un homme sur la colline avec un télescope]

*Jean regarde un homme sur la colline avec un télescope.*

- **Qui** est sur la colline ?
- **Qui** a un télescope ?
- Jean regarde [un homme sur la colline avec un télescope]
- Jean regarde [un homme sur la colline] avec un télescope

*Jean regarde un homme sur la colline avec un télescope.*

- **Qui** est sur la colline ?
- **Qui** a un télescope ?
- Jean regarde [un homme sur la colline avec un télescope]
- Jean regarde [un homme sur la colline] avec un télescope
- Jean regarde un homme [sur la colline avec un télescope]

*Jean regarde un homme sur la colline avec un télescope.*

- **Qui** est sur la colline ?
- **Qui** a un télescope ?
- Jean regarde [un homme sur la colline avec un télescope]
- Jean regarde [un homme sur la colline] avec un télescope
- Jean regarde un homme [sur la colline avec un télescope]

*Jean regarde un homme sur la colline avec un télescope.*

- **Qui** est sur la colline ?
- **Qui** a un télescope ?
- Jean regarde [un homme sur la colline avec un télescope]
- Jean regarde [un homme sur la colline] avec un télescope
- Jean regarde un homme [sur la colline avec un télescope]

## Un défi à l'intuition

Moi "... Vous avez oublié la pièce jointe..."

Le client mail :

2 mots-clés de pièce jointe trouvés : **pièce jointe, jointe**

[Ajouter une pièce jointe...](#)

[Me le rappeler plus tard](#)

*Tous les hommes aiment une femme.*

→ **Chaque** homme aime une femme ou

→ **tous** les hommes aiment la même femme ?

*Tous les hommes aiment une femme.*

→ **Chaque** homme aime une femme ou

→ **tous** les hommes aiment la même femme ?

*The Wayward Pines Trilogy (2012–2014) is a mystery/thriller/science fiction novel series by American author Blake Crouch.*

→ **Book** = "The Wayward Pines Trilogy", **Date** = "2012-2014",  
**Author** = "Blake Crouch"

*Tous les hommes aiment une femme.*

→ **Chaque** homme aime une femme ou

→ **tous** les hommes aiment la même femme ?

*The Wayward Pines Trilogy (2012–2014) is a mystery/thriller/science fiction novel series by American author Blake Crouch.*

→ **Book** = "The Wayward Pines Trilogy", **Date** = "2012-2014",  
**Author** = "Blake Crouch"

*Jean-Marie Roughol, 47 ans, sort un livre écrit à quatre mains avec le président du Conseil constitutionnel, Jean-Louis Debré*

→ **Book** = "un livre écrit à quatre mains", **Date** = "Today",  
**Author** = "Jean-Marie Roughol, Président du Conseil Constitutionnel"



Idiosyncrasie : Ox-Beef VS Bœuf

Idiosyncrasie : Ox-Beef VS Bœuf

## Différences de représentation et de combinaison

Русский язык	Français
первый	Premier
второй	Deuxième
этаж	Etage
??	Premier étage

Idiosyncrasie : Ox-Beef VS Bœuf

## Différences de représentation et de combinaison

Русский язык	Français
первый	Premier
второй	Deuxième
этаж	Etage
??	Premier étage
второй этаж	Premier étage
первый этаж	Rez-de-chaussée
резь-де-Chaussee	Rez-de-chaussée

## **Analyser les caractères**

---

En pratique pour la machine ? Les difficultés de programmation algorithmique sur des données textuelles :

- Comment la machine reconnaît une chaîne de caractère comme un mot ?
- Comment la machine reconnaît des phrases ?

Puis TD3 : Identification de Langue exploitant des propriétés génériques

- Pourquoi ?
- Usages ?
- Méthodes ?