

Éléments de linguistique

Béatrice Daille

Faculté des Sciences et Techniques de Nantes
Département Informatique

2020-2021

- **Définition**
- **Niveaux de traitement de la langue**
- **Ambiguïtés**
- **Axes d'analyse**

Définition

Le TALN s'intéresse aux traitements informatisés mettant en jeu du matériau linguistique.

langage naturel \neq langage informatique

Niveaux de traitement de la langue

Modèle de langue :

- ① Phonétique et Phonologie
- ② Morphologie
- ③ Syntaxe
- ④ Sémantique
- ⑤ Pragmatique

Phonétique : étude des sons d'une langue

Phonologie :

- étude des sons qui ont une valeur distinctive ;
Phonème : la plus petite unité de son qui permet de distinguer une unité de sens d'une autre.
- règles qui régissent les sons d'une langue, les associations possibles entre eux et les relations avec les niveaux suivants.

Morphologie : étude des mots indépendamment de la phrase.

Morphème : unité significative minimale.

Étude des contraintes entre les catégories morphosyntaxiques des mots pour former des phrases **correctes**.

Train a ne pour la le

La fille ont dormi

Je peux que je parte

En français, une forme syntaxique admissible est SUJET VERBE
OBJET

- ① Grammaires de constituants
- ② Grammaires de dépendances

Étude du sens des mots, des relations et des contraintes entre les sens des mots d'une phrase pour que celle-ci soit **compréhensible**.

Étude du sens en dehors de tout contexte particulier.

Catégories sémantiques : classes regroupant de mots dont le sens est proche ou des mots qui possèdent certaines propriétés sémantiques communes.

Relations sémantiques :

- hyperonymie/hyponymie : *vaisseau/artère*
- méronymie (partie d'un tout) : *vaisseau/système cardiovasculaire*
- antonymie : *malin/benin*

D'incolores idées vertes dorment furieusement.

La sincérité admire Pierre.

La vertu est triangulaire.

Les géants sont petits.

Jean mange une pomme

Contraintes sémantiques :

manger : 2 arguments ; 2 rôles thématiques différents

- **agent** : animé
- **objet** : solide, comestible

Étude des relations entre le sens effectif de ce qui est énoncé et son interprétation.

Étude du sens dans un contexte particulier.

- **Implicite**

Le mois de septembre était très beau. Le vin sera bon cette année.

→ Le vin est fait avec du raisin. Les vendanges ont souvent lieu en septembre. Le soleil fait mûrir les raisins.

- **Mécanisme du discours**

- **Cohésion** : continuité du texte : anaphores, homogénéité du thème, ellipses, etc.
- **Cohérence** : structure de discours, relations causales, temporelles, etc.

Niveaux de traitement en TALN

- Autant niveaux de traitement que de niveaux de description de la langue :
 - **Identification des unités de la langue** : pré-syntaxe, pré-traitement linguistique, détection des phrasèmes
 - **Analyser les mots**
 - **Analyser les phrases**
 - **Construire une représentation du sens**
 - **Analyser le discours**

Remarques :

- ambiguïtés à chaque niveau ;
- interférence entre les niveaux.

Ambiguïtés

- **graphie/son :**

x → [ks] *axe*; [s] *six*; [z] *sixième*; [gz] *exact*

ε → *thèse, fête, laid, tes, taies*, etc.

- **homographes/homophones :**

fils/foie, fois, foi

- **homonymes :** *produits, vase, avocat*

Ambiguïtés au niveau de la phrase

- **Ambiguïtés phonétiques :**

*Gall, amant de la reine, alla, tour magnanime,
galamment de l'arène à la tour Magne à Nîmes.
Victor Hugo*

- **Ambiguïtés lexicales :**

Il faut le pouvoir. La belle ferme le cache.

Ambiguïtés au niveau de la phrase

- **Ambiguïtés structurales :**

- 1. Attachement prépositionnel

- Luc a rapporté un vase de Chine.*

- Luc a parlé de ce problème à Noël*

- 2. Complétive/Relative

- J'ai l'impression que vous ne mangez pas*

- Je comprends la forte impression qu'il vous a faite*

- **Ambiguïtés sémantiques :**

- David veut épouser une Norvégienne.*

- **Ambiguïtés pragmatiques :**

- Références pronominales

Le professeur a puni l'élève. Il lui avait lancé des boulettes.

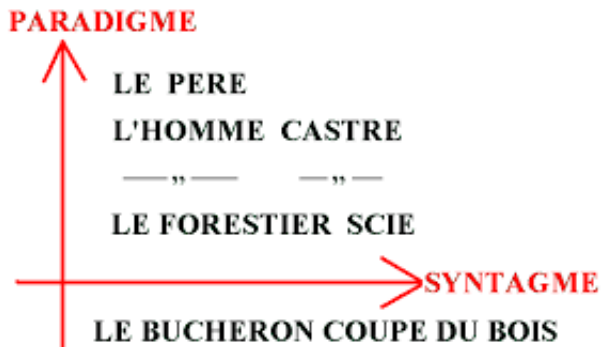
Quand Sue va dîner chez Ève, elle mange du cassoulet.

Quand Sue va dîner chez Ève, elle sert du cassoulet.

Axes d'analyse

F. de Saussure Cours de linguistique générale (1910)

Axe syntagmatique/ axe paradigmatique



Deux approches en TALN

- **analyse syntagmatique, approche par analyse des co-occurrences**

Identifier les mots qui apparaissent dans une même portion de texte, dans une phrase, dans un syntagme

- **analyse distributionnelle** Identifier les mots qui partagent les contextes sans nécessairement apparaître ensemble

Méthodes de construction de classes sémantiques : algorithme des k-moyennes