

# Traitement Automatique du Langage

## Cours 1 - Syntaxe

Antoine Rozenknop  
antoine.rozenknop@lipn.univ-paris13.fr

24 septembre 2008

# Plan

Introduction

La syntaxe

Notions de Grammaticalité et d'Interprétabilité

Définitions :

Les constituants syntaxiques

Organisation de la phrase en syntagmes

Les tests classiques

Organisation hiérarchique des syntagmes

Grammaires génératives et règles de réécriture

Règles de réécriture

Les arbres syntaxiques

Les parenthèses étiquetées

Quelques difficultés du traitement syntaxique

# Plan

## Introduction

### La syntaxe

Notions de Grammaticalité et d'Interprétabilité

Définitions :

### Les constituants syntaxiques

Organisation de la phrase en syntagmes

Les tests classiques

Organisation hiérarchique des syntagmes

### Grammaires génératives et règles de réécriture

Règles de réécriture

Les arbres syntaxiques

Les parenthèses étiquetées

### Quelques difficultés du traitement syntaxique

# Traitements de texte

*Le président des antialcooliques mangeait une orange avec un couteau.*

- **Segmenter** en unités lexicales

*“le”, “président”, “des”, ...*

# Traitements de texte

*Le président des antialcooliques mangeait une orange avec un couteau.*

- ▶ **Segmenter** en unités lexicales
- ▶ **analyse lexicale** : identifier les composants lexicaux et leurs propriétés

*“le” : article défini, masculin, singulier*

# Traitements de texte

*Le président des antialcooliques mangeait une orange avec un couteau.*

- ▶ **Segmenter** en unités lexicales
- ▶ **analyse lexicale** : identifier les composants lexicaux et leurs propriétés
- ▶ **analyse syntaxique** : identifier des groupes et leurs relations  
*“le président”, sujet de “mangeait une orange avec un couteau”*

# Traitements de texte

*Le président des antialcooliques mangeait une orange avec un couteau.*

- ▶ **Segmenter** en unités lexicales
- ▶ **analyse lexicale** : identifier les composants lexicaux et leurs propriétés
- ▶ **analyse syntaxique** : identifier des groupes et leurs relations
- ▶ **analyse sémantique** : représentation du sens

*il y a un président qui mange une orange avec un couteau*

# Traitements de texte

*Le président des antialcooliques mangeait une orange avec un couteau.*

- ▶ **Segmenter** en unités lexicales
- ▶ **analyse lexicale** : identifier les composants lexicaux et leurs propriétés
- ▶ **analyse syntaxique** : identifier des groupes et leurs relations
- ▶ **analyse sémantique** : représentation du sens
- ▶ **analyse pragmatique** : identifier enfin la fonction de l'énoncé dans le contexte

*le président Bernard Poivrot mange l'orange  
emoisonnée avec le couteau magique qu'il a trouvé  
au fond de son trou*



# Plan

Introduction

La syntaxe

Notions de Grammaticalité et d'Interprétabilité

Définitions :

Les constituants syntaxiques

Organisation de la phrase en syntagmes

Les tests classiques

Organisation hiérarchique des syntagmes

Grammaires génératives et règles de réécriture

Règles de réécriture

Les arbres syntaxiques

Les parenthèses étiquetées

Quelques difficultés du traitement syntaxique

# Plan

Introduction

La syntaxe

Notions de Grammaticalité et d'Interprétabilité

Définitions :

Les constituants syntaxiques

Organisation de la phrase en syntagmes

Les tests classiques

Organisation hiérarchique des syntagmes

Grammaires génératives et règles de réécriture

Règles de réécriture

Les arbres syntaxiques

Les parenthèses étiquetées

Quelques difficultés du traitement syntaxique

# Grammatical ?

- ▶ 1. Cette valise est pleine à craquer.  
2. \* A craquer, cette valise est pleine.
- ▶ 1. Ce nom est difficile à prononcer.  
2. A prononcer, ce nom est difficile.

# Ambigu ?

- ▶ 1. Tout le monde admire son enfant.  
2. Son enfant est admiré par tout le monde.
- ▶ 1. Tous les enfants ont vu une sorcière.  
2. Une sorcière a été vue par tous les enfants.

# Plan

Introduction

La syntaxe

Notions de Grammaticalité et d'Interprétabilité

Définitions :

Les constituants syntaxiques

Organisation de la phrase en syntagmes

Les tests classiques

Organisation hiérarchique des syntagmes

Grammaires génératives et règles de réécriture

Règles de réécriture

Les arbres syntaxiques

Les parenthèses étiquetées

Quelques difficultés du traitement syntaxique

# Définitions

La **syntaxe** est l'étude scientifique de la construction des phrases. Elle vise à formuler les régularités sous-jacentes à leur organisation, c'est-à-dire qu'elle vise à déterminer les règles (principes) qui gouvernent les relations de combinaison et de dépendance entre les mots et les groupes de mots au sein de la phrase.

La **grammaire** est la description des contraintes caractéristiques d'**une langue donnée**. La grammaire est donc constituée des principes syntaxiques universels augmentés des principes spécifiques à une langue.

# Pourquoi la syntaxe en TAL ?

- ▶ Pour la **génération** automatique d'énoncés ;
- ▶ Pour l'**interprétation** d'énoncés ?

# Pourquoi la syntaxe en TAL ?

- ▶ Pour la **génération** automatique d'énoncés ;
- ▶ Pour l'**interprétation** d'énoncés :
  - ▶ Désambiguïser les étiquettes morphosyntaxiques

Exemple :

|               |               |              |               |                     |
|---------------|---------------|--------------|---------------|---------------------|
| <b>La</b>     | <b>petite</b> | <b>ferme</b> | <b>le</b>     | <b>voile</b>        |
| <i>Art</i>    | <i>Nom</i>    | <i>Nom</i>   | <i>Art</i>    | <i>Nom féminin</i>  |
| <i>Pronom</i> | <i>Adj.</i>   | <i>Verbe</i> | <i>Pronom</i> | <i>Verbe</i>        |
| <i>Nom</i>    |               | <i>Adj.</i>  |               | <i>Nom masculin</i> |



# Pourquoi la syntaxe en TAL ?

- ▶ Pour la **génération** automatique d'énoncés ;
- ▶ Pour l'**interprétation** d'énoncés :
  - ▶ Désambiguïser les étiquettes morphosyntaxiques
  - ▶ Profiter de la structure des phrases.

# Plan

Introduction

La syntaxe

Notions de Grammaticalité et d'Interprétabilité

Définitions :

Les constituants syntaxiques

Organisation de la phrase en syntagmes

Les tests classiques

Organisation hiérarchique des syntagmes

Grammaires génératives et règles de réécriture

Règles de réécriture

Les arbres syntaxiques

Les parenthèses étiquetées

Quelques difficultés du traitement syntaxique

# Plan

Introduction

La syntaxe

Notions de Grammaticalité et d'Interprétabilité

Définitions :

Les constituants syntaxiques

Organisation de la phrase en syntagmes

Les tests classiques

Organisation hiérarchique des syntagmes

Grammaires génératives et règles de réécriture

Règles de réécriture

Les arbres syntaxiques

Les parenthèses étiquetées

Quelques difficultés du traitement syntaxique

# Mise en évidence des syntagmes

► Mon frère déteste cet homme mesquin.

► Mon frère trouve cet homme mesquin.

# Mise en évidence des syntagmes

- ▶ Mon frère déteste cet homme mesquin.
  - C'est **cet homme mesquin** que mon frère déteste.
  - Cet homme mesquin**, Mon frère **le** déteste.
  - Cet homme mesquin** est détesté par mon frère.
  - Qui ton frère déteste-t-il ?
  - **Cet homme mesquin**.
- ▶ Mon frère trouve cet homme mesquin.

# Mise en évidence des syntagmes

- ▶ Mon frère déteste cet homme mesquin.  
C'est **cet homme mesquin** que mon frère déteste.  
**Cet homme mesquin**, Mon frère **le** déteste.  
**Cet homme mesquin** est détesté par mon frère.
  - Qui ton frère déteste-t-il ?
  - **Cet homme mesquin.**
- ▶ Mon frère trouve cet homme mesquin.  
C'est **cet homme** que mon frère trouve **mesquin**  
**Cet homme**, mon frère **le** trouve **mesquin**
  - Qui ton frère trouve-t-il **mesquin** ?
  - **Cet homme.**

# Tête de syntagme

L'étudiant dort sur son bureau  
Sujet-----> verbe <----- C. C. L.

Paul                donne                une fleur                à Marie  
Sujet -----> verbe <----- C.O.D.                C.O.I.  
  
                                <-----

---

# Plan

Introduction

La syntaxe

Notions de Grammaticalité et d'Interprétabilité

Définitions :

Les constituants syntaxiques

Organisation de la phrase en syntagmes

Les tests classiques

Organisation hiérarchique des syntagmes

Grammaires génératives et règles de réécriture

Règles de réécriture

Les arbres syntaxiques

Les parenthèses étiquetées

Quelques difficultés du traitement syntaxique



# Tests d'identification

## ► La substitution

# Tests d'identification

- ▶ La substitution
  - ▶ La pronominalisation

# Tests d'identification

- ▶ La substitution
  - ▶ La pronominalisation
  - ▶ La variation paradigmaticque

# Tests d'identification

- ▶ La substitution
- ▶ Le déplacement

# Tests d'identification

- ▶ La substitution
- ▶ Le déplacement
  - ▶ Détachement

# Tests d'identification

- ▶ La substitution
- ▶ Le déplacement
  - ▶ Détachement
  - ▶ Le clivage

# Tests d'identification

- ▶ La substitution
- ▶ Le déplacement
  - ▶ Détachement
  - ▶ Le clivage
  - ▶ La dislocation (à gauche ou à droite)

# Tests d'identification

- ▶ La substitution
- ▶ Le déplacement
- ▶ La transformation



# Tests d'identification

- ▶ La substitution
- ▶ Le déplacement
- ▶ La transformation
  - ▶ Le passif

# Tests d'identification

- ▶ La substitution
- ▶ Le déplacement
- ▶ La transformation
  - ▶ Le passif
  - ▶ L'Interrogatif

# Tests d'identification

- ▶ La substitution
- ▶ Le déplacement
- ▶ La transformation
  - ▶ Le passif
  - ▶ L'Interrogatif
    - ▶ Interrogation partielle

# Tests d'identification

- ▶ La substitution
- ▶ Le déplacement
- ▶ La transformation
  - ▶ Le passif
  - ▶ L'Interrogatif
    - ▶ Interrogation partielle
    - ▶ Interrogation totale

# Tests d'identification

- ▶ La substitution
- ▶ Le déplacement
- ▶ La transformation
- ▶ La conjonction (de coordination)

# Plan

Introduction

La syntaxe

Notions de Grammaticalité et d'Interprétabilité

Définitions :

Les constituants syntaxiques

Organisation de la phrase en syntagmes

Les tests classiques

Organisation hiérarchique des syntagmes

Grammaires génératives et règles de réécriture

Règles de réécriture

Les arbres syntaxiques

Les parenthèses étiquetées

Quelques difficultés du traitement syntaxique

# Au calme

- Marie pense au calme

# Au calme

► Marie pense au calme

1. Marie pense à quelque chose ; ce quelque chose c'est le calme.



# Au calme

## ► Marie pense au calme

1. Marie pense à quelque chose ; ce quelque chose c'est le calme.
2. Marie pense et elle le fait au calme.

# Au calme

- ▶ Marie pense au calme
  1. Marie pense à quelque chose ; ce quelque chose c'est le calme.
  2. Marie pense et elle le fait au calme.
- ▶ J'aspire au calme !

# Plan

Introduction

La syntaxe

Notions de Grammaticalité et d'Interprétabilité

Définitions :

Les constituants syntaxiques

Organisation de la phrase en syntagmes

Les tests classiques

Organisation hiérarchique des syntagmes

Grammaires génératives et règles de réécriture

Règles de réécriture

Les arbres syntaxiques

Les parenthèses étiquetées

Quelques difficultés du traitement syntaxique

# Définition d'une phrase

La phrase est l'unité privilégiée en syntaxe, l'unité maximale de la description. Elle est une suite d'unités *hiérarchisées* ; ces unités hiérarchisées sont liées entre elles par des règles de “réécriture” ou règles “syntagmatiques”.

# Plan

Introduction

La syntaxe

Notions de Grammaticalité et d'Interprétabilité

Définitions :

Les constituants syntaxiques

Organisation de la phrase en syntagmes

Les tests classiques

Organisation hiérarchique des syntagmes

Grammaires génératives et règles de réécriture

Règles de réécriture

Les arbres syntaxiques

Les parenthèses étiquetées

Quelques difficultés du traitement syntaxique

# Exemple de règles de réécriture

## Règles

- 1)  $P \longrightarrow SN \quad SV$
- 2)  $SV \longrightarrow V$
- 3)  $SV \longrightarrow V \quad SN$
- 4)  $SN \longrightarrow \text{Dét} \quad \text{Adj.} \quad N$

## Exemple de règles de réécriture

## Règles

- 1) P  $\longrightarrow$  SN SV
- 2) SV  $\longrightarrow$  V
- 3) SV  $\longrightarrow$  V SN
- 4) SN  $\longrightarrow$  Dét Adj. N

## Dérivation

|   |   |     |       |        |         |      |        |      |  |
|---|---|-----|-------|--------|---------|------|--------|------|--|
|   |   | P   |       |        |         |      |        |      |  |
| 1 | ⇒ | SN  |       |        | SV      |      |        |      |  |
| 4 | ⇒ | Dét | Adj.  | N      | SV      |      |        |      |  |
| 3 | ⇒ | Dét | Adj.  | N      | V       | SN   |        |      |  |
| 4 | ⇒ | Dét | Adj.  | N      | V       | Dét. | Adj.   | N    |  |
|   | ⇒ | Le  | petit | garçon | regarde | la   | grosse | dame |  |

# Plan

Introduction

La syntaxe

Notions de Grammaticalité et d'Interprétabilité

Définitions :

Les constituants syntaxiques

Organisation de la phrase en syntagmes

Les tests classiques

Organisation hiérarchique des syntagmes

Grammaires génératives et règles de réécriture

Règles de réécriture

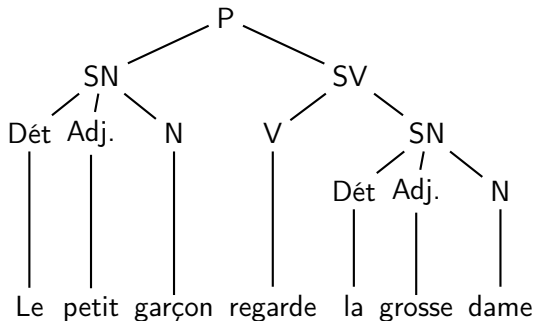
Les arbres syntaxiques

Les parenthèses étiquetées

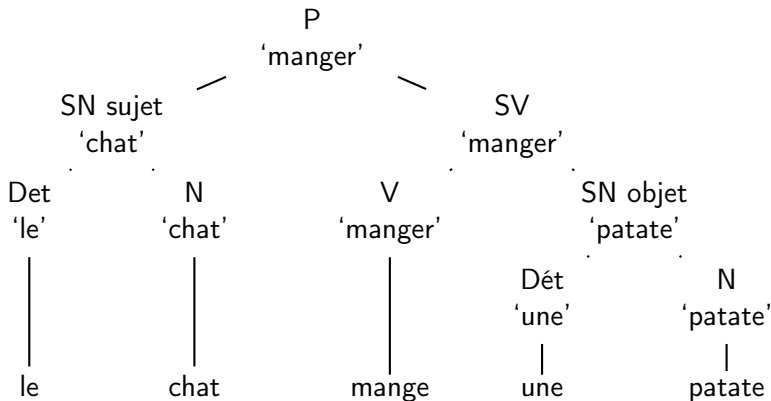
Quelques difficultés du traitement syntaxique



# Représentation sous forme d'arbre



# Autre représentation sous forme d'arbre



# Plan

Introduction

La syntaxe

Notions de Grammaticalité et d'Interprétabilité

Définitions :

Les constituants syntaxiques

Organisation de la phrase en syntagmes

Les tests classiques

Organisation hiérarchique des syntagmes

Grammaires génératives et règles de réécriture

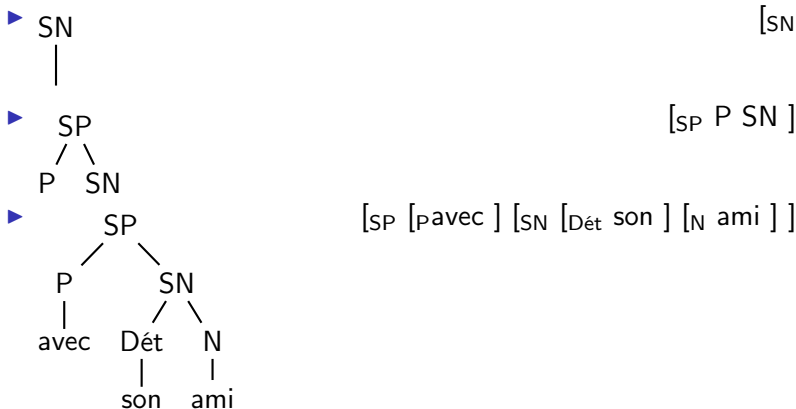
Règles de réécriture

Les arbres syntaxiques

Les parenthèses étiquetées

Quelques difficultés du traitement syntaxique

# Représentation en parenthèses étiquetées



# Plan

Introduction

La syntaxe

Notions de Grammaticalité et d'Interprétabilité

Définitions :

Les constituants syntaxiques

Organisation de la phrase en syntagmes

Les tests classiques

Organisation hiérarchique des syntagmes

Grammaires génératives et règles de réécriture

Règles de réécriture

Les arbres syntaxiques

Les parenthèses étiquetées

Quelques difficultés du traitement syntaxique

# Principales difficultés

- ▶ lutter contre la prolifération des ambiguïtés ;
- ▶ décrire des phénomènes complexes (et subtiles).

# Ambiguïtés lexicales

- ▶ porte : nom féminin, verbe conjugué
- ▶ mousse : nom féminin, nom masculin, verbe conjugué
- ▶ la : déterminant, ou pronom personnel féminin singulier, ou nom masculin (note de musique)

Si l'on se limite aux simples catégories syntaxiques de base, environ 50% des mots d'un texte sont ambigus → une phrase de 20 mots a environ  $2^{10}$  interprétations différentes au niveau des étiquettes morphosyntaxiques.

## Phénomènes de translation

Il s'agit d'un phénomène grammatical qui permet à une forme de changer de catégorie morphosyntaxique. Ainsi, tous les adjectifs, participes passés sont rendus ambigus si l'on veut que la grammaire soit capable de prendre en compte ce phénomène. Ex :

- ▶ vache : il est vraiment vache (emploi adjectival d'un nom)
- ▶ vert : mangez du vert
- ▶ blessé : il est blessé (vient du participe passé : il a blessé), le blessé



## Schémas de sous-catégorisation

- ▶ Jean parle.
- ▶ Jean parle à Marie.
- ▶ Jean parle de Paul.
- ▶ Jean parle de Paul à Marie.

← quatre entrées lexicales pour “parler”.

Parfois, cela provoque une ambiguïté dans l'énoncé :

- ▶ “je parle à la maîtresse de Marie”

# Ambiguïtés de rattachement

Ambiguïté purement syntaxique.

- ▶ Elle mange une pomme avec les doigts / Elle mange une pomme avec la peau
- ▶ Elle mange une glace à la fraise / Elle mange une glace à la plage
- ▶ C'est la fille du cousin qui boit
- ▶ Il a parlé de déjeuner avec Paul.