

# **Prédiction Structuré et Analyse Sémantique**

# Prédiction Structuré

**Nous avons des algorithmes de classification avec différents architectures**

**La plupart du temps on cherche à prédire une structure:**

- **La syntaxe → Un arbre**
- **La sémantique → Un graphe avec des entités et des relations**
- **Génération de texte → Une phrase que respecte les règles grammaticales/lexicales**

**Souvent il y a une structure implicite, lorsque un label dépends des autres**

- **POS Tagging**
- **Système de dialogue**

# Détection de chunks ou entités nommés

Une structure simple est l'encodage de étiquettes multi-mots (IO, BIO, BIOW, BIEOW)

## Chunker

<i>Wolfgang</i>	<i>Amadeus</i>	<i>Mozart</i>	,	<i>né</i>	<i>à</i>	<i>Salzburg</i>	<i>est</i>	<i>un</i>	<i>compositeur</i>
I	I	I	O	O	O	I	O	O	O
B	I	I	O	O	O	B	O	O	O
B	I	I	O	O	O	W	O	O	O
B	I	E	O	O	O	W	O	O	O

## Name Entity Recognition

<i>Wolfgang</i>	<i>Amadeus</i>	<i>Mozart</i>	,	<i>né</i>	<i>à</i>	<i>Salzburg</i>	<i>est</i>	<i>un</i>	<i>compositeur</i>
B-PER	I-PER	I-PER	O	O	O	B-LOC	O	O	O

# Analyse Sémantique

La sémantique étudiée ici :

- Sémantique lexicale : comprendre le sens des mots et leur rattachement dans une phrase

Faire une analyse du texte brut pour extraire :

- Qui? quoi? à qui? Quand? Où? Pourquoi? De quelle façon? De quelle manière?

Cette analyse comprend :

- Les évènements → La souris mange le fromage
- Les relations → La souris est un mammifère
- Les états → La voiture accidentée sur la D76

Un problème de prédiction structurée

- Certains mots sont des déclencheurs → manger, être, accidentée
- Ces mots ont des arguments associés → Qui(La souris), Quoi(le fromage), Où(sur la D76)
- Possible ambiguïté lexicale à tous les niveaux

# Analyse Sémantique



J' ai commandé une glace au serveur



J' ai commandé une glace au serveur



J' ai commandé une glace au serveur  
J' ai demandé une glace au serveur



J' ai commandé une glace au serveur  
J' ai demandé une glace au serveur

# Analyse Sémantique

## La représentation sémantique idéale doit :

- Décrire clairement les actions, évènements, relations et états
- Définir les arguments sémantiques possibles pour chaque cas
- Définir l'importance des arguments (Core / Non-core)
- Gérer la polysémie → commander une armée  $\neq$  commander une glace
- Gérer la synonymie → commander une glace = demander une glace
- Gérer les équivalences → John achète X à Frank = Frank vend X à John
- Compréhension des chaînes de coréférence
- Modélisation des relations spatiales et temporelles
- Modélisation des structures logiques (négation, quantification, etc..)

# Analyse Sémantique

Il existe une grande variété de modèles sémantiques

Le plus populaires

- **PropBank**
- **FrameNet**
- **AMR**
- **UCCA**
- **UDS...**

# PropBank

## Représentation sémantique commune pour les différentes réalisations syntaxiques

$[John]_{ARG0} \text{ broke } [the\ window]_{ARG1}$   
 $[The\ window]_{ARG1} \text{ broke}$

## Peu d'arguments possibles.

ARG0	agent	ARG3	starting point, benefactive, attribute
ARG1	patient	ARG4	ending point
ARG2	instrument, benefactive, attribute	ARGM	modifier

## Annotation en deux étapes.

- **Prédicats**
- **Arguments**

*The man joined the club with his friend.*

ARG0: The man

REL: joined

ARG1: the club

ARGM-COM: with his friend

*John and Mary met at a cocktail party.*

ARG0: John and Mary

REL: met

ARGM-LOC: at a cocktail party



# FrameNet

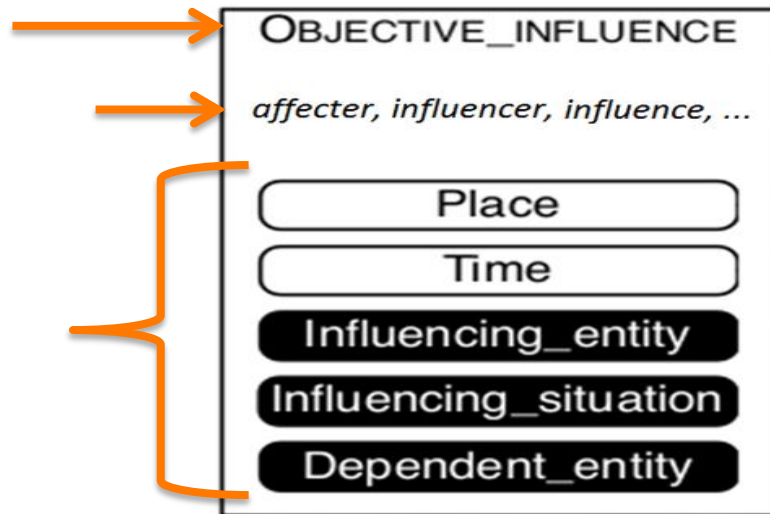
**Frame = situation prototypique d'un évènement,  
une relation, un état ou une entité**

**Evocable via des unités lexicales (déclencheurs)**

**Un cadre sémantique est caractérisé par des  
rôles sémantiques appelés Frame Elements.**

**Les cadres peuvent être liés entre eux par divers  
types de relations (*inheritance, using, ...*)**

**Rôles sémantiques indépendants de la syntaxe**



# FrameNet

## Politique de non tolérance à l'ambiguïté

[John]<sub>New\_member</sub> replaced [me]<sub>Old\_member</sub> != [John]<sub>Agent</sub> replaced [the telephone]<sub>Theme</sub>

[John]<sub>Speaker</sub> told [me]<sub>Listener</sub> [stuff]<sub>Message</sub> != [The book]<sub>Theme</sub> says [it]<sub>Message</sub>

## Beaucoup plus d'arguments possibles

### Annotation en deux étapes.

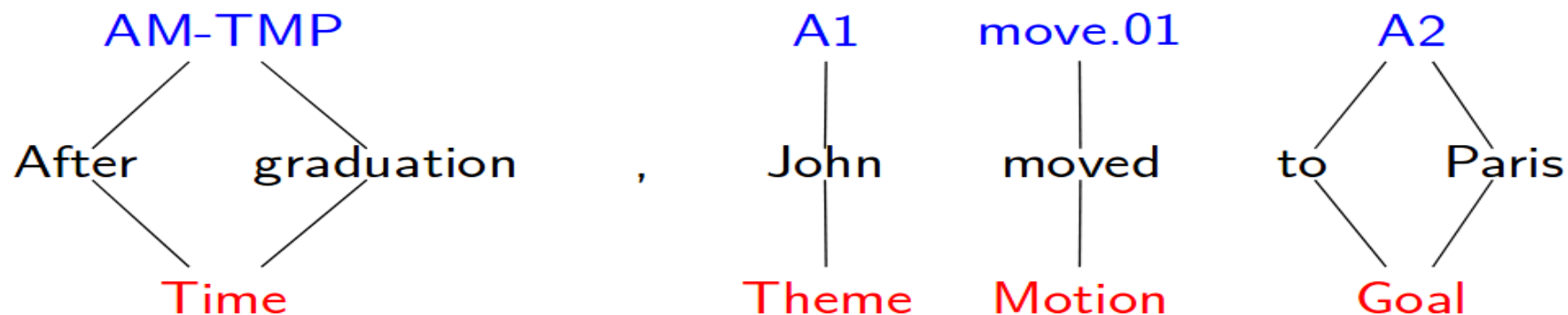
- **Prédicat (Unité Lexicale) → choix de la Frame**
- **Arguments → choix des FE (dépend de la Frame)**

### Exemples :

<https://framenet.icsi.berkeley.edu/fndrupal/frameIndex>

# PropBank vs FrameNet

PropBank



FrameNet

# AMR

## Propbank dans un graphe

Un peu plus de détail sur les coréférences, la polarité et les types de arguments.

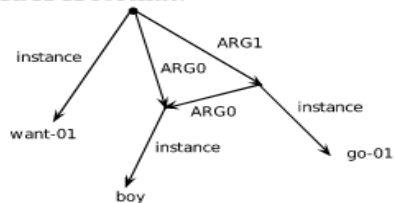
### LOGIC format:

```
∃ w, b, g:  
instance(w, want-01) ∧ instance(g, go-01) ∧  
instance(b, boy) ∧ arg0(w, b) ∧  
arg1(w, g) ∧ arg0(g, b)
```

### AMR format (based on PENMAN):

```
(w / want-01  
 :arg0 (b / boy)  
 :arg1 (g / go-01  
       :arg0 b))
```

### GRAPH format:



```
(t / think-01  
 :arg0 (b / boy)  
 :arg1 (w / win-01  
       :arg0 (t / team)  
       :polarity -))
```

The boy doesn't think the team will win.  
The boy thinks the team won't win.

(a / and

**:op1** (a2 / apple)

**:op2** (b / banana))

*Apples and bananas*

# UCCA

Modèle à base de graphes

Pas de précision sur les arguments

Plus proche de la syntaxe

Utilisé pour la génération  
et la traduction

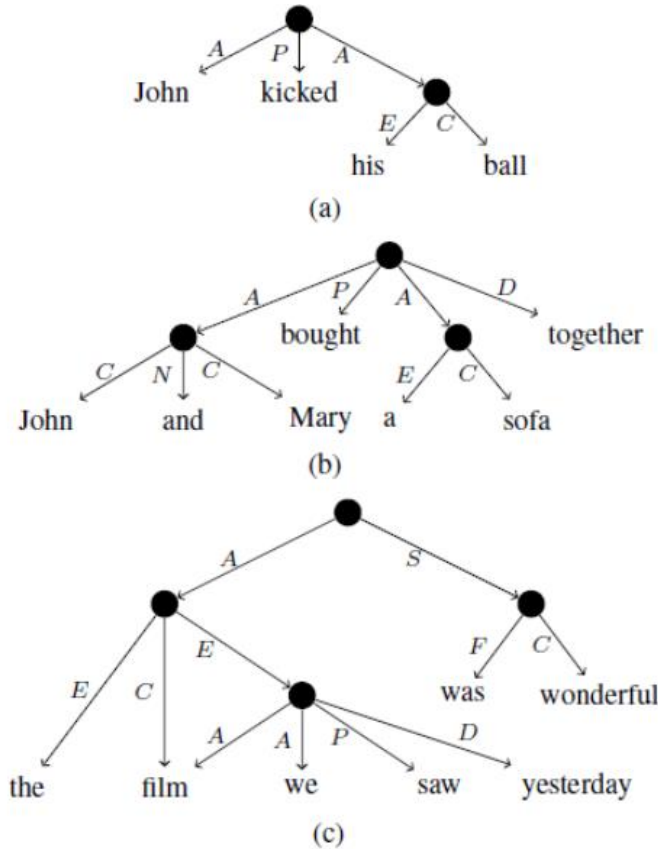


Abb.	Category
P	Process
S	State
A	Participant
D	Adverbial
C	Center
E	Elaborator
N	Connector
R	Relator
H	Parallel Scene
L	Linker
G	Ground
F	Function

# Prédiction Structuré

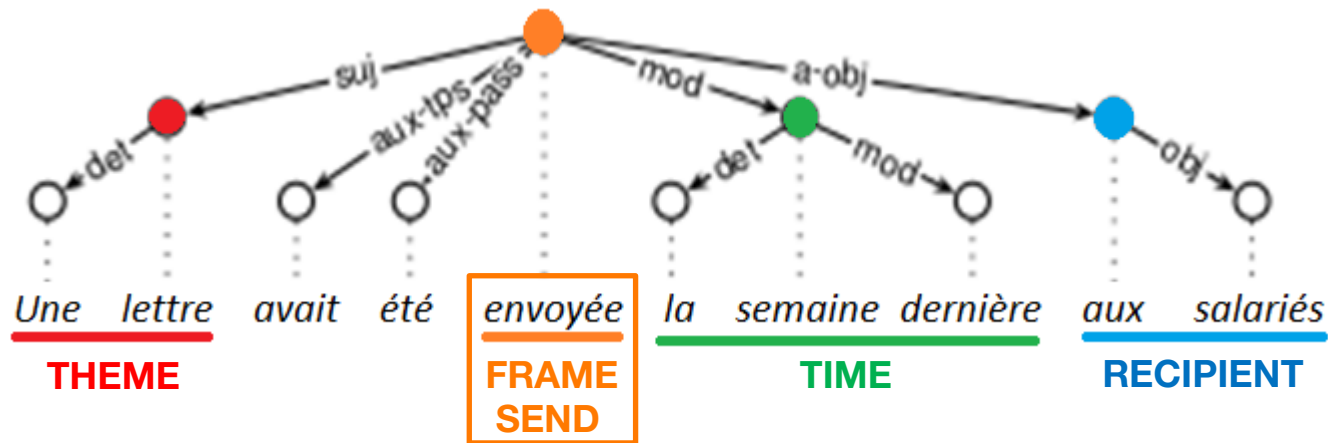
## Approches pour traiter un problème structuré

- Algorithmes à base de Transition → Analyse en dépendances
- Systèmes en cascade → Analyse en Frames
- Techniques de décodage optimal → Analyse PropBank
- Beam Search → Analyse en dépendances
- Integer Linear Programming → Analyse en Frames
- Multitask Learning → Analyse sémantique
- CRF → Analyse PropBank
  
- Neural Graph Learning

Les choix techniques vont dépendre du type de structure

# Prédiction Structuré → Analyse Sémantique par Etapes

- La mode jusqu'en 2016
- Systèmes: SEMAFOR 2016, Michalon ASFALDA 2016
- Besoin d'une analyse syntaxique au préalable
- Approche en 4 étapes
  - Détection de Cible → Lemmatiseur + Lexique
  - Sélection de Cadre → Classifieur linear
  - Détection des Rôles → Classifieur sur l'arbre syntaxique, ILP, optimisation sous contraintes
  - Sélection des Rôles → Classifieur sur l'arbre syntaxique, ILP, optimisation sous contraintes



## Prédiction Structuré → Analyse Sémantique par Transitions (SLING 2017)

Many	people	now	claim	to	have	predicted	Black	Monday
------	--------	-----	-------	----	------	-----------	-------	--------



4

3

2

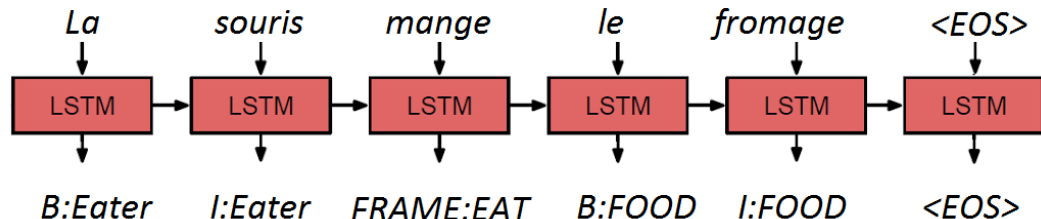
1

0

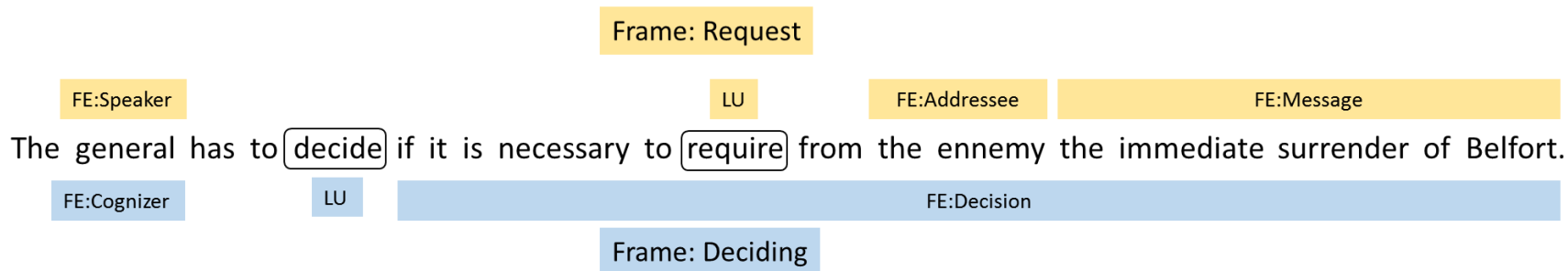


# Prédiction Structuré → Analyse Sémantique avec décodage optimal

- Etiquetage de Séquences mot à mot, à l'aide des CRF ou LSTM



- Problème des contraintes BIO et Frame/FE
- Problème Multi-label
- Problème du Linking



# Décodage Viterbi

Nous considérons les probabilités de chaque label

Mais aussi les probabilités de transition

$$V_{1,k} = P(y_1 | k) \cdot \pi_k$$

$$V_{t,k} = \max_{x \in S} (P(y_t | k) \cdot a_{x,k} \cdot V_{t-1,x})$$

# Beam Search

Faire un décodage par étapes

Garder seulement les K options les plus plausibles

Exemple: Analyse syntaxique par transitions

# Merci