

Structures de dépendances

Phrase

- Une phrase est une séquence de mots :

$$S = w_0 w_1 \dots w_n$$

- On considère que la décomposition en mots de la chaîne de caractères qui constitue la phrase est faite et ne varie pas.
 - l'université \rightarrow |l' | université|
 - aujourd'hui \rightarrow | aujourd'hui |
- le mot w_0 est un mot artificiel ajouté au début de la phrase pour des raisons techniques qu'on comprendra plus tard.

Graphe de dépendance

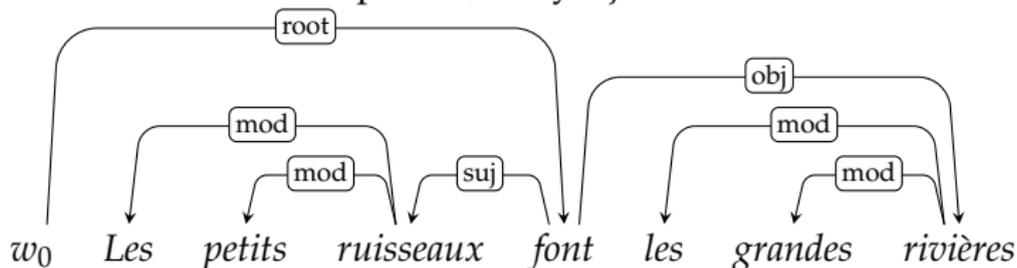
- On note $\mathcal{L} = \{l_1 \dots l_m\}$ l'ensemble des étiquettes des relations syntaxiques qui peuvent lier les mots d'une phrase (sujet, objet, attribut ...).
- Un **graphe de dépendance** $G = (V, A)$ est un graphe orienté étiqueté, constitué d'un ensemble de sommets V et d'un ensemble d'arcs A .
- tel que pour une phrase $S = w_0 w_1 \dots w_n$ et un ensemble d'étiquettes \mathcal{L} :
 - $V \in \{w_0, \dots, w_n\}$
 - $A \subseteq V \times \mathcal{L} \times V$
 - si $(w_i, l, w_j) \in A$ alors $(w_i, l', w_j) \notin A, \forall l' \neq l$
(en d'autres termes, il ne peut y avoir plus d'un arc entre deux sommets)
- l'ensemble des nœuds V correspondent aux mots d'une phrase
- l'ensemble A représente les arcs de dépendance étiquetés.

Arbre de dépendance

- Etant donné une phrase $S = w_0w_1 \dots w_n$, on note $V_S = \{w_0, \dots, w_n\}$ l'ensemble des occurrences des mots de la phrase.
- Un **arbre de dépendance** pour S est un arbre orienté $G = (V, A)$ dont la racine est w_0 et tel que $V = V_S$.
- On note G_S l'ensemble de tous les arbres de dépendance pour S .

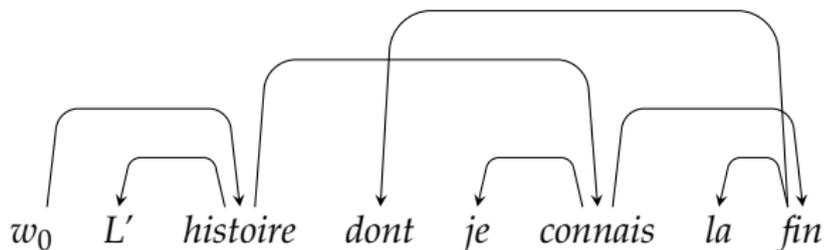
Projectivité

- La **projectivité** est une propriété de certains arbres de dépendance.
- Elle se traduit par le fait que lorsqu'on trace les dépendances au dessus des mots de la phrase, il n'y a jamais de croisement.



Projectivité (2)

- La grande majorité des phrases françaises sont projectives, mais il existe certains cas de non projectivité.



- Le degré de non projectivité varie d'une langue à l'autre
- Il est élevé pour le tchèque, le néerlandais ou le turc, par exemple.
- Il dépend aussi des choix linguistiques faits pour la définition des dépendances.

Projectivité (3)

Notations

- Etant donné une phrase S et un arbre de dépendance (V, A)
- on note $w_i \rightarrow w_j$ s'il existe, dans l'arbre (V, A) , une dépendance ayant w_i pour gouverneur et w_j pour dépendant.
en d'autres termes $w_i \rightarrow w_j \Leftrightarrow (w_i, l, w_j) \in A$ avec $l \in \mathcal{L}$
- $w_i \xrightarrow{*} w_j$ représente la clôture réflexive transitive de la relation de dépendance dans un arbre (V, A) .
- $w_i \xrightarrow{*} w_j$ si et seulement si $i = j$ (réflexivité) ou
 $w_i \xrightarrow{*} w_{i'}$ et $w_{i'} \rightarrow w_j$

Projectivité (4)

- un arc $(w_i, l, w_j) \in A$ dans un arbre de dépendance $G = (V, A)$ est projectif si et seulement si $w_i \xrightarrow{*} w_k$
 $\forall k \ i < k < j$ si $i < j$
 $\forall k \ j < k < i$ si $j < i$
- En d'autres termes, une dépendance $w_i \rightarrow w_j$ est projective, s'il existe un chemin allant de w_i à tous les mots situés entre les deux mots de la dépendance.
- un arbre de dépendance est projectif si toutes ses dépendances sont projectives.
- un arbre de dépendance n'est pas projectif si au moins une de ses dépendances n'est pas projective.

Projectivité (5)

- la projectivité diminue drastiquement le nombre d'arbres pouvant être associés à une phrase S (G_S) et donc l'espace de recherche d'un analyseur en dépendances.

nb nœuds	3	4	5	6	7	8	9	10
arbres proj	7	30	143	728	3876	21318	120001	690690
arbres dep	9	64	628	7776	117649	2×10^6	43×10^6	10^9