

# CYK Probabiliste

Arthur Lapraye & Korantin Lévêque & Mélanie Viegas

Paris VII

29 juin 2016

- Les CFG : un quadruplet  $(\Sigma, V, S, P)$

- Les CFG : un quadruplet  $(\Sigma, V, S, P)$
- Les CFG pondérées : ajout d'une fonction  $f : p \mapsto \alpha, p \in P, \alpha \in \mathbb{R}$

- Les CFG : un quadruplet  $(\Sigma, V, S, P)$
- Les CFG pondérées : ajout d'une fonction  $f : p \mapsto \alpha, p \in P, \alpha \in \mathbb{R}$
- Les CFG probabilistes : toutes les réécritures d'un non-terminal doivent sommer à 1.

$$\forall X \in V \sum_{X \rightarrow \alpha \in R} p(X \rightarrow \alpha) = 1$$

- Les CFG : un quadruplet  $(\Sigma, V, S, P)$
- Les CFG pondérées : ajout d'une fonction  $f : p \mapsto \alpha, p \in P, \alpha \in \mathbb{R}$
- Les CFG probabilistes : toutes les réécritures d'un non-terminal doivent sommer à 1.  
$$\forall X \in V \sum_{X \rightarrow \alpha \in R} p(X \rightarrow \alpha) = 1$$
- Les CFG probabilistes représentent un modèle de prédiction déduit à partir du corpus dont elles sont extraites.

# L'Algorithme CYK

- Un algorithme de parsing ascendant

# L'Algorithme CYK

- Un algorithme de parsing ascendant
- Complexité  $\mathcal{O}(|G|n^3)$

# L'Algorithme CYK

- Un algorithme de parsing ascendant
- Complexité  $\mathcal{O}(|G|n^3)$
- Parsing tabulaire



# L'Algorithme CYK

- Un algorithme de parsing ascendant
- Complexité  $\mathcal{O}(|G|n^3)$
- Parsing tabulaire
- Extension aux grammaire hors-contexte probabilistes (PCFG)

# L'Algorithme CYK

# Transformer la grammaire en CNF



# Transformer la grammaire en CNF



# Transformer la grammaire en CNF



# Transformer la grammaire en CNF



# Le corpus Sequoia

- Un corpus diversifié



# Notre implémentation du CYK



# Evaluation

# References



Brian Roark, Richard Sproat.

*Computational Approaches to Morphology and Syntax.*  
Oxford University Press, 2007.



Mariana Romanyshyn, Vsevolod Dyomkin.

*The Dirty Little Secret of Constituency Parser Evaluation*, 2014.  
<http://tech.grammarly.com/blog/posts/The-Dirty-Little-Secret-of-Constituency-Parser-Evaluation.html>



Martin Lange, Hans Leiss

« To CNF or not to CNF : An Efficient Yet Presentable Version of the CYK Algorithm », 2009

*Informatica Didactica* N° 8

("A Procedure for Quantitatively Comparing the Syntactic Coverage of English Grammars", E. Black, S. Abney et al. 1991, DARPA Speech and Natural Language Workshop, pp.306-311. Morgan Kaufmann )