

## Chapitre 3 : Etiquetage

- L'étiquetage est le fait d'opposer une étiquette à chaque mot d'une phrase
- Une phrase est donc l'entrée de l'étiquetage et est techniquement une suite de mots commençant par une majuscule et se terminant par un point.

- On souhaite trouver la catégorie grammaticale (POS-tagging - POS: Part of Speech)

ex: L'astrophysicien Stephen Hawking est mort à 76 ans.

DET NC NP NP V ADJ PRÉP NUM NC

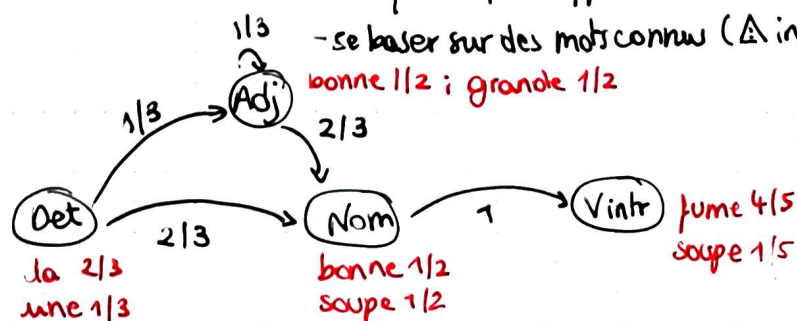
- On peut le faire aussi sous la forme d'entités nommées pour mettre du sens à la phrase

B-fore → I-fore ; O

- Problème étiquetage : associer chaque mot à une étiquette car choix des étiquettes difficile, il faut un manuel d'annotation (ex: Paris)

- ambiguïté de l'étiquette donc il est besoin de plus de données étiquetées pour apprendre

- se baser sur des mots connus (Δ inconnus), cas des mots peu fréquents...



proba unigram  $P(\text{grande} | \text{ADJ}) = \frac{\text{count}(\text{grande} | \text{adj})}{\text{count}(\text{adj})}$

proba trigram  $\rightarrow P(\text{ADJ} | \text{ADJ}, X) = \frac{\text{count}(\text{ADJ} | \text{ADJ}, X)}{\text{count}(\text{ADJ} | \text{ADJ})}$

- Entité nommée est une expression linguistique référentiel. Elle est souvent associée aux NP / NC. Une expression nominale décrivant un individu, un objet déterminé de manière unique

- Les objets considérés :
  - personnes humaines, réelles ou fictives, contemporaines ou historiques
  - lieux, entités localisées géographiquement toponyme anthroponyme
  - organisations: sociétés, institutions, gouvernements... ergonyme
  - dates, quantités ...

À regarder: [cs.columbia.edu/~mcollins/hmms-spring2013.pdf](http://cs.columbia.edu/~mcollins/hmms-spring2013.pdf)

NLP - 05. Tagging Problems and HMM (Youtube)