Compte rendu TP TAL

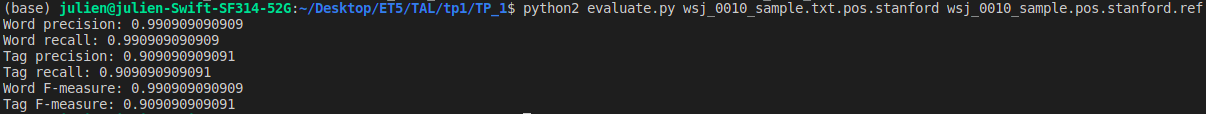
# TP 1 - Analyse linguistique avec la plateforme Stanford CoreNLP

## I. Installation et évaluation de l’outil de désambiguïsation morphosyntaxique de l’université de Stanford

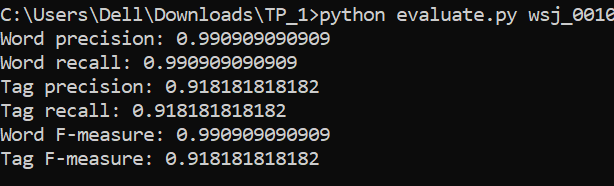
### 2. Evaluation

c

Evaluation :



e.



f. A partir de l’analyse des deux documents et du résultat de l’évaluation, on peut faire les conclusions suivantes:

* On observe une différence d’étiquetage morpho-syntaxique de 8%, et donc le Stanford donne des résultats assez similaires à la référence.
* Cette différence n’est pas forcément négative comme par exemple pour le cas “more” où on a un meilleur étiquetage pour le Stanford (ref:more\_ADJ, Stanford: more\_ADV).
* On observe presque aucune différence de segmentation entre les deux documents. Cela montre que l’outil de tokenisation de Stanford est très efficace.

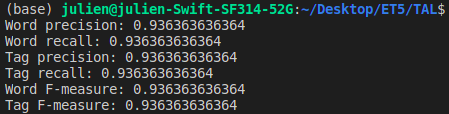
## II. Installation et utilisation de l’outil de reconnaissance d’entités nommées de l’université de Stanford

### 2. Extraction d’entités nommées

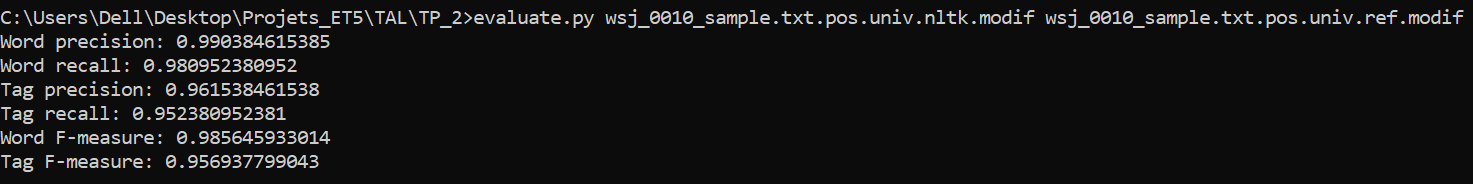
# TP 2 - Analyse linguistique avec le framework NLTK

### 1. Evaluation de l’analyse morpho-syntaxique de la plateforme NLTK

2. OK



3.



Conclusion:

* On observe une légère différence dans l'étiquetage d’environ 1%, ce qui montre que l’outil nltk est assez fiable.
* Le nltk en général, donne des résultats qui s'approchent beaucoup de la référence.

### 3. Utilisation de la plateforme NLTK pour l’extraction d’entités nommées

1. Suite à l’extraction des entitées nommées nous obtenons des arbres comme le suivant (plus facilement lisible qu’un fichier txt) :

