

# **Rapport de projet L3S5 : Linguistique de Corpus**

Ousseynou GUEYE

December 17, 2017

<b>1</b>	<b>Définition des objectifs</b>	<b>2</b>
1.1	Cahier de charge . . . . .	2
1.2	Matériels . . . . .	2
1.3	... . . . .	2
<b>2</b>	<b>Organisation des dossiers et du code</b>	<b>3</b>
2.1	Approche . . . . .	3
2.2	Organisation des dossiers . . . . .	4
2.3	Présentation des modules . . . . .	5
2.4	Processus . . . . .	6
<b>3</b>	<b>Difficultés rencontrées</b>	<b>7</b>
3.1	... . . . .	7
3.2	... . . . .	7
3.3	... . . . .	7
<b>4</b>	<b>Résultats</b>	<b>8</b>
<b>5</b>	<b>Ouverture et regrets</b>	<b>9</b>
<b>6</b>		<b>10</b>

# CHAPTER 1

---

## DÉFINITION DES OBJECTIFS

1.1 Cahier de charge

1.2 Matériels

1.3 ...

## CHAPTER 2

# ORGANISATION DES DOSSIERS ET DU CODE

### 2.1 Approche

## 2.2 Organisation des dossiers

### AVANT EXÉCUTION DU PROGRAMME

- **PROJECT\_ROOT/**
  - code\_source/
    - \* modules/
  - corpus\_litterature/
  - corpus\_professeur/
  - from\_outside\_treetagger/
  - morphalo/
  - rapport/

### APRÈS EXÉCUTION DU PROGRAMME

- **PROJECT\_ROOT/**
  - code\_source/
    - \* modules/
  - corpus\_litterature/
    - \* **statistiques/**
      - \* **xml/**
  - corpus\_professeur/
    - \* **statistiques/**
      - \* **xml/**
  - from\_outside\_treetagger/
  - morphalo/
  - rapport/
  - **resultats\_xml/**

Les noms nous paraissent assez explicites pour ne pas être tous détaillés. Néanmoins, certains dossiers nécessitent une petite explication.

- **from\_outside\_treetagger/** : au début de la programmation, l'accès aux éléments de morphalou était difficile. Pour ne pas perdre de temps, nous avons décidé d'utiliser aussi `tree_tagger`.

Celui-ci ayant besoin de fichiers particuliers pour fonctionner, nous les avons regroupés sous ce dossier.

- **corpus<sub>x</sub>/statistiques/** : Contient les statistiques de base, ainsi que les distributions de mots de chaque fichier pris séparément, puis du corpus en entier.
- **corpus<sub>x</sub>/xml/** : Contient l'arbre xml concernant le corpus.
- **resultats\_xml** : S'appuyant sur les deux dossiers précédemment cités, ce dossier contient toutes les données que nous avons pu recueillir.

## 2.3 Présentation des modules

La partie programmée est regroupée dans le dossier *code\_source*.

```
code_source/
• main.py
• settings.py
• modules/
  - big_process.py
  - ponctuation_texte.py
  - tagging.py
  - stats_0_distributions.py
  - stats_1_base.py
  - writing_in_files.py
  - others.py
```

Cette fois-ci, chaque fichier mérite que l'on vous explicite ce qu'il contient.

De même, nous aimerions préciser la logique de notre mouvement (flow). C'est en la suivant que j'ai pu, de manière systématique naviguer à travers le code<sup>1</sup>. Je pars toujours du niveau 0 vers le niveau 2.

### Niveau 0

- **main.py** : Contient l'interface. C'est le fichier qui englobe le tout.

• **settings.py** : Pour faciliter la gestion des liens, nous avons regroupé les liens vers nos dossiers dans ce fichier. Il ne contient aucune fonction.

### Niveau 1

- **modules/big\_process.py** : Ne contient que des fonctions composées d'autres fonctions des sous-modules. C'est le seul fichier du dossier *modules/* qu'appelle le *main*. On peut dire qu'il sert d'interface entre le *main.py* et les autres modules.

### Niveau 2

- **modules/ponctuation\_texte.py** :
- **modules/tagging.py** :
- **modules/stats\_0\_distributions.py** :
- **modules/stats\_1\_base.py** :
- **modules/writing\_in\_files.py** :
- **modules/others.py** :

---

<sup>1</sup>qui devenait de plus en plus gros.

## 2.4 Processus

- - [x] :
- - [x] :
- - [x] :
- - [x] :
- - [x] :
- - [x] :
- - [x] :
- - [x] :
- - [x] :
- - [x] :
- - [x] :
- - [x] :
- - [x] :
- - [x] :
- - [x] :
- - [x] :
- - [x] :

CHAPTER 3

DIFFICULTÉS RENCONTRÉES

3.1 ...

3.2 ...

3.3 ...



CHAPTER 4

RÉSULTATS

CHAPTER 5

OUVERTURE ET REGRETS

CHAPTER 6

--

## CONTENTS