# Rapport du projet de traduction automatique de la langue

## Sommaire

[Rapport du projet de traduction automatique de la langue 1](#_Toc34603319)

[Sommaire 1](#_Toc34603320)

[Objectif du projet 1](#_Toc34603321)

[Evaluation des plateformes d’analyse linguistique 2](#_Toc34603322)

[Contribution des membres du groupe 2](#_Toc34603323)

[Sullivan Honnet 2](#_Toc34603324)

[Jules Vittone-Burnel 2](#_Toc34603325)

## Objectif du projet

L’objectif affiché de ce projet était d’évaluer les résultats des analyses linguistiques réalisées par divers outils et d’en déduire lesquels étaient les plus efficaces dans quelle situation et éventuellement de proposer des pistes d’améliorations intéressantes pour ces outils.

L’objectif plus profond, abritait par ce projet, était de nous pousser à nous intéresser un domaine que nous rencontrions pour la première fois, celui du traitement automatique des langues et, par extension, la linguistique de manière générale.

## Evaluation des plateformes d’analyse linguistique

### Analyse morpho-syntaxique

### Reconnaissance d’entités nommées

## Contribution des membres du groupe

### Sullivan Honnet

Il eut la charge de l’utilisation des outils nltk et stanford, il s’est chargé de générer les fichiers à partir de ces outils.

### Jules Vittone-Burnel

Il s’est occupé de l’utilisation et de la génération du fichier lima et d’écrire le script principal. Ce script a pour rôle de formater tous les fichiers de départ pour obtenir des fichiers utilisables par la suite par le programme evaluate.py. L’une des grosses difficultés dans cette partie fut que chaque fichier suivait sa propre nomenclature et ses propres régles et qu’il n’existait pas d’uniformité entre le fichier de départ fourni et les fichiers obtenus par nltk, stanford et lima.

Le problème se présenta pour l’analyse morpho-syntaxique d’abord par le fichier source qui contenait de nombreuses lignes comportant plusieurs mots pour un seul tag alors que tous les autres programmes donnaient une équivalence un mot, un tag. Ce problème fut résolu en découpant les groupes de mots d’une ligne en plusieurs lignes contenant un mot et le tag associé auparavant au groupe de mots.

Une image contenant capture d’écran, intérieur, assis, objet

Description générée automatiquement avec une confiance faible

Figure 1 : Résultat de la reconnaissance d'entités nommées de nltk

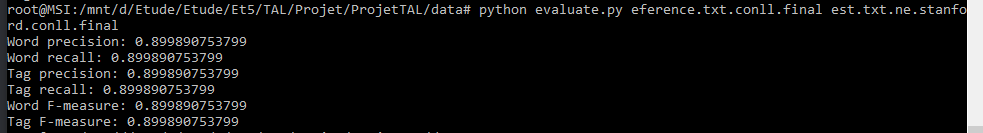


Figure 2 : Résultat de la reconnaissance d'entités nommées de stanford

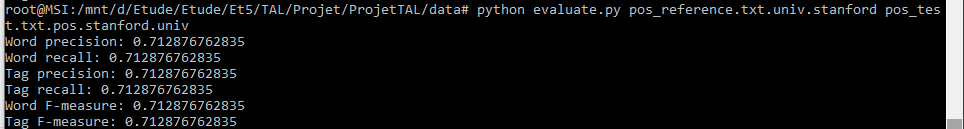


Figure 3 : Résultat de l'analyse morpho-syntaxique de stanford

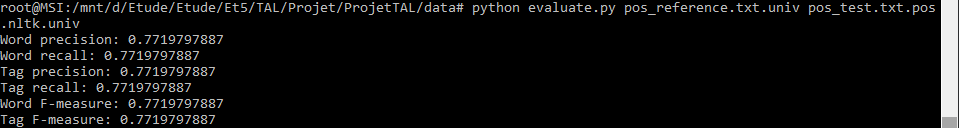


Figure 4 : Résultat de l'analyse morpho-syntaxique de nltk