

# Projets pour le TAL

## Liste des projets

L'ensemble des projets listés ci-dessous seront regroupés par la suite dans une seule application « Boite à Outils ». Ainsi, il faut utiliser la même bibliothèque d'interfaces graphiques pour tous les projets.

### 1. Translittérateur

L'outil doit permettre de convertir des caractères d'une langue naturelle vers un autre jeu de caractères. L'outil doit posséder les fonctionnalités suivantes :

- permettre à l'utilisateur de configurer le script source et cible ;
- permettre à l'utilisateur de modifier le jeu de caractères à utiliser ;
- assurer la translittération en temps réel ;
- permettre la conversion du contenu d'un fichier de type Word, RTF ou Txt ;
- assurer la conversion d'une liste de fichiers.

### 2. Concordancier

Ce projet consiste à la réalisation d'un outil qui affiche toutes les occurrences d'un mot ou d'une forme donnée avec son contexte d'apparition dans un texte ou un ensemble de textes. L'outil doit posséder des fonctionnalités permettant :

- des recherches simples et complexes (séquence de mots, expressions régulières) ;
- le choix du contexte ;
- l'affichage de la liste des fréquences des mots ;
- la liste des fichiers contenant un mot ;
- une exportation aisée des résultats sous format Txt.

### 3. Gestionnaire de corpus

L'objectif de ce projet est de réaliser un outil qui assure la gestion des fichiers d'un corpus an assurer les fonctionnalités suivantes :

- segmenter le texte en phrases et en mots ;
- présenter le résultat de la segmentation dans un fichier XML en marquant les parties : le nom du fichier, l'auteur et la dates de rédaction extrait d'un fichier Excel, ainsi que la phrase et le mot tout en précisant leur emplacement ;
- présenter les statistiques du corpus en terme de fréquence de mots et de caractères, la longueur des fichiers, longueur moyenne et maximale des mots du texte.