IUT de Nantes

Année 2015-2016

Modélisation Mathématique

Auteurs :

* Cédric Berland
* Nathan Maraval



# Introduction

Ce projet a été réalisé dans le cadre de notre formation en DUT Informatique, à l’IUT de Nantes. Il fût demandé durant le cours de Modélisation Mathématique.

Le but de ce travail est de pouvoir identifier, à partir de probabilités, le langage d’origine d’une phrase donnée, en ayant par avant utiliser des corpus d’apprentissage dans plusieurs langues.

Le projet se sépare en deux parties, une ayant pour principe de générer les probabilités des phrases, et l’autre cherchant les langages les plus probable, tout en gérant les langages d’origine inconue.

# Différents problèmes rencontrés

## Les langages inconnues

Durant ce projet, nous avons pu rencontrer quelques problèmes. Le principal de tous ces problèmes fût la gestion des langues inconnues. Dans un modèle de langage Naïf, toutes langues inconnues auront une probabilité de un, mais ce genre de modèle n’est pas très fiable – en effet, si un mot est inconnu, c’est toute la phrase qui perdra sa probabilité.

Nous devons donc utiliser un modèle dit de Laplace, qui affectera à chaque tuple de la phrase un poids, ce qui empêchera d’avoir une probabilité de 0 pour un simple mot inconnue dans une phrase.

Cependant, si une phrase est inconnue, cela veut dire que toutes les langues dans le modèle que nous avons créé auront une probabilité non-nul, dans un intervalle pouvant varier.