

ZO

ZO

ZO

ZO

ZO

SILC (Système d'Identification de la Langue et du Codage)

TextCat

Xerox Mltt Language Guesser

Lextek Language Identifier

Architecture du système

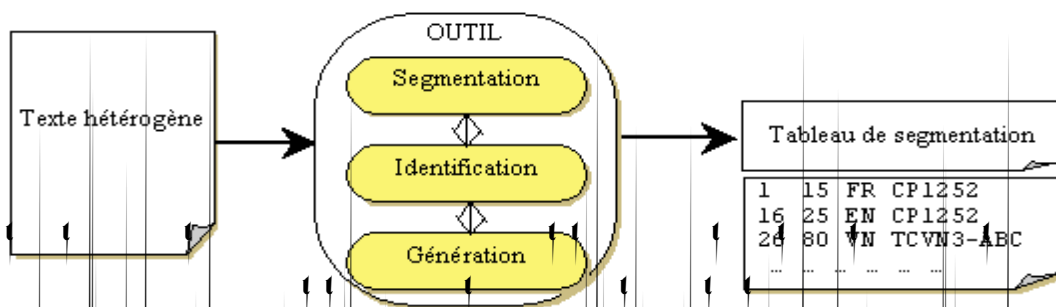


Figure 1. Outil d'analyse d'un texte hétérogène en zones homogènes

Segmentation

Identification de la langue et du codage d'une zone homogène

Architecture générale du système d'identification de la langue et du codage

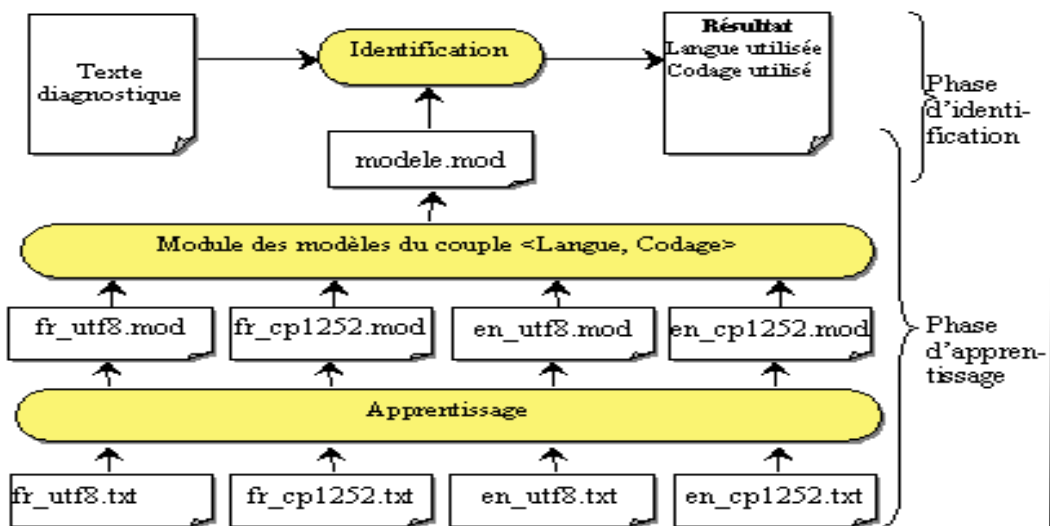
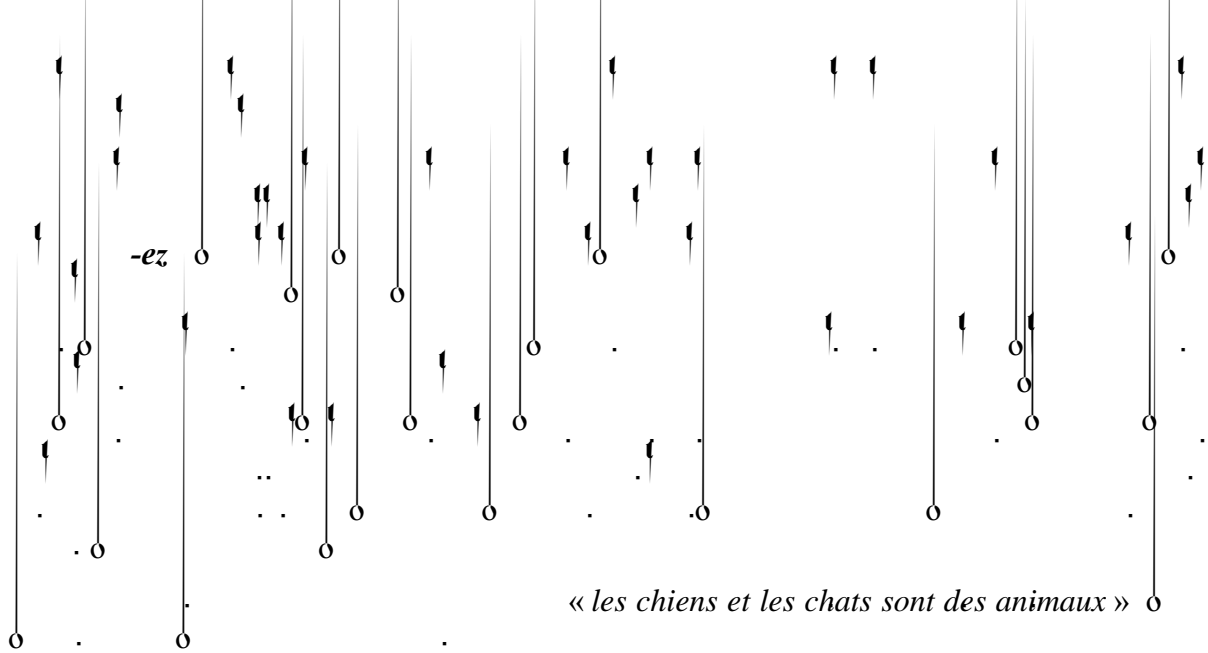


Figure 2. Architecture générale d'un identifieur de langue et de codage

Méthode n -grammes

n caractères

-ck



*Tableau 1.
Modèles de langage avec leurs séquences et leurs fréquences*



Résultat obtenu

« les chiens et les chats sont des animaux »

地定空虛混沌。湖國黑暗。禪的冥運行在水面上

...		

Tableau 2. Tableau de scores de résultat diagnostique

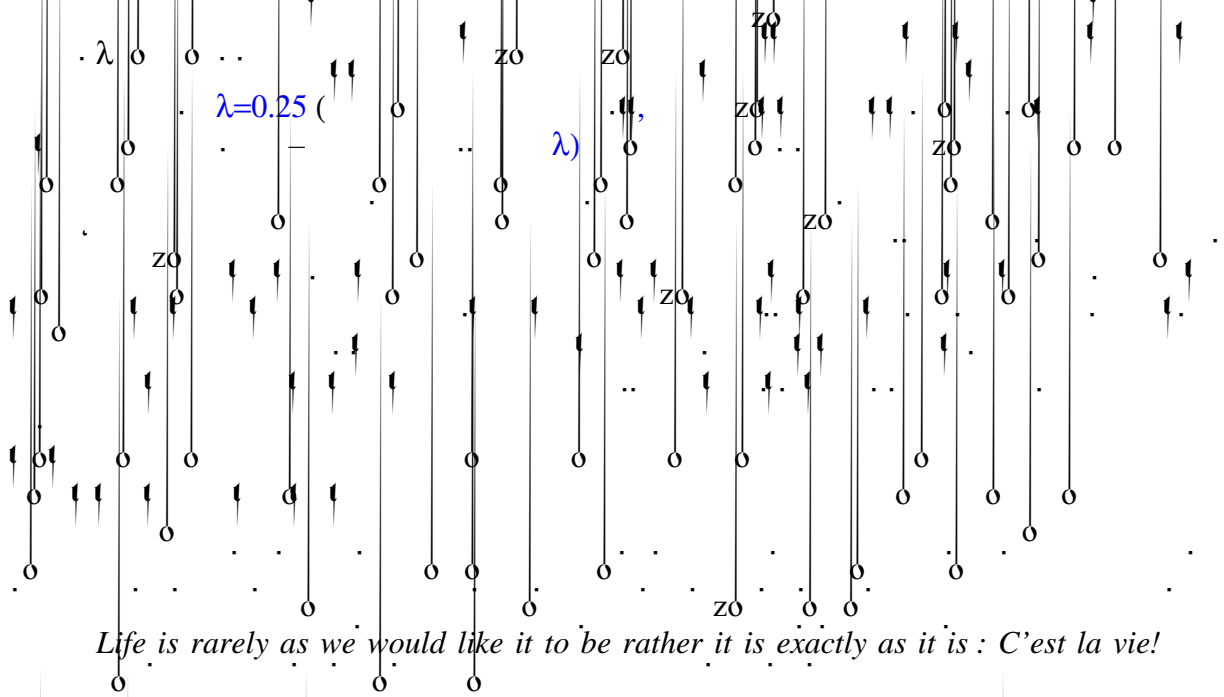
Évaluation des zones

zo

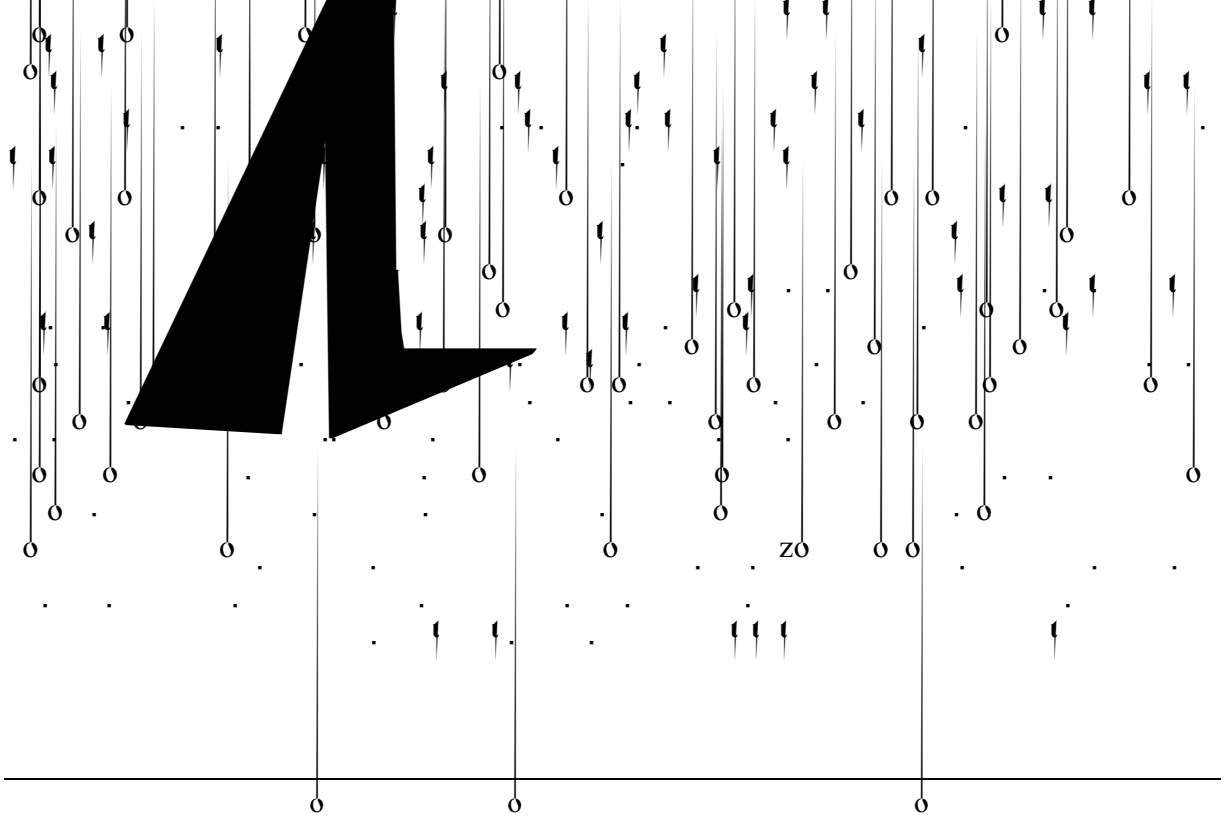
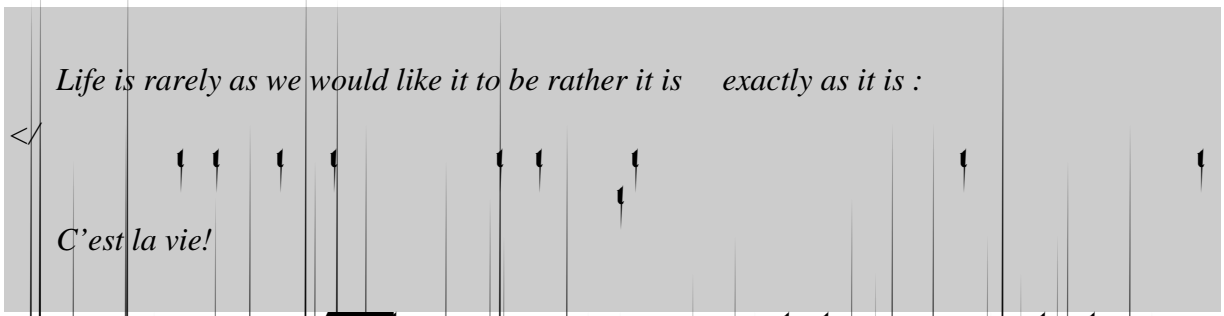
$$q = \frac{l_1 - t}{l}$$

l_1

l_2



Life is rarely as we would like it to be rather it is exactly as it is : C'est la vie!



Language at Crossroads : Proceedings of the 29th Annual Conference of the American Translators Association

Reconstruction et Utilisation de SILC

Évaluation de SILC

Symposium On Document Analysis and Information Retrieval

Symposium On Document Analysis and Information Retrieval

Méthode pour l'analyse automatique de structures formelles sur documents multilingues.

Actes des JADT 1995.

Foundations of Statistical Natural Language Processing

The QUE Language and Encoding Identification Package

Conférence MAJECSTIC'03

