Méthodologie du travail de correction des noms d’auteurs erronées

**Corrections des auteurs**

* Cible aux **auteurs problématiques :**

A partir du tableur auteur\_tri.xlsx, nous concentrons sur deux types d’information sur la chance d’être auteur : **NULL** (dont Anonymous) et **faible** (ceux qui ont l’activité de travail et aussi date de décès au 19e, 20e siècle

* Nous avons au final une liste de 43 noms d’auteurs avec chance d’être erronés plus élevée
* À priori, des auteurs avec une grande quantité de textes sont plus attentionnées (avec plus de chance d’être erronée), également sur les groupes d’auteurs (avec la possibilité comme co-auteurs ou groupe d’éditeurs)
* La plupart des textes d’un fichier dédié d’un auteur du projet vient d’un ouvrage précis sur DBNL, dans le cas d’un éditeur ou groupe d’éditeurs, la collection contient souvent des chansons traditionnelles, folkloriques dont auteurs anonymes.
* Difficulté en reconnaissance entre un éditeur et un auteur : Un auteur d’hypothèse pourrait en même temps un éditeur, musicien, poète et écrivain, pour cette raison, il faut trouver son vrai rôle dans l’ouvrage précis
* Des collections de chansons dites folkloriques ont deux possibilités soit écrite par un auteur connu, soit collectionnée par un éditeur (un musicien, un poète ect.)
* Vérifier manuellement l’information de profil d’auteur par utilisation du portail **DBNL :** collecter des donnés biographiques et aussi le nom du vrai auteur de chaque parole de chanson
* Listes des noms erronés
* Corriger manuellement des noms d’auteurs erronés par défaut
* Corriger automatiquement des noms d’auteurs erronées en masse (cas des collections des chansons folkloriques dont auteurs anonyme)

Choix de correction :

* Le cas des auteurs « Anonymous » : nous avons un choix problématique de vérifier 100 premiers textes (fichier numéro 1 à numéro 155) sur 8959 textes: 28/100 sont erronées : 28%
* Corriger seulement si l’auteur est déjà dans la liste à défaut, dans le cas d’un nouvel auteur hors de la liste, nous mettons « Anonymous »
* Nombre de nom d’auteurs erronés / total : 1024/22214 fichiers : 4,6%