Corpus Malherbe : corpus de textes versifiés du XVIIe au début du XXe

Richard Renault

corpus malherbe

Corpus de textes versifiés du XVII^e au début du XX^e

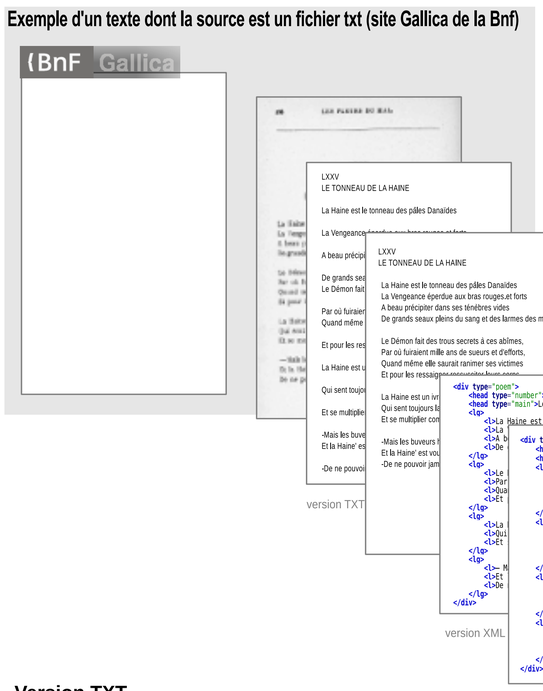
projet ANAMÈTRE (traitement automatique de textes versifiés)

responsables du projet : Éliane Delente & Richard Renault

responsable du corpus : Richard Renault

PRÉPARATION DU CORPUS

Exemple d'un texte dont la source est un fichier txt (site Gallica de la Bnf)



Préparation du texte :
de la version image à la version XML-TEI

Version TXT

- Préformatage du texte au moyen de tabulations et d'interlignes
- Vérification orthographique (lettres ligaturées, majuscules accentuées...)
- Normalisation de la ponctuation (espace devant un signe de ponctuation forte, tirets cadratins, guillemets...
- Introduction des balises XML délimitant les différentes parties du texte au moyen d'expressions régulières

Version XML

- Formatage du texte (italiques, exposants, filets...)
- Rédaction de l'en-tête éditorial
- Validation du corpus au moyen d'un schéma XSD*

*** L'élaboration du schéma de validation ainsi que la normalisation de l'ensemble du corpus a bénéficié d'un financement du consortium CORLI-ORTOLANG en 2016.**

Post-traitement

- Après une première passe d'analyse métrique automatique des vers, insertion automatique des retraits de vers en fonction de la longueur métrique du vers et de la longueur métrique maximale du poème.
- Insertion d'éléments XML-TEI de formatage métrique :
  - phonétisation de mots incomplets ou de syllabe sans voyelle (*Brrr* !) ...
  - vers sans rime, vers incomplets, vers inanalysables…
- Insertion d'éléments XML-TEI de correction lorsque l'édition imprimée comporte des erreurs.

PAGE HTML DU RECUEIL*

* http://www.crisco.unicaen.fr/verlaine/

PAGE HTML D'UN POÈME ANALYSÉ APRÈS TRAITEMENT AUTOMATIQUE

Charles Baudelaire
LES FLEURS DU MAL
1857-1861

SPLEEN ET IDÉAL

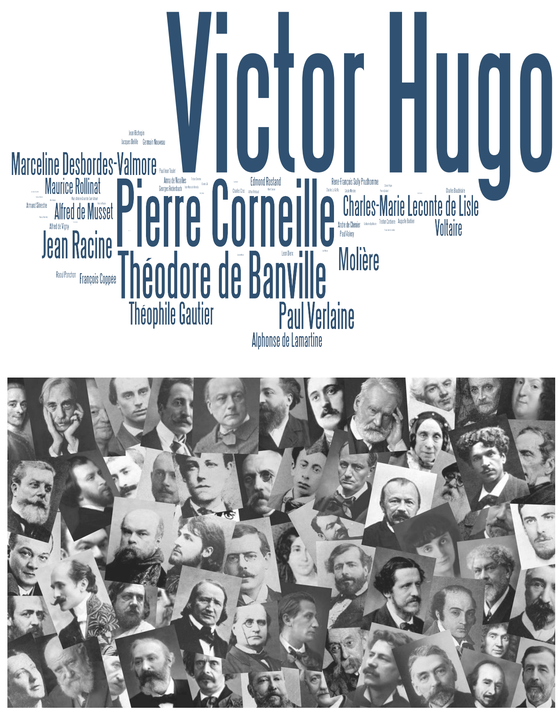
LXXVI

Le Tonneau de la haine

	La Haine est le tonneau des pâles Danaïdes ;	6+6	a
	La Vengeance éperdue aux bras rouges et forts	6+6	b
	A beau précipiter dans ses ténèbres vides	6+6	a
	De grands seaux pleins du sang et des larmes des morts,	6+6	b
5	Le Démon fait des trous secrets à ces abîmes,	6+6	c
	Par où fuiraient mille ans de sueurs et d'efforts,	6+6	d
	Quand même elle saurait ranimer ses victimes,	6+6	c
	Et pour les ressaigner ressusciter leurs corps.	6+6	d
	La Haine est un ivrogne au fond d'une taverne,	6+6	e
10	Qui sent toujours la soif naître de la liqueur	6+6	f
	Et se multiplier comme l'hydre de Lerne.	6+6	e
	— Mais les buveurs heureux connaissent leur vainqueur,	6+6	f
	Et la Haine est vouée à ce sort lamentable	6+6	g
	De ne pouvoir jamais s'endormir sous la table.	6+6	g

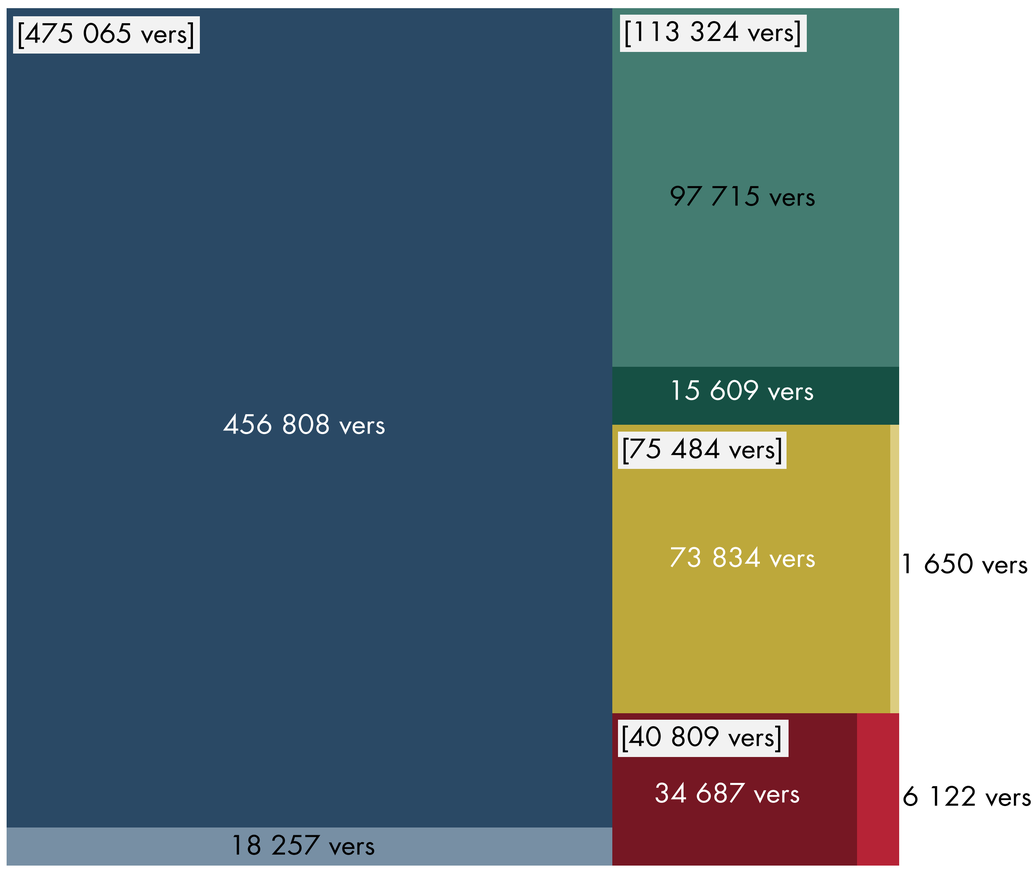
profil métrique : **6+6**

type : **sonnet**



RÉPARTITION DU CORPUS SELON LA PÉRIODE ET LE TYPE DE TEXTE

(proportionnellement au nombre de vers)

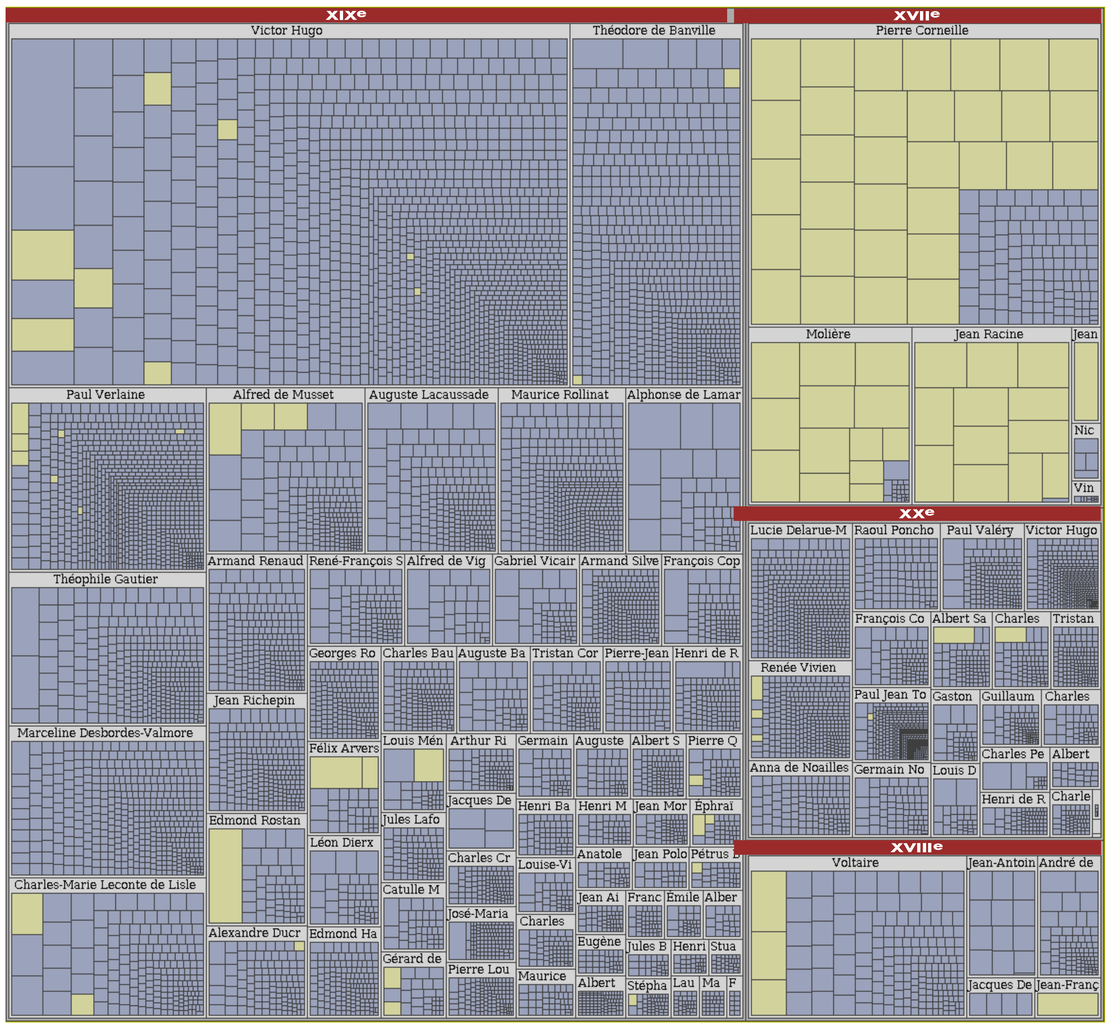


- 87 auteurs
- 322 textes
- 26 recueils de poésies
- 102 pièces de théâtre
- 12 670 poèmes
- 704 385 vers

(version 2.2, janvier 2018)

RÉPARTITION DU CORPUS SELON LA PÉRIODE ET LES AUTEURS

(proportionnellement au nombre de vers)



poème
pièce de théâtre

- 87 auteurs
- 322 textes
- 26 recueils de poésies
- 102 pièces de théâtre
- 12 670 poèmes
- 704 385 vers

(version 2.2, janvier 2018)

poster réalisé par Richard Renault

Centre de Recherches Inter-langues sur la Signification en COntexte

UNICAEN université de Caen Basse-Normandie