

Anatext, <http://phraseotext.univ-grenoble-alpes.fr/anaText>

On utilisera l'œuvre *Le Mariage inattendu de Chérubin* (Olympe de Gouges, 1786) :
https://fr.wikisource.org/wiki/Le_Mariage_inattendu_de_Ch%C3%A9rubin

Copiez-collez le texte dans un éditeur de texte. On prend toute l'œuvre, y compris la préface, mais pas l'interface de Wikisource.

Chez Cailleau, 1788 (p. 3-111).	
Le Mariage inattendu de Chérubin	Le Mariage inattendu de Chérubin ►

PRÉFACE.

Je suis femme & Auteur ; j'en ai toute l'activité. Mon premier mouvement est semblable à une tempête ; mais dès que l'explosion est faite, je reste dans un calme profond : tel est l'effet qu'éprouvent toutes les personnes vives & sensibles.

Mon *Mariage de Chérubin* est un enfant de la *Folle Journée*, qui naquit de l'enthousiasme général, c'est un de mes premiers Ouvrages, duquel je me promettois beaucoup de gloire, & encore plus de profit ; mais, hélas ! c'est bien le cas de dire :

Pauvres petits infortunés,
Vous êtes morts avant que d'être nés !

Lu à la Comédie Italienne, il y fut accueilli ; mais des considérations de Théâtre à Théâtre en ont empêché la représentation, je le présente aujourd'hui au Public, rempli de fautes, tel que doit l'être une production faite en vingt-quatre heures à laquelle je n'ai rien changé. Cependant des hommes de Lettres, ainsi que MM. les Comédiens, y ont

Dans l'éditeur, adaptez la graphie du texte pour qu'il soit analysable par un modèle de langue du français moderne : comment traitez-vous les caractères **f** et **&** ?

Consultez le texte source et le texte étiqueté.

Quels sont les lemmes les plus spécifiques du texte ?

... Par rapport à quoi sont-elles spécifiques ?

(indice : bien lire les petites notes sur la page d'accueil, et http://lexique.org/_documentation/Manuel_Lexique.3.2.pdf)

L'un des verbes les plus spécifiques est le verbe *avoyer*, à quoi correspond-il ?

Quels sont les lemmes les plus fréquents ?

Recherchez toutes les occurrences de l'adjectif *pauvre*. Combien d'occurrences ?
(si vous en trouvez 10 il y a un problème... on a dit les adjectifs)

Recherchez les concordances, les cooccurrences et les patterns pour le verbe *faire* suivi d'un adjectif. Lequel de ces outils vous semble le plus approprié pour trouver l'adjectif le plus souvent employé juste après le verbe *faire* ? Comment pourriez-vous « détourner » les autres pour obtenir aussi cette information ?

Quel est le mode verbal le plus utilisé dans ce texte ?

Frantext, <https://www.frantext.fr/>

Familiarisation avec le corpus

Quel **genre textuel**, **période** ?

Quelle **taille** ? Nombre de textes, nombre mots ? Quelles autres informations sont disponibles dans l'onglet Statistiques ?

Sous-corpus

Créer un sous-corpus avec les textes entre 1701 et 2000.

Mots en -isme

Quels sont les mots en *-isme* les plus fréquents entre 1701 et 2000 ?

Quel siècle comporte le plus de mots en *-isme* ?

Quels sont les mots en *-isme* les plus fréquents pour chaque siècle ? (faites une recherche par siècle)

S'agit-il de fréquences absolues ou bien de fréquences relatives ?

Voisinage de mots

Quels mots s'emploient le plus souvent dans la même phrase que *travail* ? Comparez avec *travaux*.

Quels mots s'emploient souvent exactement après *faire* ?

Parties du discours, bruit et silence

Recherchez le verbe *taire* (mot fléchi) au participe passé (partie du discours). Combien d'occurrences trouvez-vous ? Citez un biais qui cause du **bruit** si on n'utilise pas la catégorisation.

Quels substantifs utilise-t-on le plus souvent directement (sans article) après le verbe *faire* ? Avec un article ? Est-ce suffisant pour étudier les compléments d'objet préférés du verbe *faire*, ou bien y a-t-il du **silence** ?

Créez un corpus de 4 œuvres de Zola : *La Curée*, *Germinal*, *L'Œuvre* et *L'Argent*. Créez une liste de mots fléchis liés à la notion de dette : *dette*, *créance*, etc. Dans lequel de ces 4 romans ces mots sont-ils les plus fréquents ?

Quels sont les compléments d'objet direct préférés du verbe *faire* ? Citez un biais qui peut provoquer du **silence**.

Récapitulatif

	Type de texte	Structuration du corpus (oui/non)	Lemmes (oui/non)	Parties du discours (oui/non)
Anatext				
Frantext				
Google Ngrams				
Google Search				
Google Trends				
TXM				

Et encore plein d'autres !

- <http://explorationdecorpus.corpusecrits.huma-num.fr/>

Et plus généralement : le consortium CORLI

- <https://groupes.renater.fr/sympa/info/corli>
- <https://groupes.renater.fr/wiki/corli/index>