

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

On utilisera le corpus de démonstration (Boule de Suif).

Consultez le texte source, la sortie brute de TreeTagger, et le texte étiqueté.

Quels sont les noms/verbes/adjectifs/adverbes les plus fréquents ?

Quels sont les lemmes les plus spécifiques du texte ?

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

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

Recherchez toutes les occurrences de l'adjectif prussien.

Chercher les segments répétés contenant « sœur ». Chercher les segments répétés contenant « boule ». Y'a-t-il des erreurs ? D'où viennent-elles ?

SketchEngine, <https://www.sketchengine.eu/>

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

Familiarisation avec le corpus

Quel critère de sélection : **genre textuel, période** ?

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

Listes de mots

Créez la liste des mots en *-isme(s)*.

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

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	Lemmes	Parties du discours	Dépendances
Anatext					
Frantext					
Google Ngrams					
Google Search					
Google Trends					
ScienQuest					
SketchEngine					
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>