

Analyse du discours

Licence 2 : Lettres modernes

Séance 2 (22/02/23) : Le projet et le corpus

marine.delaborde@cyu.fr



CERGY PARIS
UNIVERSITÉ

Présentations

- <https://www.cyu.fr/marine-delaborde>
- À vous : <https://forms.gle/yKTBKuWMvQLRCWfn6>
 - Rempli ?

Organisation des séances

2 / 6

Séances	Points abordés
Séance 1 : 18/01/23 (salle 207)	Introduction
Séance 2 : 22/02/23 (salle 121)	Le projet, iTrameur et le corpus
Séance 3 : ??/??/23 (salle 121)	Ventilation et concordance
Séance 4 : ??/??/23 (salle 121)	Spécificités
Séance 5 : ??/??/23 (salle 121)	Cooccurrents et segments répétés
Séance 6 : ??/??/23 (salle 121)	Retour sur le projet

Évaluation et projet

- **Projet** : réalisation d'un dossier d'analyse du discours
 - **Constitution et exploitation de corpus** :
 - collecte de 10 articles de presse,
 - formatage du corpus : **le corpus fait partie du rendu (reproductibilité des expériences)**,
 - application d'une méthode qualitative et d'une méthode quantitative,
 - rédaction de 10 pages hors page de couverture et mise en forme du dossier
 - travail en groupe de 2 ou 3 personnes
 - **Dossier** :
 - 10 pages dactylographiées à interligne 1,5 et numérotées, précédées d'un page couverture (sujet+auteur+table des matières+date+formation)
 - Note 1 : Ni la page couverture (qui ne sera pas numérotée), ni les annexes ou la bibliographie ne seront comptées dans le nombre de 10 pages. La bibliographie est formée de références théoriques sur le sujet, d'ouvrages qui analysent le même événement.
 - Note 2 : Indiquez le corpus, c'est-à-dire la liste des articles de presse analysés dans les annexes.
- **Évaluation** : participation en cours et investissement (20%), qualité du dossier d'analyse d'un événement médiatique (80%).

Analyse textométrique

Objectifs du projet

- **Objectif : apporter des éléments de réponse à une question de recherche spécifique en s'appuyant sur les résultats de mesures textométriques appliquées à un corpus de 10 articles.**
 - Formuler une problématique en lien avec la thématique abordée dans les 10 articles.
 - Identifier des hypothèses qui seront confirmées ou infirmées par les données.
 - Si les mesures permettent d'infirmar une hypothèse, c'est quand même intéressant !
 - Ne pas modifier les hypothèses formulées au début : prise de recul pour comprendre les résultats.

Analyse textométrique

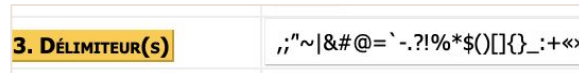
Méthodes générales de la textométrie

- Glossaire : <http://lexicometrica.univ-paris3.fr/livre/st94/STUCGLOS.pdf>
- Repérer des **régularités** / **anomalies** dans le comportement des unités d'un texte.
- Comparer la répartition des unités dans les différentes parties du corpus.

Analyse textométrique

Méthodes générales de la textométrie

- **Objectif** : explorer un corpus textuel
 - Repérer des régularités / anomalies dans le comportement des **unités** d'un texte
 - Quelles unités ?
 - Mot ? Qu'est-ce qu'un mot ?
 - Définir ce que l'on va compter
 - Racine
 - Lemme
 - Forme graphique
 - Segment
 - Cooccurrences
 - Dépend de l'analyse souhaitée et des outils utilisés
 - Segmentation : choix des délimiteurs dans iTrameur :



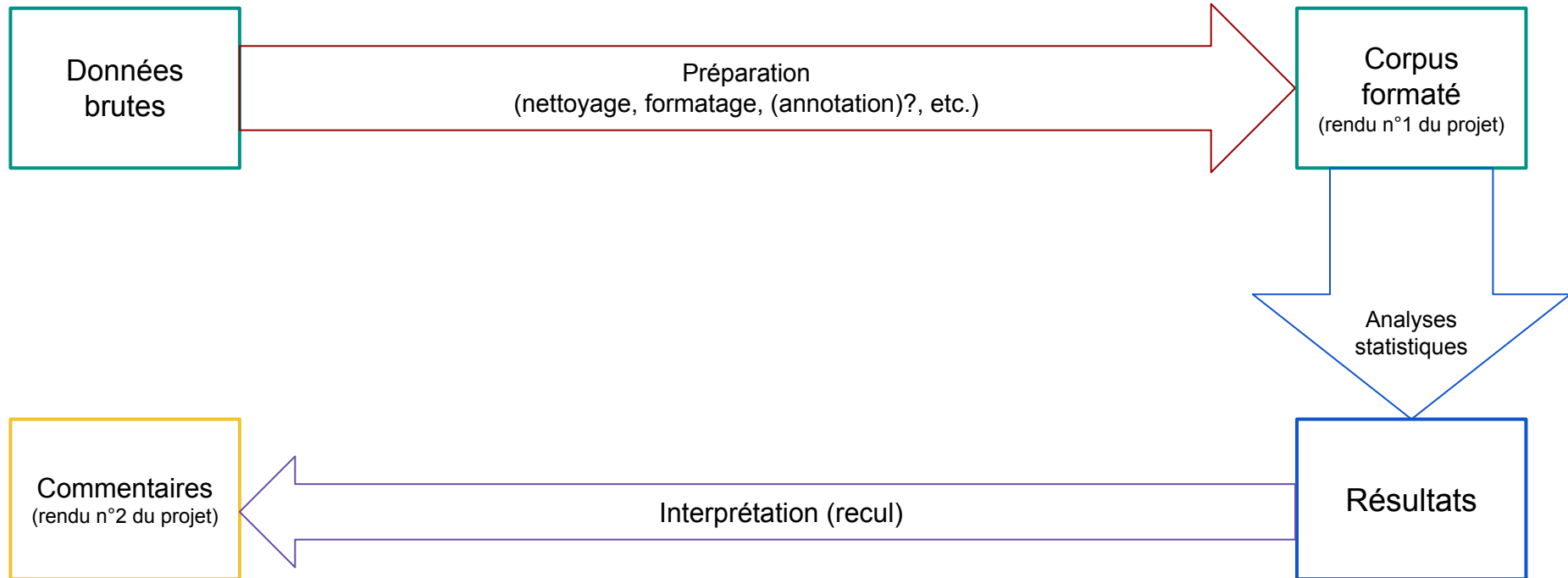
Analyse textométrique

Méthodes générales de la textométrie

- **Objectif** : explorer un corpus textuel
 - Comparer la répartition des unités dans les différentes **parties** du corpus
 - Découpage du texte en parties à l'aide de jalons textuels (balises)
 - Quel découpage (partition) ?
 - date,
 - auteur,
 - source,
 - genre,
 - courant,
 - ...
 - Dépend des données de départ et l'analyse souhaitée
 - Qu'est-ce qu'on veut comparer ?

Analyse textométrique

Chaîne de traitements



Analyse textométrique

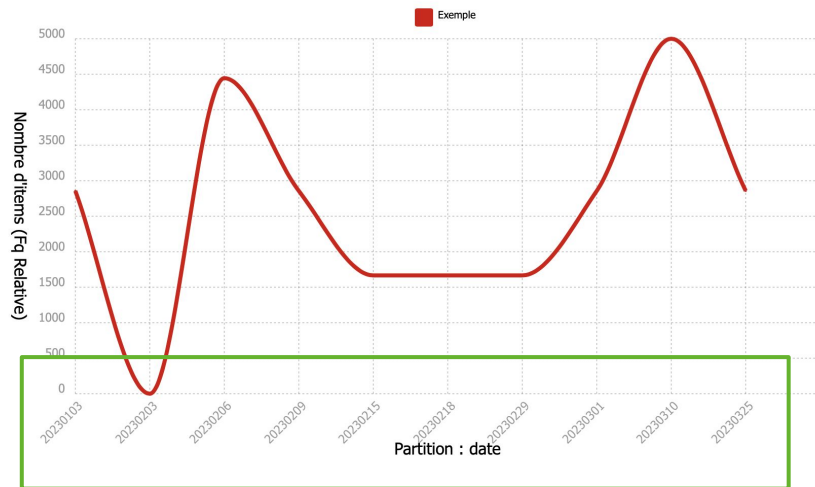
Le corpus formaté pour iTrameur

- 10 articles
- **Format texte brut** (exemple : corpus.txt) et **UTF8 !**
 - Modifications à l'aide d'un éditeur de texte brut comme **Notepad++** ou **CotEditor** par exemple (pas word ni le bloc note !)
- **Découpage en parties (partition)** : le corpus devra contenir au moins une partition (en dates), plus si nécessaire (journal, opinions politique, genre, etc.).
 - Indication à l'aide de jalons textuels : balises
- **Découpage en sections** : à minima, à la fin de chaque article et si besoin, découpage en paragraphes par exemple.
 - Indication à l'aide d'un séparateur de contexte (par défaut : §)

Analyse textométrique

Le corpus formaté pour iTraueur : exemple de partition (structure-corpus-date-genre.txt)

Les dates doivent être au format année, mois, jour (ex : 20230101) pour un affichage chronologique dans les graphiques iTraueur.



```
<genre="autrices">  
<article="Keyword1">  
<date="YYYYMMDD">  
Texte de l'article 1.  
</date>  
</article> §  
...  
<article="Keyword5">  
<date="YYYYMMDD">  
Texte de l'article 5.  
</date>  
</article> §  
</genre>  
<genre="auteurs">  
<article="Keyword6">  
<date="YYYYMMDD">  
Texte de l'article 6.  
</date>  
</article> §  
...  
<article="Keyword10">  
<date="YYYYMMDD">  
Texte de l'article 10.  
</date>  
</article> §  
</genre>
```

Analyse textométrique

Le corpus formaté pour iTrameur : la partition

- **Langage de balisage pour la partition (les règles) :**

- Les balises servent de jalon textuel pour indiquer au logiciel où commence une partie et où elle se termine.
- Il doit y avoir un jeu de balise pour chaque partition. Exemple :
 - Partition en genre = `<genre="autrice">...</genre>`
 - Partition en dates = `<date="20230101">...</date>`
 - ...
- Chaque balise ouvrante doit avoir une balise fermante.
- Chaque caractère compte (ex : `</date` est invalide).
- Les balises doivent être correctement imbriquées :
 - Exemple invalide :
 - `<genre="autrices">`
 - `<article="Keyword1">`
 - `<date="YYYYMMDD">`
 - Texte de l'article 1.
 - `</article> §`
 - `</date>`
 - Ici la balise fermante `</date>` devrait être avant la balise fermante `</article>`

Analyse textométrique

Le corpus formaté pour iTrameur : exemple de découpage en sections (structure-corpus-date-genre.txt)

```
<genre="autrices">
<article="Keyword1">
<date="YYYYMMDD">
Texte de l'article 1.
</date>
</article> §
...
<article="Keyword5">
<date="YYYYMMDD">
Texte de l'article 5.
</date>
</article> §
</genre>
<genre="auteurs">
<article="Keyword6">
<date="YYYYMMDD">
Texte de l'article 6.
</date>
</article> §
...
<article="Keyword10">
<date="YYYYMMDD">
Texte de l'article 10.
</date>
</article> §
</genre>
```

Ces délimiteurs de section permettent d'obtenir une carte des sections.

Avant d'utiliser ce caractère comme délimiteur, il faut vérifier s'il n'est pas déjà présent dans le corpus. Si oui, soit on le supprime, soit on trouve un autre caractère délimiteur de section .

Carte des sections avec délimiteur de sections : §

Clic sur section : affichage contexte | Clic-droit sur section : sélection section



Clic sur section : affichage contexte | Clic-droit sur section : sélection section

Analyse textométrique

Le corpus formaté pour iTrameur : exemple de découpage en sections (le dormeur du val)

Découpage en “lignes” (vers) :

1 Le dormeur du val §
2 C'est un trou de verdure où chante une rivière, §
3 Accrochant follement aux herbes des haillons §
4 D'argent ; où le soleil, de la montagne fière, §
5 Luit : c'est un petit val qui mousse de rayons. §
6
7 Un soldat jeune, bouche ouverte, tête nue, §
8 Et la nuque baignant dans le frais cresson bleu, §
9 Dort ; il est étendu dans l'herbe, sous la nue, §
10 Pâle dans son lit vert où la lumière pleut. §
11
12 Les pieds dans les glaïeuls, il dort. Souriant comme §
13 Sourirait un enfant malade, il fait un somme : §
14 Nature, berce-le chaudement : il a froid. §
15
16 Les parfums ne font pas frissonner sa narine ; §
17 Il dort dans le soleil, la main sur sa poitrine, §
18 Tranquille. Il a deux trous rouges au côté droit. §
19
20 | | Arthur Rimbaud §
21

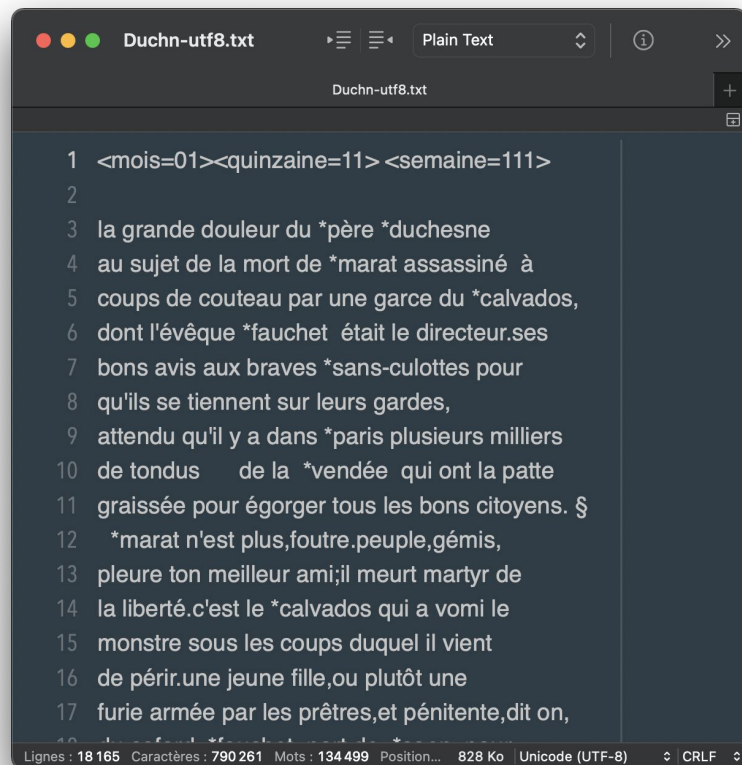


Carte des sections correspondante :

[illegible]

Analyse textométrique

Le corpus formaté pour iTraqueur : exemple de découpage en sections (Le Père Duchesne)



```
1 <mois=01><quinzaine=11> <semaine=111>
2
3 la grande douleur du *père *duchesne
4 au sujet de la mort de *marat assassiné à
5 coups de couteau par une garce du *calvados,
6 dont l'évêque *fauchet était le directeur.ses
7 bons avis aux braves *sans-culottes pour
8 qu'ils se tiennent sur leurs gardes,
9 attendu qu'il y a dans *paris plusieurs milliers
10 de tondus de la *vendée qui ont la patte
11 graissée pour égorger tous les bons citoyens. §
12 *marat n'est plus,foutre.peuple,gémis,
13 pleure ton meilleur ami;il meurt martyr de
14 la liberté.c'est le *calvados qui a vomi le
15 monstre sous les coups duquel il vient
16 de périr.une jeune fille,ou plutôt une
17 furie armée par les prêtres,et pénitente,dit on,
```

Lignes : 18 165 Caractères : 790 261 Mots : 134 499 Position... 828 Ko Unicode (UTF-8) CRLF

Section = paragraphe

Codage de caractères majuscules avec *