



Voyant Tools

Ljudmila PETKOVIC

Introduction aux humanités numériques (L1HN001)
Mineure « Humanités numériques », licence Lettres
Paris, le 30 novembre 2023, année 2023-2024

Les projets en humanités numériques

Quatre principales étapes :

- 1** Acquisition d'un objet d'étude
- 2** Traitement d'un objet d'étude
- 3** Exploitation d'un objet d'étude
- 4** Publication des résultats



NB:

Les frontières entre ces étapes sont très poreuses et l'ordre des opérations ne respecte pas toujours cette progression logique.

Les projets en humanités numériques

Quatre principales étapes :

- 1** Acquisition d'un objet d'étude
- 2** Traitement d'un objet d'étude
- 3** **Exploitation d'un objet d'étude**
- 4** Publication des résultats



NB:

Les frontières entre ces étapes sont très poreuses et l'ordre des opérations ne respecte pas toujours cette progression logique.

Voyant Tools



VOYANT
see through your text

Analyse de texte assistée par ordinateur

- environnement d'analyse, de lecture et de visualisation de textes numériques
- suite d'outils d'analyse de textes accessible [par navigateur](#) ou [en local](#)
- adapté à l'interprétation de textes structurés (XML, HTML) ou non structurés (TXT)
- logiciel de fouille de texte (angl. *text mining*) et de visualisation statistique
- son élaboration s'inscrit dans le projet [Hermeneuti.ca](#) (Rockwell & Sinclair, 2016)

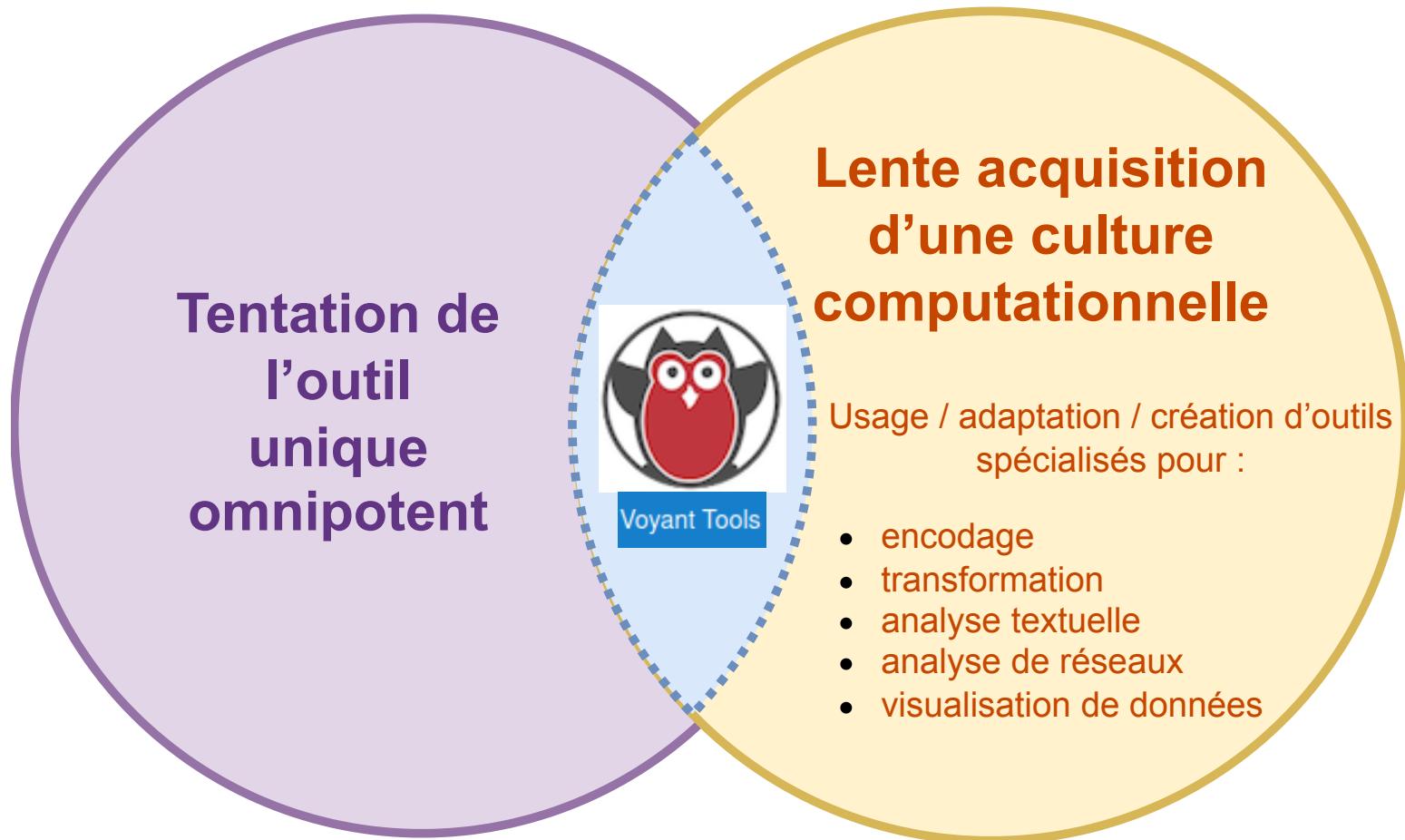
Les outils numériques ne sont jamais seulement des outils.

Principes

- Voyant Tools repose sur un principe d'ouverture
 - son code est publié en code source ouvert (angl. *open source*)
- manipuler, explorer, fouiller les corpus textuels
 - intéragir avec les données de manière rapide, aisée, voire ludique, pour apprendre du processus et produire des séries de résultats
- la théorie et la pratique sont intimement mêlées

Les outils analytiques sont des instantiations de méthodes interprétatives qui peuvent être étroitement liées à d'autres éléments herméneutiques, comme le texte (trad. de Rockwell & Sinclair, 2016, p. 4).

Modularité et caractère évolutif



Attentes vs. réalité dans le contexte de l'utilisation des outils numériques.

Macro-objectifs

1. se familiariser avec l'interface graphique
2. analyser les corpus
3. réfléchir sur le sens et l'intérêt de ces explorations

Page d'accueil



Ajouter des textes

Entrez une URL (ou plusieurs sur des lignes séparées) ou bien copiez tout un texte.

Ouvrir Charger Révéler

Voyant Tools est un environnement en ligne de lecture et d'analyse de textes numériques.
Traduction française d'Aurélien Berra

Import du corpus

1. rendez-vous sur la page d'accueil <https://voyant-tools.org/>
2. deux corpus d'essai sont proposés par défaut et déjà chargés sur la plateforme :
 - pièces de théâtre de William Shakespeare
 - romans de Jane Austen (nous allons tester celui-là)

Cliquer sur **Ouvrir** > **Choisir un corpus** > **Austen's Novels** > **✓ Ouvrir**

Pour explorer votre propre corpus, vous pouvez :

- l'importer dans un format courant (texte brut, HTML ou XML, etc.)
- connaître l'URL d'une page où le texte est accessible sans restriction

Configuration et navigation

- ensemble d'outils (modules), qui sont complémentaires et parfois coordonnés
- des panneaux supplémentaires sont présents lorsque l'on travaille sur une collection de textes

Principes de fonctionnement d'une vue :

- à chaque **outil** correspond un **panneau**, que vous pouvez réduire ou agrandir
- pour chaque panneau, des **options** sont disponibles
- chaque panneau peut être manipulé ou exploré d'une façon qui lui est propre
- chaque panneau peut modifier le contenu d'autres panneaux

Panneaux d'outils



Corpus ▾ Termes:

Résumé Documents Syntagmes

Ce corpus contient 8 documents, 781,763 mots et 15,368 formes verbales uniques. Il a été créé il y a environ 2 années.

Taille du document :

- Ordre décroissant : [1815 Emma](#) (160996); [1814 Mansfield Park](#) (160461); [1813 Pride and Prejudice](#) (122204); [1811 Sense and Sensibility](#)... (119957); [1818 Persuasion](#) (83657)
- Ordre croissant : [1805 Lady Susan](#) (23149); [1790 Love And Freindship](#) (33559); [1818 Northanger Abbey](#) (77780); [1818 Persuasion](#) (83657); [1811 Sense and Sensibilit...](#) (119957)

Densité du vocabulaire :

- Ordre décroissant : [1805 Lady Susan](#) (n 127); [1790 Love And Freindship](#) (n 126); [1818 Northanger Abbey](#) (n 079);

Élément

Voyant Tools

Cirrus Termes Liens Lecteur Grappe de termes Tendances Termes du document

1790 Love And Freindship

LOVE AND FREINDSHIP AND OTHER EARLY WORKS
(Love And Friendship And Other Early Works)
A Collection of Juvenile Writings
By Jane Austen
Transcriber's Note: A few very small changes have been made to this version: Italics have been converted to capitals. The British 'pound' symbol has been converted to 'L'; but in general the author's erratic spelling, punctuation and capitalisations have been retained.
CONTENTS.
Love and Freindship
Lesley Castle
The History of England

Corpus (documents)

Réinitialiser Afficher

Contextes Lignes de bulles Collocations

Document	Gauche	Terme	Droite
1790 Lo...	genteel family since we came.	mr	and Mrs Marlowe are very
1790 Lo...	A brother of Mrs Marlowe,	mr	Cleveland is with them at
1790 Lo...	thirded by the entreaties of	mr	. Fitzgerald who is certainly one
1790 Lo...	a Concert or a Ball.	mr	Marlowe is so desirous of
1790 Lo...	Kickabout's; we were escorted by	mr	Fitzgerald who is a very
1790 Lo...	of my Charlotte at Bristol.	mr	and Mrs M. were the

3,117 Contexte Extract Corpus

Panneaux d'outils

Voyant Tools

Cirrus

Corpus ▾ Termes:

Résumé **Documents** **Syntagmes**

Ce corpus contient 8 documents, 781,763 mots et 15,368 formes verbales uniques. Il a été créé il y a environ 2 années.

Taille du document :

- Ordre décroissant : [1815 Emma](#) (160996); [1814 Mansfield Park](#) (160461); [1813 Pride and Prejudice](#) (122204); [1811 Sense and Sensibility](#)... (119957); [1818 Persuasion](#) (83657)
- Ordre croissant : [1805 Lady Susan](#) (23149); [1790 Love And Freindship](#) (33559); [1818 Northanger Abbey](#) (77780); [1818 Persuasion](#) (83657); [1811 Sense and Sensibilit...](#) (119957)

Densité du vocabulaire :

- Ordre décroissant : [1805 Lady Susan](#) (n 127); [1790 Love And Freindship](#) (n 126); [1818 Northanger Abbey](#) (n 079);

Élément

Lecteur

1790 Love And Freindship

LOVE AND FREINDSHIP AND OTHER EARLY WORKS
(Love And Friendship And Other Early Works)
A Collection of Juvenile Writings
By Jane Austen
Transcriber's Note: A few very small changes have been made to this version: Italics have been converted to capitals. The British 'pound' symbol has been converted to 'L'; but in general the author's erratic spelling, punctuation and capitalisations have been retained.
CONTENTS.
Love and Freindship
Lesley Castle
The History of England

Corpus ▾ ? Réinitialiser Afficher

Tendances

Frequencies relatives

Corpus (documents)

mr mrs said miss think

Corpus (documents)	mr	mrs	said	miss	think
1790 Love...	0.0015	0.0005	0.0030	0.0015	0.0010
1805 Lady...	0.0035	0.0025	0.0010	0.0010	0.0005
1811 Sense...	0.0065	0.0045	0.0030	0.0025	0.0015
1813 Pride...	0.0068	0.0030	0.0025	0.0025	0.0015
1814 Mansf...	0.0025	0.0045	0.0030	0.0035	0.0020
1815 Emma	0.0072	0.0040	0.0025	0.0035	0.0020
1818 North...	0.0025	0.0035	0.0020	0.0025	0.0015
1818 Persu...	0.0015	0.0035	0.0020	0.0015	0.0015

Contextes

Document	Gauche	Terme	Droite
1790 Lo...	genteel family since we came.	mr	and Mrs Marlowe are very
1790 Lo...	A brother of Mrs Marlowe,	mr	Cleveland is with them at
1790 Lo...	thirded by the entreaties of	mr	. Fitzgerald who is certainly one
1790 Lo...	a Concert or a Ball.	mr	Marlowe is so desirous of
1790 Lo...	Kickabout's; we were escorted by	mr	Fitzgerald who is a very
1790 Lo...	of my Charlotte at Bristol.	mr	and Mrs M. were the

3,117 Contexte Extract Corpus

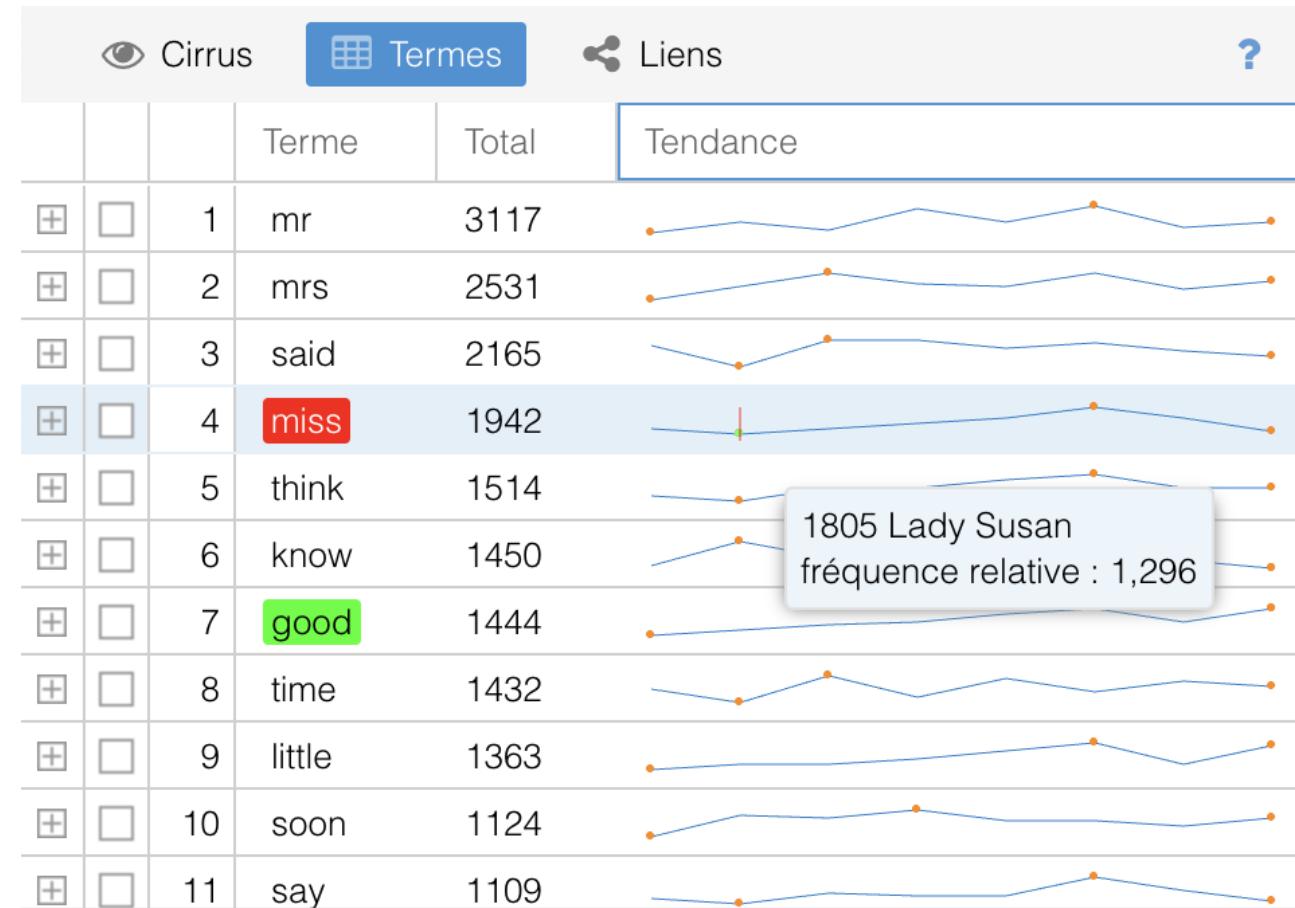
Cirrus

- nuage de mots montrant les termes les plus fréquents dans *Pride and Prejudice*



Termes (*Terms*)

- la fondamentale table des fréquences

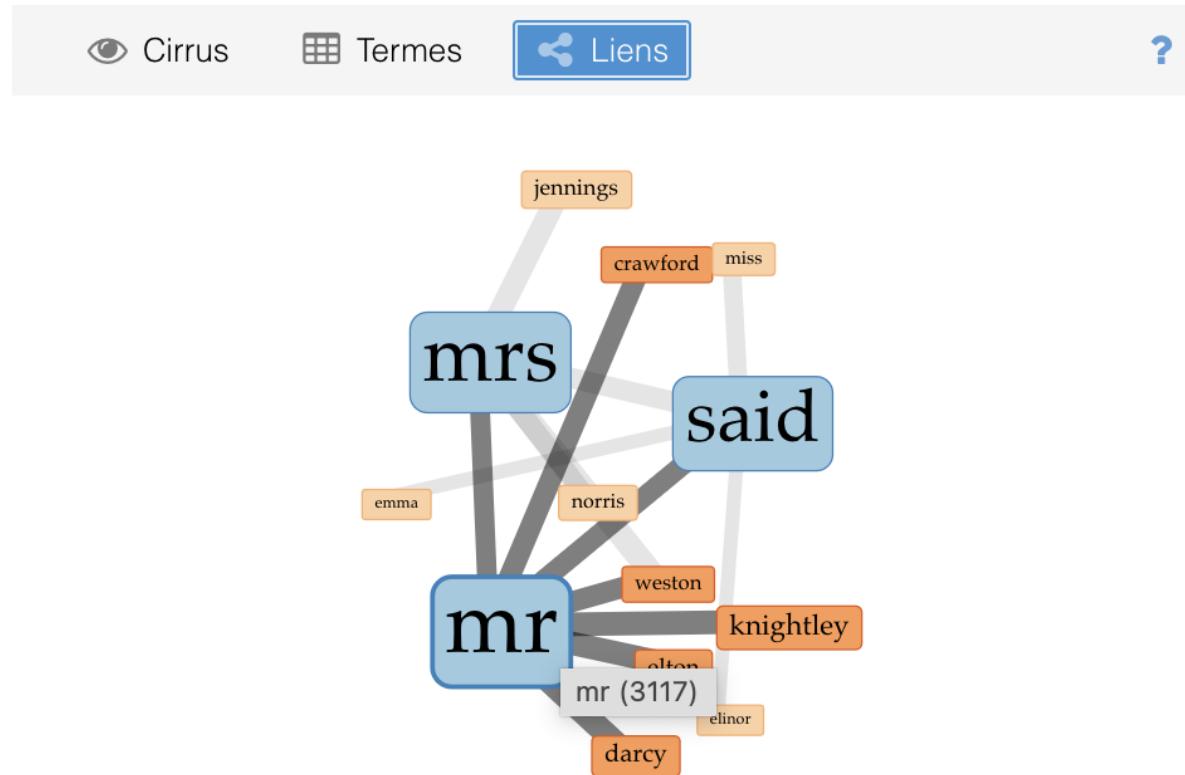


Termes (*Terms*)

- **Terme** : ce terme dans le corpus
- **Total** : fréquence absolue pour un terme dans le corpus entier
- **Tendance** : représentation de la tendance des fréquences relatives pour chaque terme de chaque document du corpus
 - le graphique de type ***sparkline***
 - résume dans la colonne de droite les tendances des termes au fil du corpus, en indiquant les valeurs minimale, maximale et finale
 - vous pouvez survoler la sparkline pour voir des résultats plus précis

Liens (Links)

- visualisation des termes fréquemment liés (double clic sur un terme → autres liens)



Liens (*Links*)

- analyse des réseaux (angl. *social network analysis*)
 - **nœuds** (sommets) : en l'occurrence, les termes
 - **arêtes** (arcs, si graphe orienté) : liens / relations entre ces nœuds (termes)

Lecteur (Reader)

- lecteur de corpus qui récupère des segments de texte lorsque vous faites défiler (formatage minimal)
- passez le curseur de la souris sur des mots pour observer leurs fréquences
- recherche des mots (cf. onglet collocations)

1813 Pride and Prejudice

PRIDE AND PREJUDICE

By Jane Austen

Chapter 1

It is a truth universally acknowledged, that a single man in possession of a good fortune, must be in want of a wife.

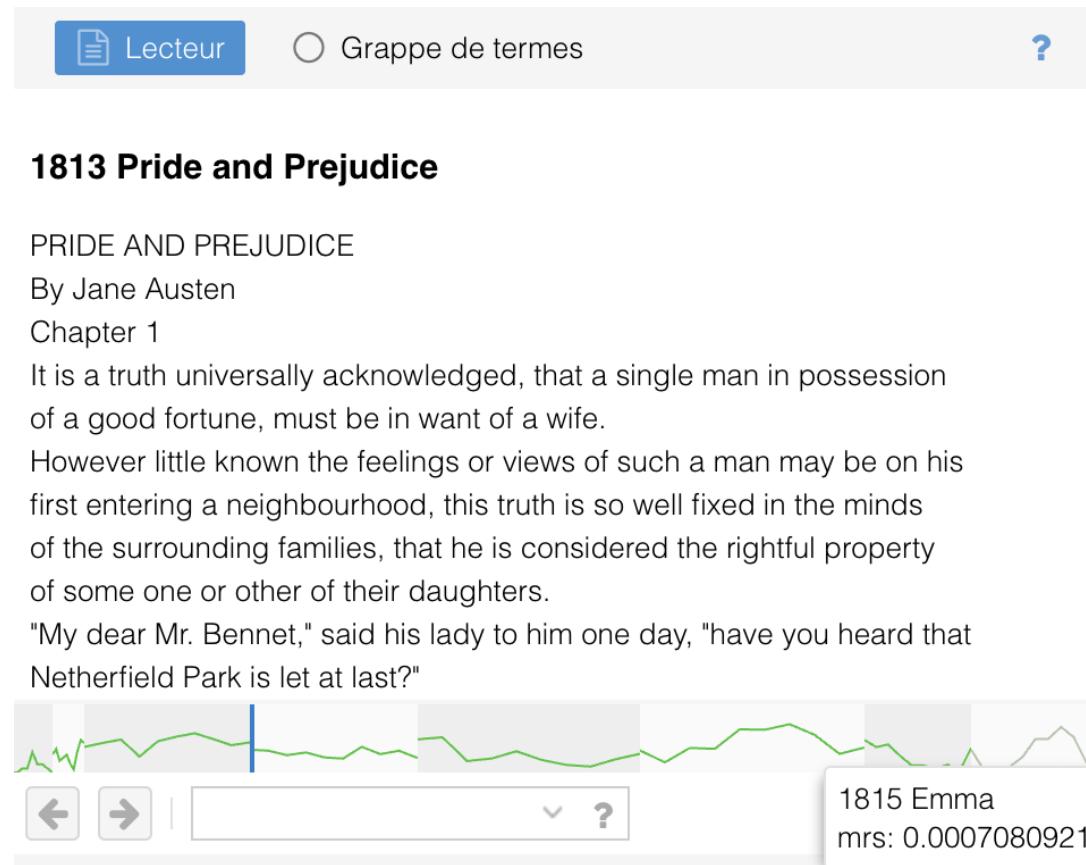
However little known the feelings or views of such a man may be on his first entering a neighbourhood, this truth is so well fixed in the minds of the surrounding families, that he fréquence dans le document : 86

of some one or other of their daughters.

"My dear Mr. Bennet," said his lady to him one day, "have you heard that Netherfield Park is let at last?"

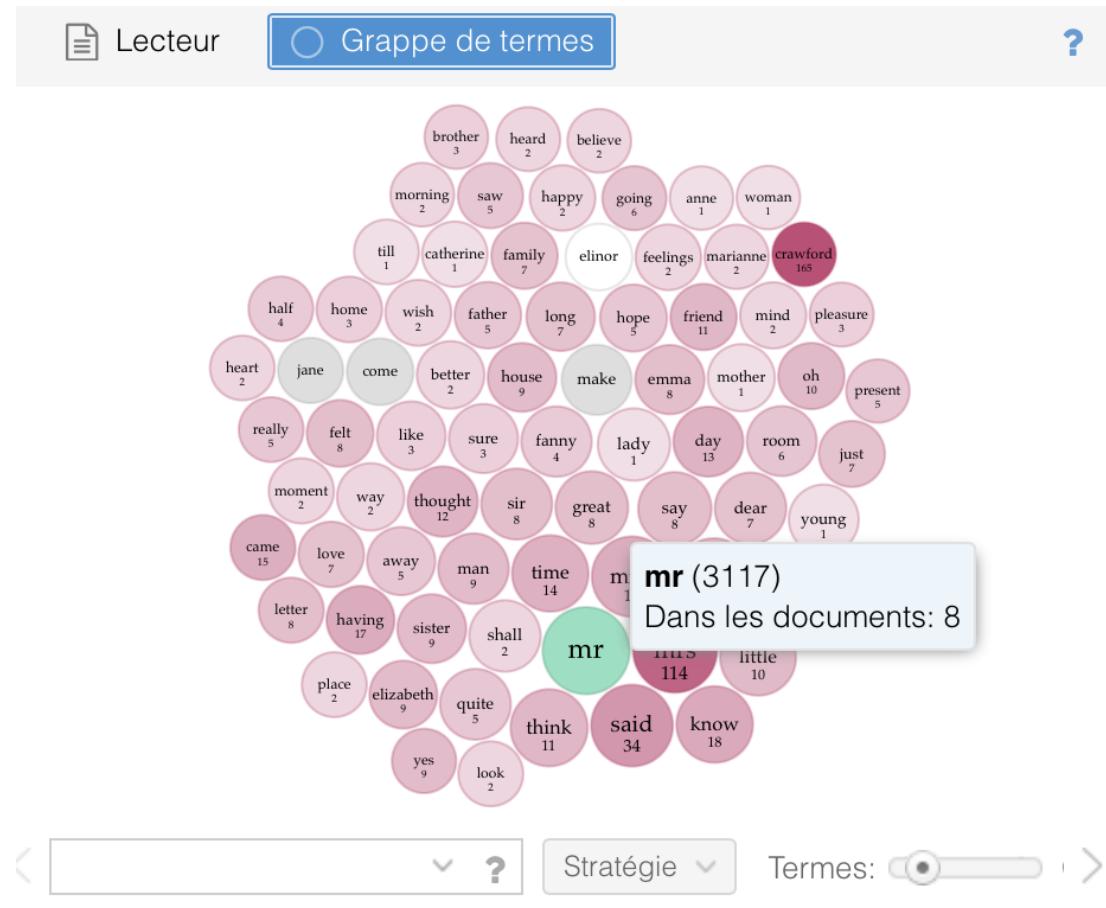
Lecteur (Reader)

- **frise chronologique** : répartition d'un terme avec sa fréquence relative à travers les documents d'un corpus



Grappe de termes (*TermsBerry*)

- explore les termes fréquents et leurs co-occurrences (mots qui apparaissent à proximité)

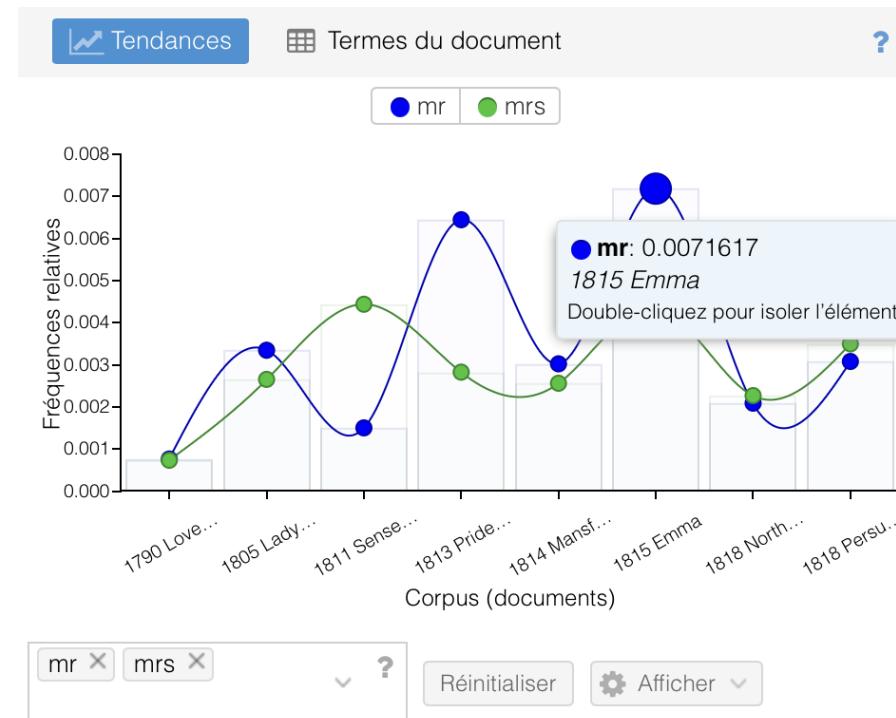


Grappe de termes (*TermsBerry*)

- grappe ou en angl. *cluster*
- les termes les plus fréquents (ou les termes les plus distincts, si vous modifiez les options) apparaissent au milieu et dans des bulles plus grandes
- les termes sont répartis en spirale vers l'extérieur
- la noirceur des termes représente la proportion de documents où le terme apparaît
 - plus foncé signifie qu'il apparaît dans plus de documents
 - il n'y aura pas de différenciation s'il n'y a qu'un seul document dans le corpus

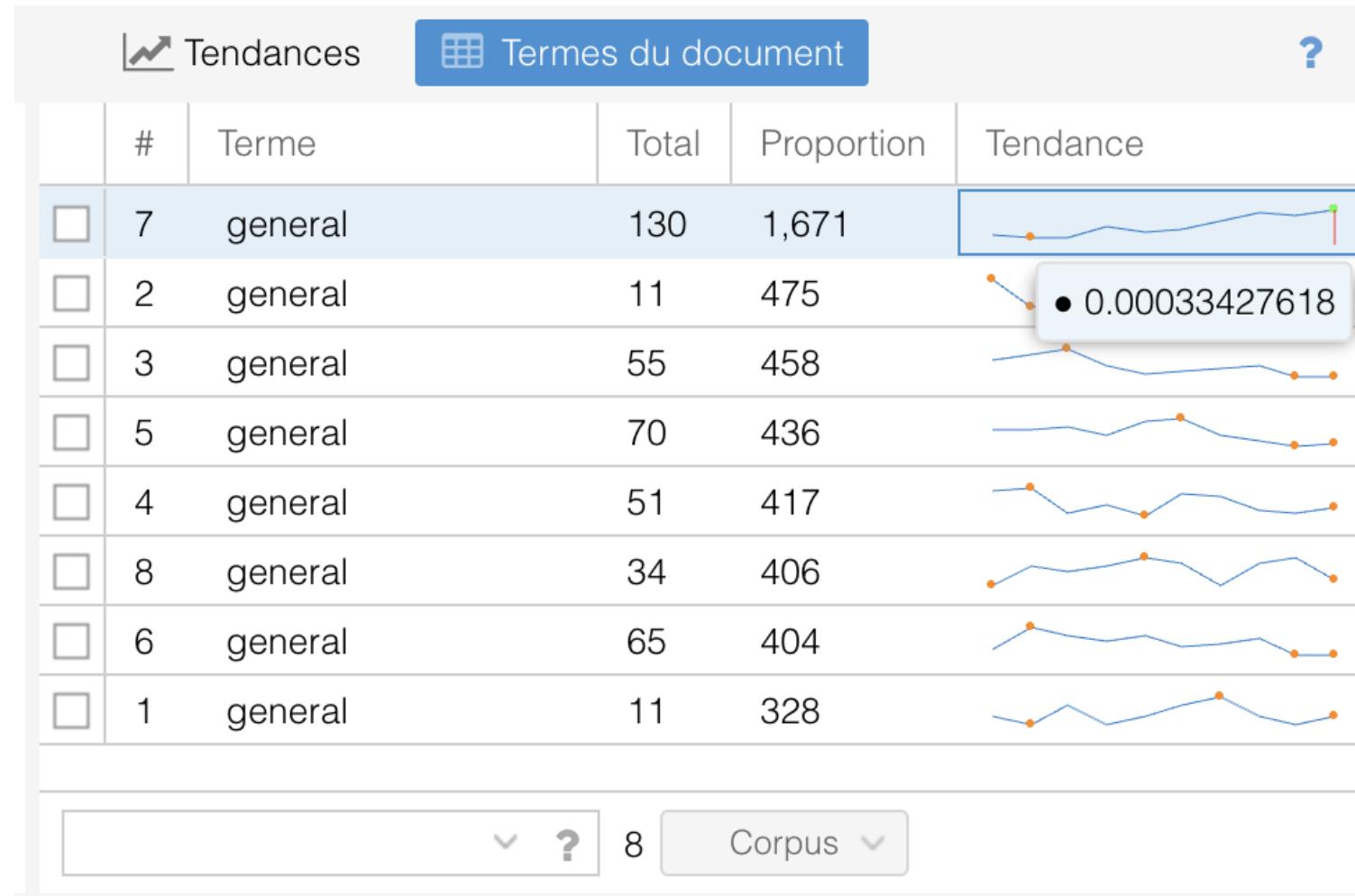
Tendances (*Trends*)

- graphe de distribution du ou des termes sélectionnés au sein du corpus (ou dans un document) → « cadence » d'un mot (Tournier, 1980)
- le corpus est découpé en fonction des documents qui le constituent ou en segments d'égale longueur



Termes du document (*Document Terms*)

- vue sous forme de tableau des fréquences des termes pour chaque document



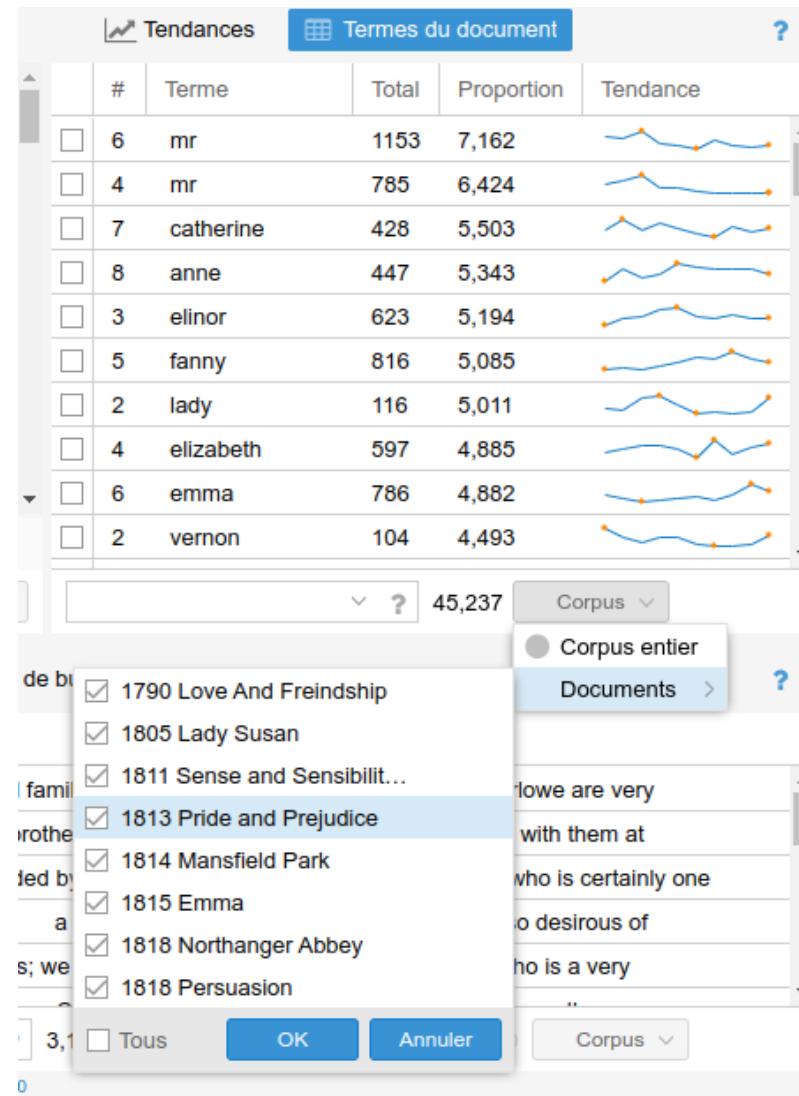
Termes du document (*Document Terms*)

- **#** : numéro (position) du document dans le corpus
 - 1 – *Love and Friendship*, 2 – *Lady Susan*... 8 – *Persuasion*
- **Terme** : terme dans un document unique
- **Total** : fréquence absolue de ce terme dans le document
- **Proportion** : fréquence relative (par million) de ce terme dans chaque document
- **Tendance** : graphique *sparkline*, distribution du terme au sein de segments continus du document (par défaut, 10 segments de taille égale)

Termes de document (*Document Terms*)

- vous pouvez observer les numéros associés aux documents en cliquant sur le bouton **Corpus** en bas, puis **Documents**
- vous pouvez cocher des cases correspondant à un / plusieurs documents que vous voulez prendre en compte, ou au corpus entier
- les autres calculs (**Total**, **Proportion**, **Tendance**) sont donc effectués en fonction de chaque document :
 - p. ex. le terme *mr* est présent 1153 fois (**fréquence absolue**), avec une **fréquence relative** de 7,162 dans le document n° 6, *Emma*
 - si vous cochez la case à gauche du premier terme listé *mr*, l'onglet **Contexte** en bas s'actualise, en indiquant que le numéro 6 est bien celui d'*Emma*
 - par conséquent, le graphique **sparkline** de l'onglet Tendance reflète la distribution de ce terme au sein de segments continus de ce document

Termes de documents (*Document Terms*)



Résumé (Summary)

- synthèse quantitative du corpus

Résumé Documents Syntagmes ?

Ce corpus contient 8 documents, 781,763 mots et 15,368 formes verbales uniques. Il a été créé il y a environ 2 années.

Taille du document : 

- Ordre décroissant : [1815 Emma](#) (160996); [1814 Mansfield Park](#) (160461); [1813 Pride and Prejudice](#) (122204); [1811 Sense and Sensibilit...](#) (119957); [1818 Persuasion](#) (83657)
- Ordre croissant : [1805 Lady Susan](#) (23149); [1790 Love And Freindship](#) (33559); [1818 Northanger Abbey](#) (77780); [1818 Persuasion](#) (83657); [1811 Sense and Sensibilit...](#) (119957)

Densité du vocabulaire : 

- Ordre décroissant : [1805 Lady Susan](#) (0.127); [1790 Love And Freindship](#) (0.126); [1818 Northanger Abbey](#) (0.079); [1818 Persuasion](#) (0.070); [1811 Sense and Sensibilit...](#) (0.054)
- Ordre croissant : [1815 Emma](#) (0.046); [1814 Mansfield Park](#) (0.050); [1813 Pride and Prejudice](#) (0.054); [1811 Sense and Sensibilit...](#) (0.054); [1818 Persuasion](#) (0.070)

Moyenne des mots par phrase : 

- Ordre décroissant : [1790 Love And Freindship](#) (25.8); [1805 Lady Susan](#) (25.2); [1811 Sense and Sensibilit...](#) (23.9); [1814 Mansfield Park](#) (23.6); [1818 Persuasion](#) (23.3)
- Ordre croissant : [1815 Emma](#) (19.2); [1813 Pride and Prejudice](#) (20.7); [1818 Northanger Abbey](#) (22.2); [1818 Persuasion](#) (23.3); [1814 Mansfield Park](#) (23.6)

Résumé (Summary) II

- synthèse quantitative du corpus

Readability Index: 

- Ordre décroissant : [1811 Sense and Sensibility \(7.842\)](#); [1790 Love And Freindship \(7.769\)](#); [1818 Northanger Abbey \(7.744\)](#); [1818 Persuasion \(7.734\)](#); [1813 Pride and Prejudice \(7.456\)](#)
- Ordre croissant : [1805 Lady Susan \(7.036\)](#); [1815 Emma \(7.164\)](#); [1814 Mansfield Park \(7.404\)](#); [1813 Pride and Prejudice \(7.456\)](#); [1818 Persuasion \(7.734\)](#)

Mots les plus fréquents :

- [mr](#) (3117); [mrs](#) (2531); [said](#) (2165); [miss](#) (1942); [think](#) (1514)

Mots distinctifs (par rapport au reste du corpus) :

- [1790 Love And Freindship](#) : [tho](#) (50), [freind](#) (40), [lesley](#) (39), [augustus](#) (23), [laura](#) (44).
- [1805 Lady Susan](#) : [vernon](#) (104), [reginald](#) (74), [frederica](#) (72), [courcy](#) (51), [churchill](#) (46).
- [1811 Sense and Sensibilit...](#) : [elinor](#) (623), [marianne](#) (492), [jennings](#) (199), [dashwood](#) (231), [willoughby](#) (181).
- [1813 Pride and Prejudice](#) : [darcy](#) (373), [bennet](#) (294), [bingley](#) (257), [elizabeth](#) (597), [wickham](#) (162).
- [1814 Mansfield Park](#) : [crawford](#) (493), [edmund](#) (364), [fanny](#) (816), [bertram](#) (240), [rushworth](#) (185).
- [1815 Emma](#) : [weston](#) (389), [emma](#) (786), [knightley](#) (356), [elton](#) (319), [woodhouse](#) (277).
- [1818 Northanger Abbey](#) : [tilney](#) (196), [thorpe](#) (126), [morland](#) (125), [allen](#) (116), [catherine](#) (428).
- [1818 Persuasion](#) : [elliot](#) (254), [wentworth](#) (191), [russell](#) (118), [walter](#) (123), [uppcross](#) (77).

Élement

Résumé (*Summary*)

- **mots (tokens)** : nombre de mots individuels trouvés dans le document (p. ex., chaque occurrence de « le » est comptée)
- **formes verbales uniques (types)** : nombre de formes de mots trouvées dans le document (p. ex., toutes les occurrences de « le » → une seule forme de mot)
- **taille du document** : nombre de mots du document
- **densité du vocabulaire** : ratio de mots uniques dans le document (mots uniques divisés par le nombre total de mots)
- **moyenne des mots par phrase** : nombre moyen de mots par phrase
- **readability index (indice de lisibilité)** : évalue la compréhensibilité d'un texte
- **mots les plus fréquents**
- **mots distinctifs** : mots fréquents et relativement distinctifs d'un document

Documents

- rappelle la structure du corpus

The screenshot shows a software interface for corpus analysis. At the top, there are four tabs: 'Résumé' (Summary), 'Documents' (selected), 'Syntagmes', and a question mark icon. Below the tabs is a table with the following data:

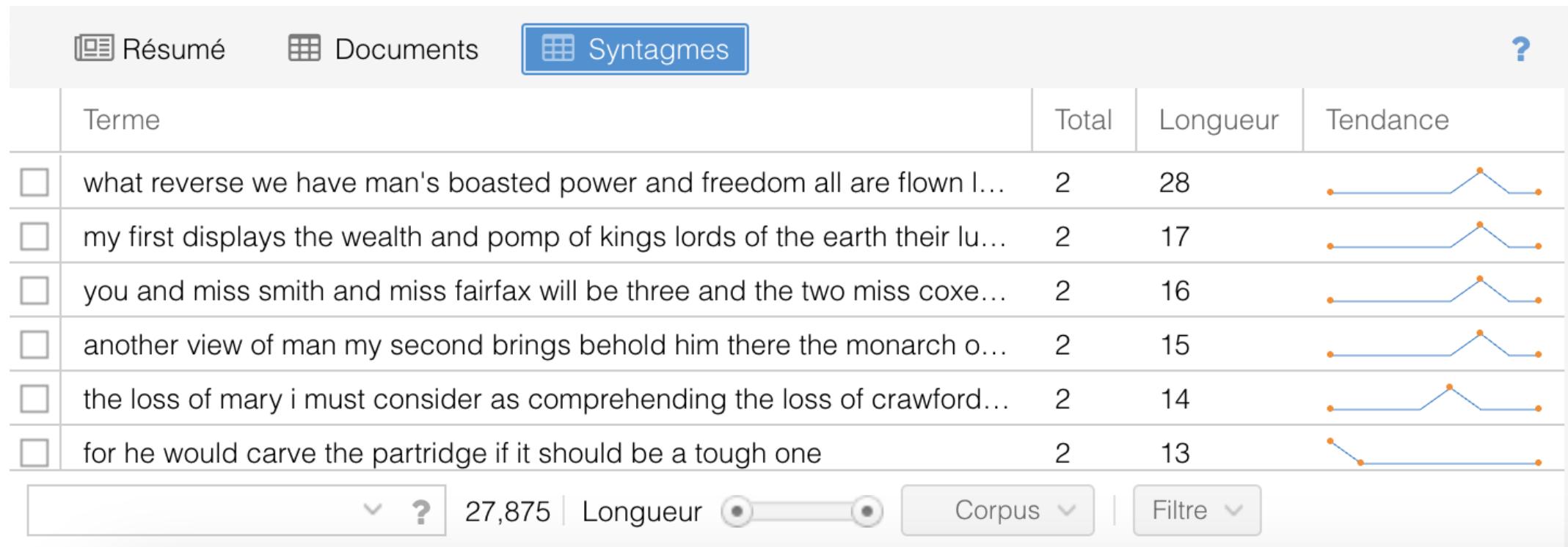
	Titre	Mots	Types	Proportion	Mots/Phrase
1	1790 Love And Freindship	33,559	4,235	13%	25.8
2	1805 Lady Susan	23,149	2,929	13%	25.2
3	1811 Sense and Sensibility	119,957	6,419	5%	23.9
4	1813 Pride and Prejudice	122,204	6,538	5%	20.7
5	1814 Mansfield Park	160,461	8,077	5%	23.6
6	1815 Emma	160,996	7,356	5%	19.2
7	1818 Northanger Abbey	77,780	6,132	8%	22.2
8	1818 Persuasion	83,657	5,858	7%	23.3

Documents

- **Titre** : titre du document (ou son nom de fichier si aucun titre n'a été trouvé)
- **Mots** : (*tokens*) : nombre de mots individuels trouvés dans le document (p. ex., chaque occurrence de « le » est comptée)
- **Types** : formes verbales uniques, nombre de formes de mots trouvées dans le document (p. ex., toutes les occurrences de « le » → une seule forme de mot)
- **Proportion** : rapport types/tokens, exprimé en pourcentage ; des nombres plus élevés → une plus grande diversité de vocabulaire
- **Mots/Phrase** : estimation de la moyenne des mots par phrase

Syntagmes (*Phrases*)

- propose un affichage sous forme de tableau de syntagmes ou groupes de mots



Syntagmes (*Phrases*)

- **Terme** : syntagme, ou expression-clé (forme générique, qui peut varier légèrement entre les occurrences)
- **Total** : nombre d'occurrences du syntagme dans le corpus
- **Longueur** : le nombre de mots dans cette phrase
- **Tendances** : graphique *sparkline*, répartition des fréquences absolues entre les documents du corpus (si le corpus contient plus d'un document)

Syntagmes (*Phrases*)

Même principe que pour l'outil Termes du document :

- cliquer sur **Corpus** > **Corpus entier** ou **Documents** > *1813 Pride and Prejudice* (ou un autre document) pour sélectionner le(s) document(s) à partir desquels les calculs **Total**, **Longueurs** et **Tendance** seront effectués

The screenshot shows the Voyant Tools interface with the 'Syntagmes' tab active. A dropdown menu is open, listing various documents with checkboxes. The document '1813 Pride and Prejudice' is selected. Other checked documents include '1790 Love And Freindship', '1805 Lady Susan', '1811 Sense and Sensibilit...', '1814 Mansfield Park', '1815 Emma', '1818 Northanger Abbey', and '1818 Persuasion'. The 'Corpus entier' option is also selected in the dropdown.

Terme	Total	Longueur
what reverse we have man's boasted power and freedom ...	2	28
my first displays the wealth and pomp of kings lords of the...	2	17
you and miss smith and miss fairfax will be three and the t...	2	16
another view of man my second brings behold him there t...	2	15
the loss of mary i must consider as comprehending the lo...	2	

27,875 | Longueur

<https://voyant-tools.org/?corpus=austen#>

Contextes (Contexts)

- angl. *Keywords in Context (KWIC)* : concordance qui montre chaque occurrence d'un mot-clé avec un peu de contexte

The screenshot shows a software interface for keyword search. At the top, there are three tabs: "Contextes" (selected), "Lignes de bulles", and "Collocations". Below the tabs is a header row with columns: "Document", "Gauche", "Terme", and "Droite". The "Terme" column is highlighted with a blue border. The main area contains a list of search results, each consisting of a snippet from a document (Document), the word being searched (Terme), and some context (Gauche and Droite). The results are listed vertically, showing various instances of the keyword "Marlowe" and its context in the text.

Document	Gauche	Terme	Droite
1790 Lo...	gentee family since we came.	mr	and Mrs Marlowe are very
1790 Lo...	A brother of Mrs Marlowe,	mr	Cleveland is with them at
1790 Lo...	thirded by the entreaties of	mr	. Fitzgerald who is certainly one
1790 Lo...	a Concert or a Ball.	mr	Marlowe is so desirous of
1790 Lo...	Kickabout's; we were escorted by	mr	Fitzgerald who is a very
1790 Lo...	of my Charlotte at Bristol.	mr	and Mrs M. were the
1790 Lo...	handsome?) The elegant address of	mr	Cleveland, his polished Manners and
1790 Lo...	conspicuous in the conversation of	mr	Cleveland. The approach of Sir
1790 Lo...	We have since often met	mr	and Mrs Marlowe but no
1790 Lo...	and whose only ones now	mr	Whitaker, Mrs Lefroy, Mrs Knight
1790 Lo...	particulars of his life, to	mr	Sheridan's play of the Critic
1790 Lo...	shall proceed by degrees.--Tomorrow	mr	Stanly's family will drink tea
1790 Lo...	a private Ball given by	mr	Ashburnham. As my Mother never
1790 Lo...	was engaged to dance by	mr	Bernard, but just as we
1790 Lo...	so exposed before every one--	mr	Bernard however soon returned and
1790 Lo...	freind We dined yesterday with	mr	Evelyn where we were introduced
1790 Lo...	she was a relation of	mr	Evelyn, and that her name
1790 Lo...	I beleive." But how will	mr	and Mrs Grenville bear your
1790 Lo...	expect a fortune with you.	mr	Musgrove is so far from

Contextes (*Contexts*)

- **Document** : document où se trouve l'occurrence
- **Gauche** : contexte gauche du mot-clé
- **Terme** : mot-clé du contexte
- **Droite** : contexte droit du mot-clé

Lignes de bulles (*Bubblelines*)

- visualise la fréquence et la distribution des termes dans un corpus
- répartition et (ré)apparitions du mot au sein de chaque texte



Lignes de bulles (*Bubblelines*)

- chaque document du corpus est représenté sous forme de ligne horizontale et divisé en segments d'égale longueur (50 segments par défaut)
- chaque mot sélectionné est représenté sous forme de bulle dont la taille indique la fréquence du mot dans le segment de texte correspondant
- plus la bulle est grande, plus le mot apparaît fréquemment

Collocations (*Collocates*)

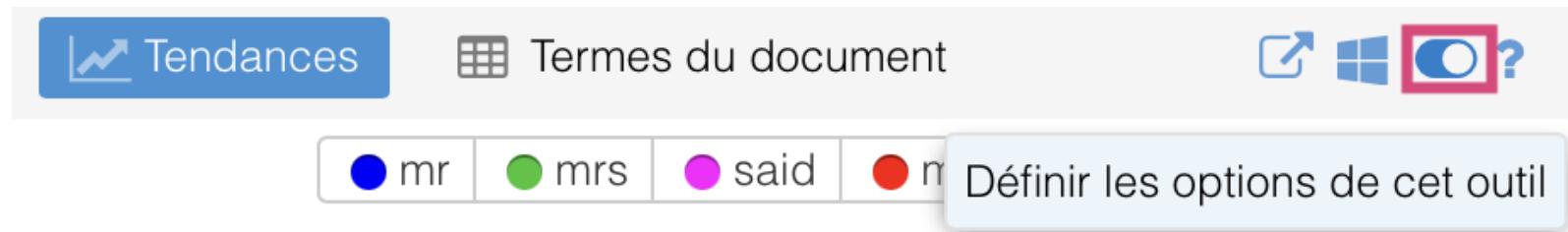
- liste des paires de mots apparaissant souvent dans les mêmes contextes
- **total (contexte)** : nombre d'occurrences de la **collocation** près du mot-clé (**terme**) dans le corpus

Terme	Collocation	Total (contexte)
mr	knightley	321
mr	darcy	254
mrs	weston	247
miss	crawford	229
mr	elton	218
mrs	jennings	202
mr	crawford	180
miss	woodhouse	179
mr	weston	176
said	mrs	167
said	mr	165

21,437 Contexte Corpus

Options

- manipuler les paramètres d'un outil (panneau)
 - afficher les fréquences absolues ou relatives
 - modifier la liste des *stopwords* enlevés en fonction de la langue
 - les options varient en fonction des outils utilisés



Aide

- vous pouvez consulter les explications de chaque outil

L'outil *Termes* montre sous la forme d'un tableau, dans l'ensemble du corpus, les mots qui y apparaissent. À noter :

- il est possible de réarranger le tableau en prenant pour critère de tri le terme ou le nombre de termes (cliquez sur les en-têtes des colonnes)
- un graphique *sparkline* des tendances de fréquence dans la totalité du corpus (si le corpus comprend plusieurs documents) ou dans la totalité du document (si le corpus ne comprend qu'un seul document)
- il est possible d'ajouter des colonnes (fréquence relative, *peakedness*, ou coefficient de finesse de la distribution, et *skew*, ou coefficient de dissymétrie de la distribution) en cliquant sur la flèche qui apparaît au survol d'un en-tête
- une case de recherche est disponible (l'aide de la syntaxe des requêtes apparaît au survol de l'icône représentant une loupe)

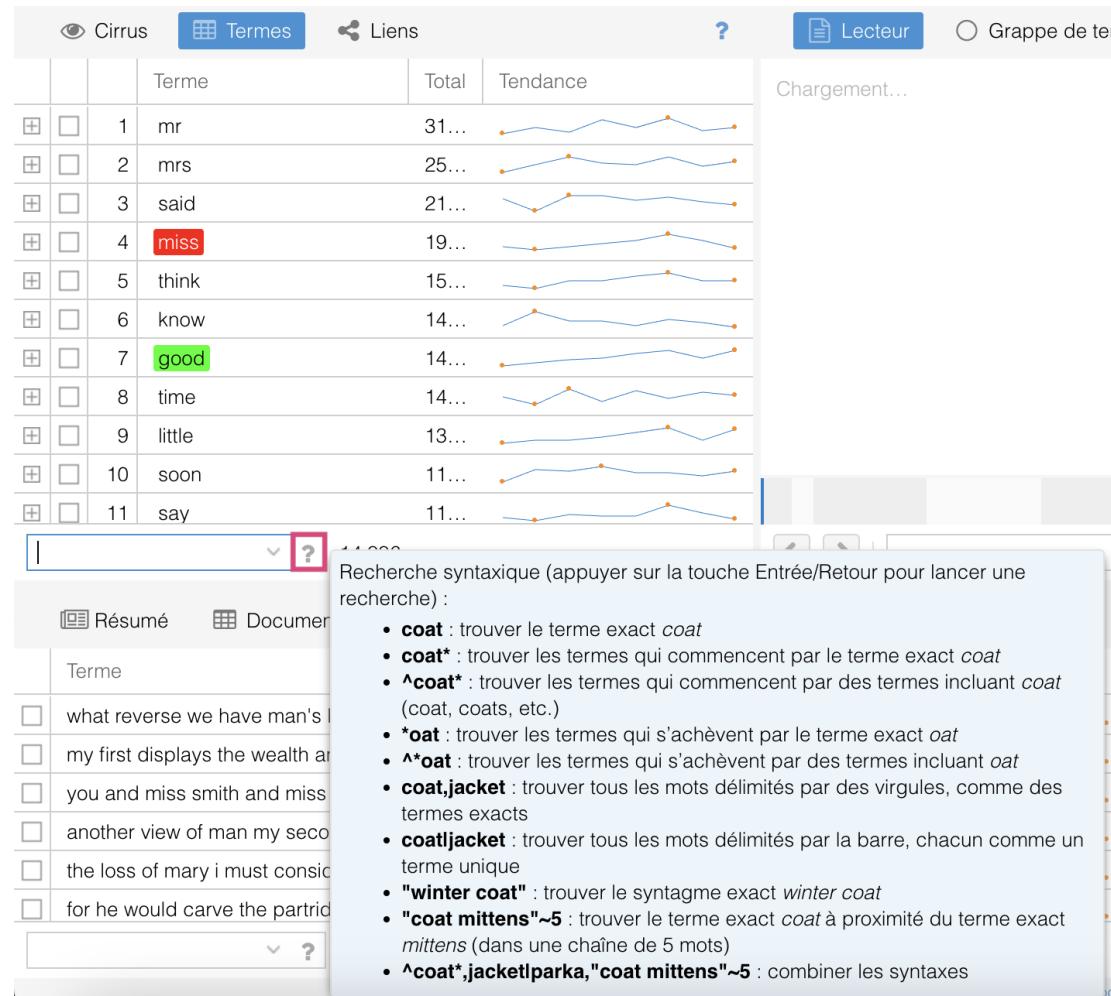
		Terme	Total	Tendance
1	mr	31...		
2	mrs	25...		
3	said	21...		
4	miss	19...		
5	think	15...		
6	know	14...		
7	good	14...		
8	time	14...		
9	little	13...		
10	soon	11...		
11	say	11...		

14,996

Résumé Documents Syntagmes Contextes Lignes de bulles

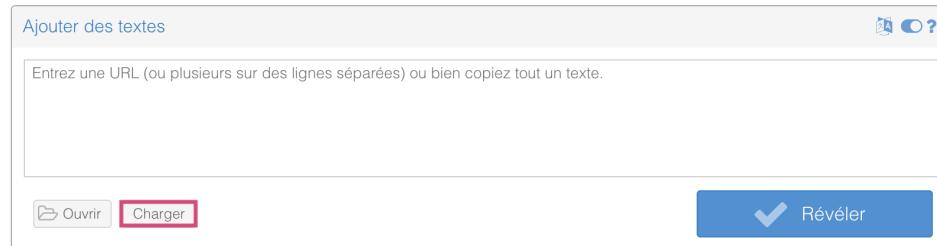
Regex

- recherche avancée avec les **expressions régulières (regex)**



Exercice 1

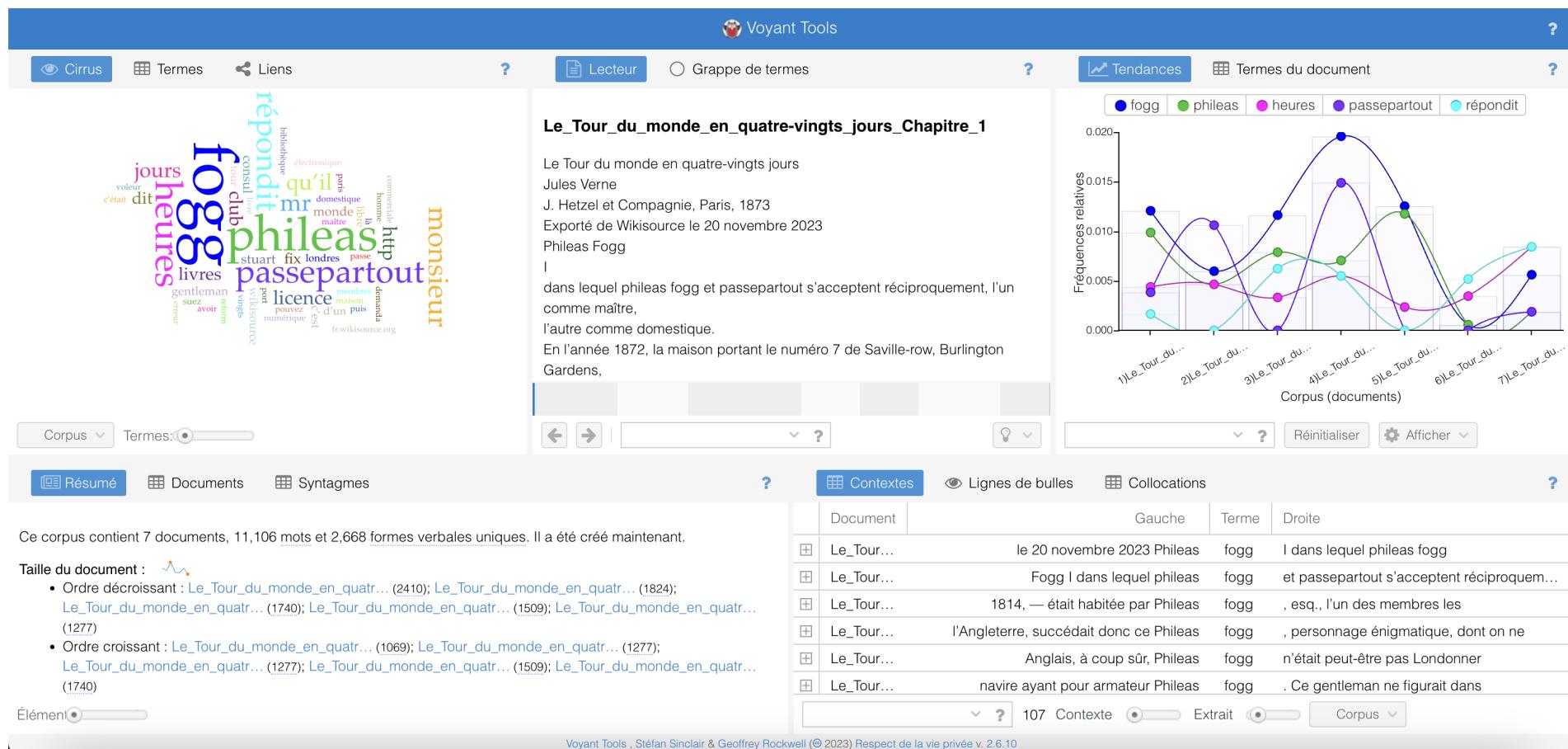
- Téléchargez le corpus *Le tour du monde en quatre-vingts jours* utilisé lors de la séance du 23 novembre 2023, « Reconnaissance d'entités nommées ».
- Importez-le dans Voyant Tools depuis la page d'accueil (bouton **Charger**).



Voyant Tools est un environnement en ligne de lecture et d'analyse de textes numériques.
Traduction française d'Aurélien Berra

Exercice 2

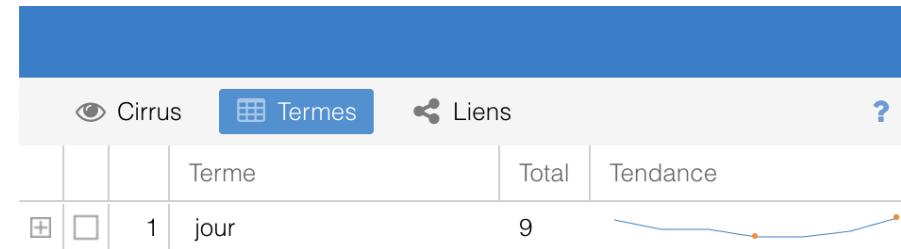
- Observez les résultats et choisissez un ou plusieurs aspects à analyser, en manipulant les différents éléments présentés lors de cette séance.



Exercice 3

- Formulez une regex qui permet de récupérer tous les termes qui commencent par le terme exact *jour*.

Solution de l'exercice 3



Tutoriels sur Voyant Tools (sélection)

Vidéo

- [Tuto VOYANT](#) (Riguet, 2018)

Écrits / visuels

- [Introduction à Voyant Tools](#) (Berra, 2023)
- « [Tutorial/Workshops](#) ». Voyant Tools Help (Rockwell & Sinclair, s.d.)

Références

- **Berra, A.** (2023). « Introduction à Voyant Tools » [dépôt GitHub].
https://github.com/aurelberra/voyant_tools/blob/master/tutorial/voyant_tools_intro_f_r.md
- **Galleron, I.** (2021). « Introduction aux humanités numériques (L1HN001). [diapositives en interne].
- **Riguet, M.** (2018). *Tuto VOYANT* [vidéo]. YouTube.
<https://www.youtube.com/watch?app=desktop&v=W4EWdzJzCFQ>
- **Rockwell, G., & Sinclair, S.** (2016). *Hermeneutica. Computer-Assisted Interpretation in the Humanities*. Cambridge, Massachusetts, MIT Press, 2016.
<http://mitpress.mit.edu/9780262545891/>
- **Rockwell, G., & Sinclair, S.** (s.d.). « Tutorial/Workshops ». Voyant Tools Help.
<https://voyant-tools.org/docs/#!/guide/tutorial>

Références II

- **Tournier, M.** (1980). En souvenir de Lagado. *Mots. Les langages du politique*, 1(1), 5-9. https://www.persee.fr/doc/mots_0243-6450_1980_num_1_1_1000