

**Journée OSIDMESH  
Montpellier – 23/10/2009**

# ***Analyses de textes avec TreeCloud et Lexico3***

**Philippe Gambette**



# Plan

---

- **Nuages de mots**
- **Nuages améliorés**
- **Analyses plus fines**

# Plan

---

- **Nuages de mots**
- Nuages améliorés
- Analyses plus fines













# Plan

---

- Nuages de mots
- **Nuages améliorés**
- Analyses plus fines

# Nuages de mots améliorés

Ajouter de l'information extraite du texte :

- pâleur pour exprimer la désuétude dans Amazon
- tags partagés en rouge dans del.icio.us
- regrouper les tags cooccurrents sur la même ligne  
Hassan-Montero & Herrero-Solana, InScit'06
- optimiser l'espace vide et la proximité sémantique  
Kaser & Lemire, WWW'07
- “topigraphy”: placement 2D d'après la cooccurrence  
Fujimura, Fujimura, Matsubayashi, Yamada & Okuda, WWW'08

# Nuages de tags/mots améliorés

Ajouter de l'information extraite du texte :

- pâleur pour exprimer la désuétude dans Amazon
- tags partagés en rouge dans del.icio.us
- regrouper les tags cooccurrents sur la même ligne
- optimiser l'espace vide et la proximité sémantique
- “topigraphy”: placement 2D d'après la cooccurrence

Hassan-Montero & Herrero-Solana, InScit'06

Kaser & Lemire, WWW'07

Fujimura, Fujimura, Matsubayashi, Yamada & Okuda, WWW'08

## Most Popular Tags (What's this?)

Welcome to the Amazon.com tag cloud. Tags are labels customers can use to classify a product. More frequently used tags are larger and more recently used tags will appear darker.

1080p action adventure american history animation anime art baby best canceled tv shows  
biography blu-ray book business canon children childrens books christian christianity  
christmas classic classic movie classic rock classical music comedy comics cookbook cooking  
defectivebydesign digital camera disney drama dvd erotica exercise family fantasy fiction stress fun  
games gift idea graphic novel harry potter hd dvd hdtv health hip hop historical fiction historical  
romance history horror humor inspirational ipod jazz kids kindle love magic manga  
meditation memoir metal movie mp3 player music mystery nonfiction paranormal  
romance pc game philosophy photography playstation 3 poetry politics progressive rock psychology  
religion rock romance rpg science science fiction self-help sex soundtrack  
spirituality suspense thriller toys travel tv series vampire vampire romance video  
games wii women world war ii xbox 360

# Nuages de tags/mots améliorés

Ajouter de l'information extraite du texte :

- pâleur pour exprimer la désuétude dans Amazon
- tags partagés en rouge dans del.icio.us
- regrouper les tags cooccurrents sur la même ligne  
Hassan-Montero & Herrero-Solana, InScit'06
- optimiser l'espace vide et la proximité sémantique  
Kaser & Lemire, WWW'07
- “topigraphy”: placement 2D d'après la cooccurrence  
Fujimura, Fujimura, Matsubayashi, Yamada & Okuda, WWW'08





# Nuages de tags/mots améliorés

Ajouter de l'information extraite du texte :

- pâleur pour exprimer la désuétude dans Amazon
- tags partagés en rouge dans del.icio.us
- regrouper les tags cooccurrents sur la même ligne

Hassan-Montero & Herrero-Solana, InScit'06

- optimiser l'espace vide et la proximité sémantique

Kaser & Lemire, WWW'07

- “topigraphy”: placement 2D d'après la cooccurrence

Fujimura, Fujimura, Matsubayashi, Yamada & Okuda, WWW'08



# Extraire l'information sémantique d'un texte

- analyse littéraire :

*approche philologique* : se concentrer sur le texte

Brody

- analyse du discours :

analyse arborée, graphe de cooc., projection géodésique

Brunet (Hyperbase), Viprey (Astartex)

- fouille de texte :

graphe sémantique

Grimmer (Wordmapper), Viegas et al. (IBM Many Eyes)

- traitement des langues naturelles :

désambiguïsation

Véronis (Hyperlex)

# Extraire l'information sémantique d'un texte

- analyse littéraire :

*approche philologique* : se concentrer sur le texte

Brody

- analyse du discours :

analyse arborée, graphe de cooc., projection géodésique

Brunet (Hyperbase), Viprey (Astartex)

- fouille de texte :

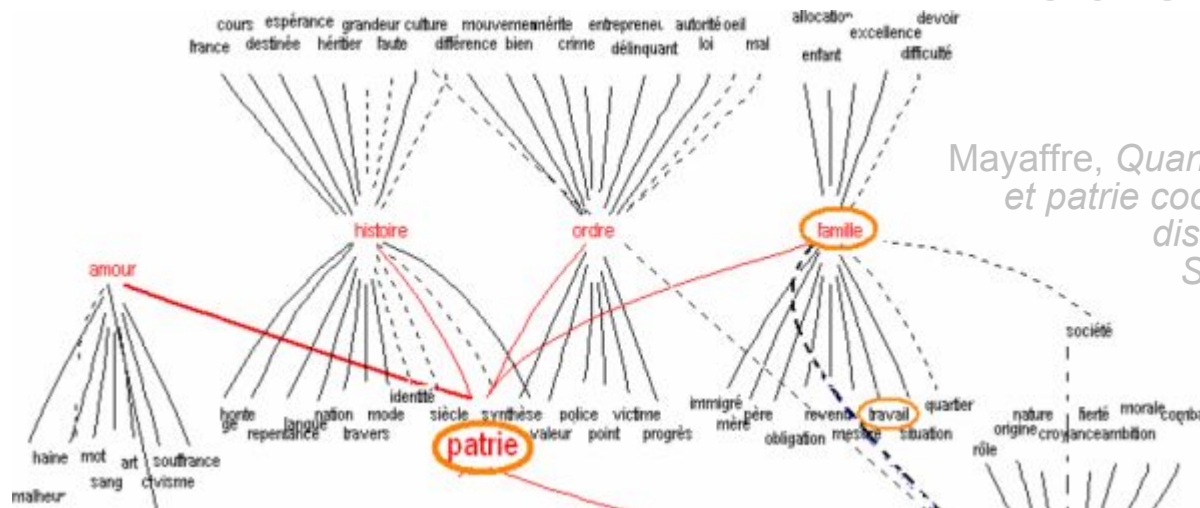
graphe sémantique

Grimmer (Wordmapper), Viegas et al. (IBM Many Eyes)

- traitement des langues naturelles :

désambiguïsation

Véronis (Hyperlex)



Mayaffre, *Quand travail, famille, et patrie cooccurrent dans le discours de Nicolas Sarkozy*, JADT'08



# Extraire l'information sémantique d'un texte

- analyse littéraire :

*approche philologique* : se concentrer sur le texte

Brody

- analyse du discours :

analyse arborée, graphe de cooc., projection géodésique

Brunet (Hyperbase), Viprey (Astartex)

- fouille de texte :

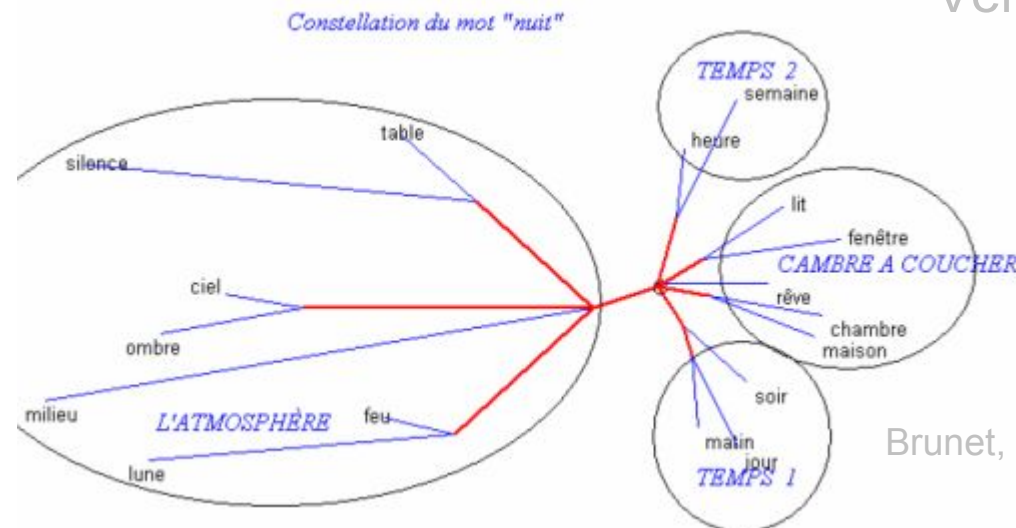
graphe sémantique

Grimmer (Wordmapper), Viegas et al. (IBM Many Eyes)

- traitement des langues naturelles :

désambiguïsation

Véronis (Hyperlex)



Brunet, *Les séquences (suite)*,  
JADT'08

# Extraire l'information sémantique d'un texte

- analyse littéraire :

*approche philologique* : se concentrer sur le texte

Brody

- analyse du discours :

analyse arborée, graphe de cooc., projection géodésique

Brunet (Hyperbase), Viprey (Astartex)

- fouille de texte :

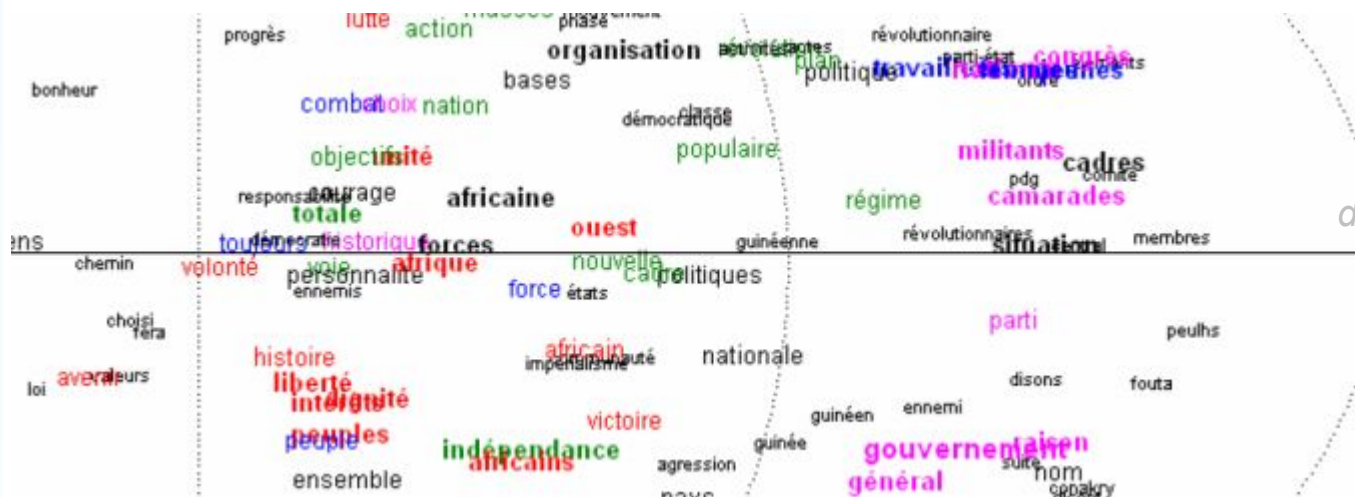
graphe sémantique

Grimmer (Wordmapper), Viegas et al. (IBM Many Eyes)

- traitement des langues naturelles :

désambiguïsation

Véronis (Hyperlex)



Barry, Viprey,  
*Approche comparative  
des résultats d'exploration  
textuelle des discours  
de deux leaders africains  
Keita et Touré,  
JADT'08*

# Extraire l'information sémantique d'un texte

- analyse littéraire :

*approche philologique* : se concentrer sur le texte

Brody

- analyse du discours :

analyse arborée, graphe de cooc., projection géodésique

Brunet (Hyperbase), Viprey (Astartex)

- fouille de texte :

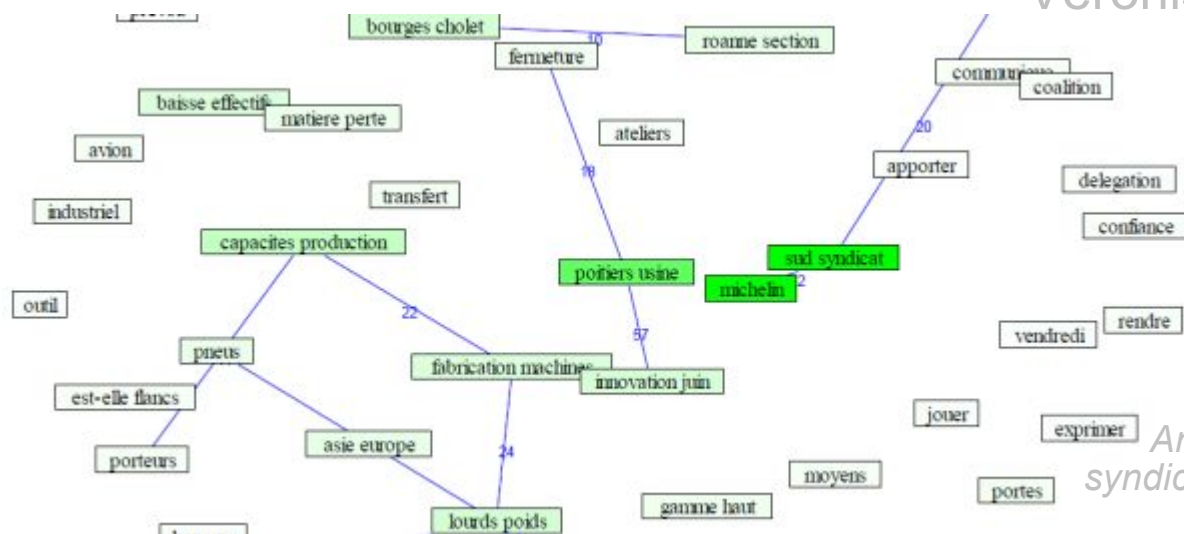
graphe sémantique

Grimmer (Wordmapper), Viegas et al. (IBM Many Eyes)

- traitement des langues naturelles :

désambiguïsation

Véronis (Hyperlex)



Peyrat-Guillard,  
*Analyse du discours  
syndical sur l'entreprise*,  
JADT'08

# Extraire l'information sémantique d'un texte

- analyse littéraire :

*approche philologique* : se concentrer sur le texte

Brody

- analyse du discours :

analyse arborée, graphe de cooc., projection géodésique

Brunet (Hyperbase), Viprey (Astartex)

- fouille de texte :

graphe sémantique

Grimmer (Wordmapper), Viegas et al. (IBM Many Eyes)

- traitement des langues naturelles :

désambiguïsation



Véronis (Hyperlex)

Visualisation PhraseNet de paroles des Beatles créé avec Many Eyes (IBM) <http://many-eyes.com>

# Extraire l'information sémantique d'un texte

- analyse littéraire :

*approche philologique* : se concentrer sur le texte

Brody

- analyse du discours :

analyse arborée, graphe de cooc., projection géodésique

Brunet (Hyperbase), Viprey (Astartex)

- fouille de texte :

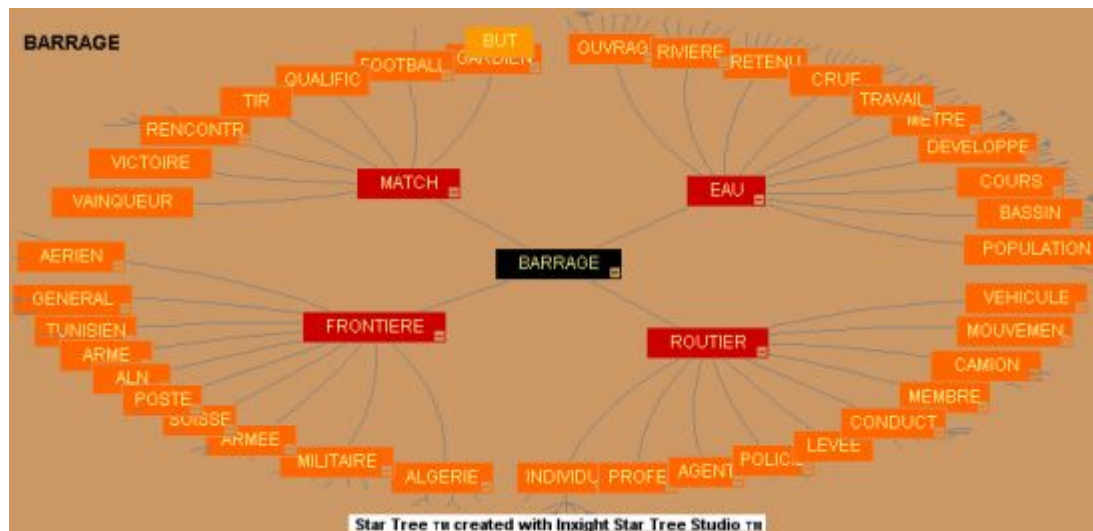
graphe sémantique

Grimmer (Wordmapper), Viegas et al. (IBM Many Eyes)

- traitement des langues naturelles :

désambiguïsation

Véronis (Hyperlex)



Désambiguïsation du mot  
"barrage".

Véronis, *HyperLex: Lexical Cartography for Information Retrieval*, 2004





# Plan

---

- Nuages de mots
- Nuages améliorés
- **Analyses plus fines**



# Analyses textuelles fines

## Lexico 3

- Laboratoire SYLED-CLA2T de Paris 3 Sorbonne Nouvelle
- Logiciel pour PC sous Windows
- Version d'évaluation gratuite
- Analyses fines, retour au texte

# Analyses textuelles fines

## Lexico 3

- Laboratoire SYLED-CLA2T de Paris 3 Sorbonne Nouvelle
- Logiciel pour PC sous Windows
- Version d'évaluation gratuite
- Analyses fines, retour au texte
- Démo sur les avis cinéma de Monique Pantel



# Analyses textuelles fines

**Treecloud**

Ce programme est une interface graphique pour Treecloud, et permet de construire le nuage arboré d'un texte. Téléchargez Treecloud, ainsi que son code source Python, sur <http://www.treecloud.org>.

Emplacement de Python (télécharger version 2.X sur [www.python.org](http://www.python.org))  
C:\Python26\python.exe

Emplacement de SplitsTree (télécharger version 4.X sur [www.splitstree.org](http://www.splitstree.org))  
C:\Treecloud\Splittree.lnk

*Le chemin de fichier ne doit pas contenir d'espace. Sinon, créez un raccourci vers SplitsTree dont le nom de fichier ne contient pas d'espace.*

Texte à visualiser : Ouvrir un fichier texte

<note=00> <critique=10233> Oh non ça moi j'ai passé l'âge, c'est trop mignon ça pour moi. Eh ben j'aurais dû aller le voir parce que j'adore Pierre Arditi  
<critique=10442> C'était le film à voir moi évidemment je l'ai pas vu.  
<critique=9279> Ca je sais pas non plus, il paraît que c'est beau.  
<critique=8454> Ah oui ça j'avoue humblement que je l'ai pas vu, hein. C'est peut-être rigolo parce que c'est des histoires plus ou moins de cul, et j'aur  
<critique=6539> Non, ça non, hein, j'avoue que je les ai pas vues les Spice Girls !  
<critique=16220> J'ai écrit un énorme papier sur elle Demi Moore et j'ai raté le film. Au lieu d'aller voir le film (y avait qu'une seule projection) je suis all  
<critique=16874> On m'a fait sauter mon avis dans la page cinéma de France Soir et c'était très gentil ce que je disais. Mocky est venu à France So  
<critique=16467> Ah oui je l'ai pas vu non plus, il paraît que c'est très joli. Je vois énormément de films, il en sort 10 par semaine, je peux pas aller les  
<critique=16464> Il y a le petit stagiaire d'ici qui l'a vu, qui m'a dit justement que c'était un style un peu à la Tarantino. J'adore Tarantino, c'est mon derni  
<critique=17511> Je l'ai pas vu celui-là. Ptêt que j'aurais dû.

Antidictionnaire : C:\Sites\GambetteLirmm\TreecloudDistribution\StoplistFrench.txt Français

Nombre de mots du nuage arboré : 100 ou nombre minimal d'occurrences pour apparaître dans le nuage arboré :

Ligne de commande :  
splitstreepath=C:\Treecloud\Splittree.lnk unit=1 nbwords=100 sepchar=critique distance=liddell  
color=chronology C:\Sites\Monique.Pantel\Pantel.txt


Distance  
 liddell  
 gmean  
 jaccard  
 dice  
 ms  
 zscore  
 hyperlex  
 chisquared  
 poissonstirling  
 loglikelihood  
 oddsratio  
 ngd  
 mi

Treecloud  
 Français  
 English

fenêtre glissante  
Taille :  
Déplacement :  
Séparateur : critique

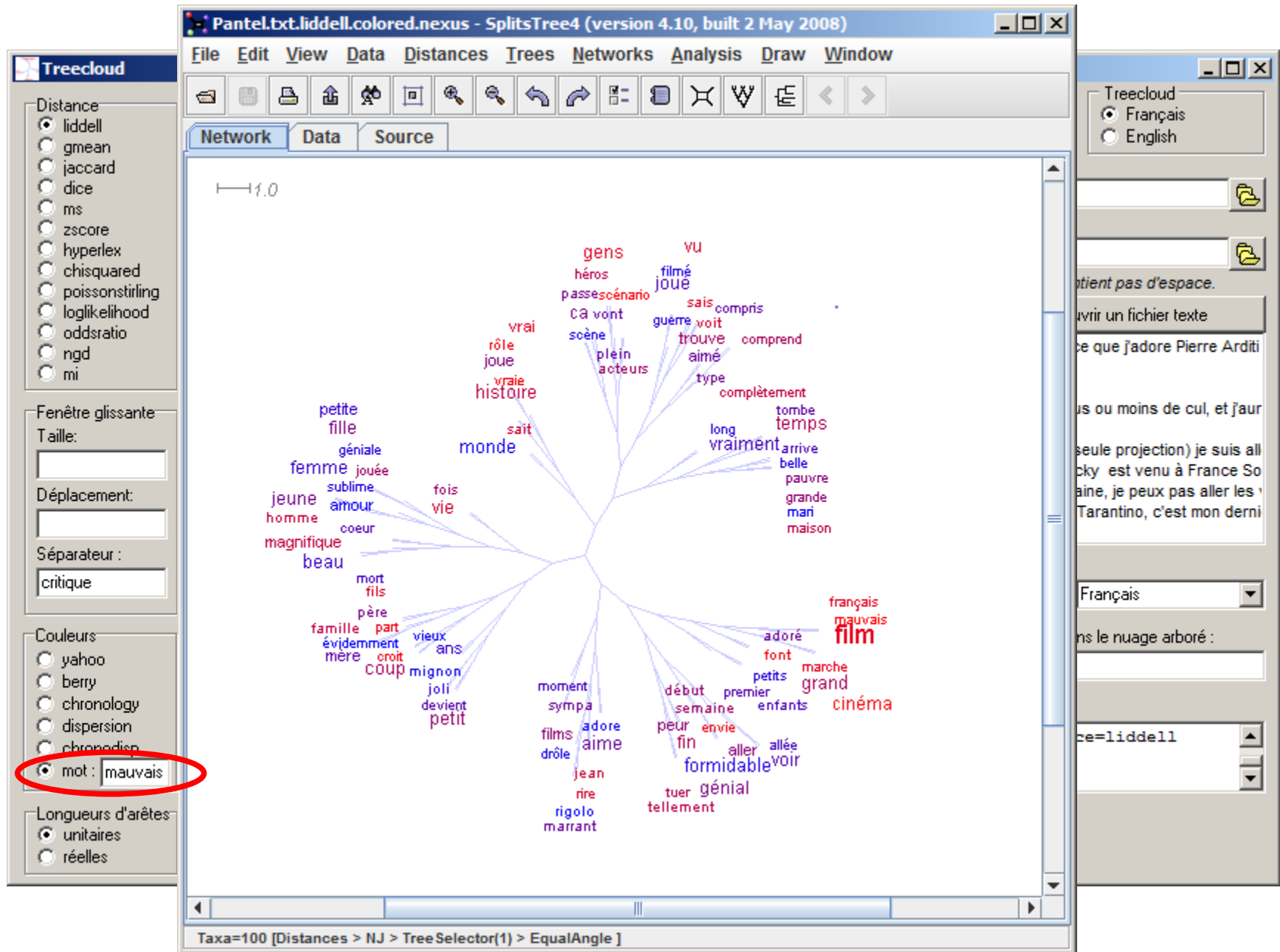
Couleurs  
 yahoo  
 berry  
 chronology  
 dispersion  
 chronodisp  
mot : cible

Longueurs d'arêtes  
 unitaires  
 réelles

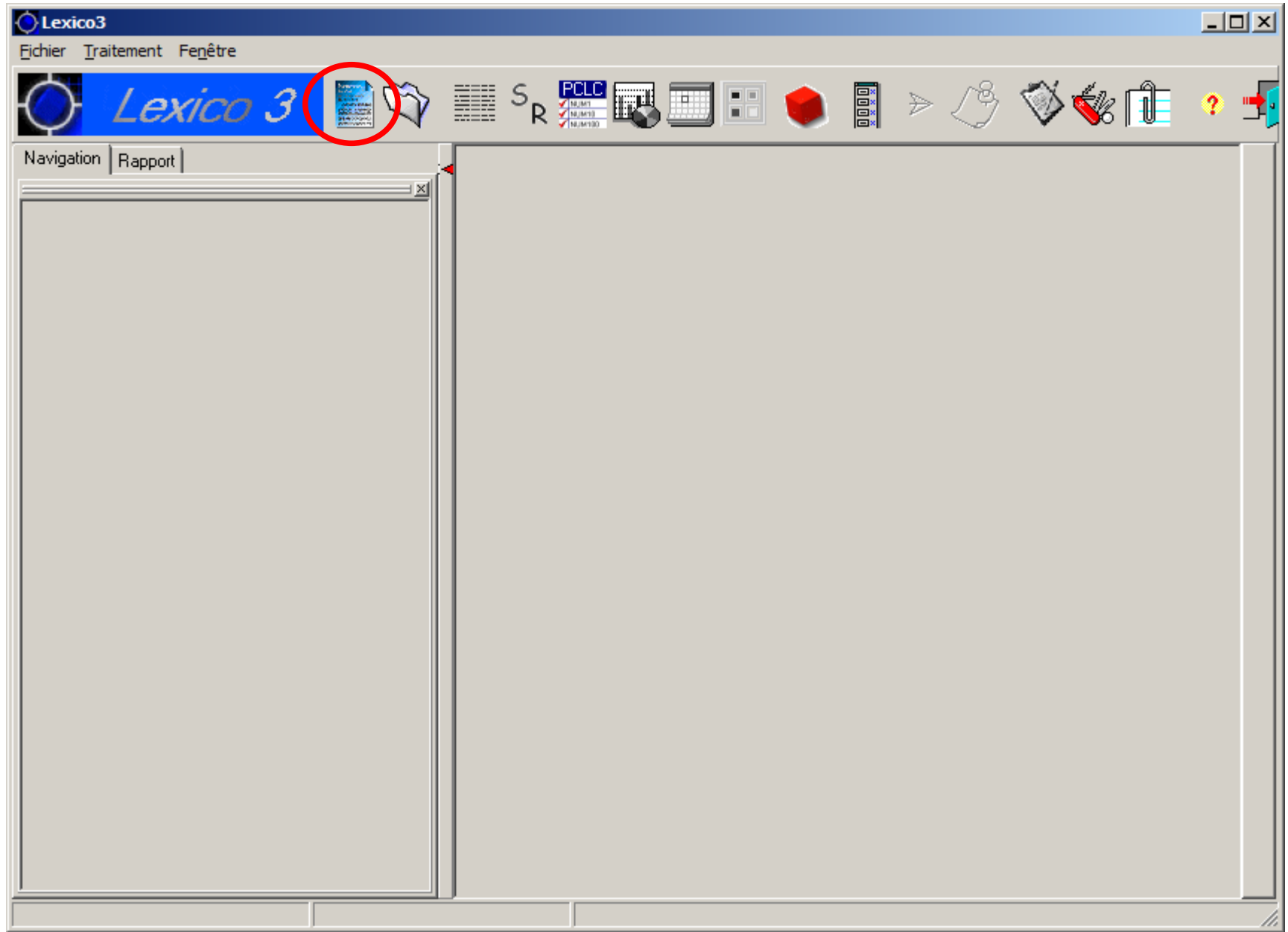
 Calcule le nuage arboré avec TreeCloud !



# Analyses textuelles fines



# Analyses textuelles fines



# Analyses textuelles fines

The screenshot shows the Lexico3 software interface. The title bar reads "Lexico3 - [TextPloreur]". The menu bar includes "Fichier", "Traitement", and "Fenêtre". The toolbar contains various icons for file operations and analysis. The interface is divided into several sections:

- Navigation**: Includes "Rapport" and "Dictionnaire" tabs.
- Recherche**: A search box with the text "Sélectionnez une couleur :".
- Tableau de fréquences**: A table showing the frequency of various forms (ordre lexicométrique).
- Texte analysé**: A large text area containing several paragraphs of text with highlighted annotations.

Formes (ordre lexicométrique)	Fréquence
est	8461
de	3727
c	3525
il	3202
un	3116
le	3055
et	2866
la	2782
qui	2393
C	2165
pas	2097
a	2038
très	1781
les	1743
à	1733
l	1596
en	1521
que	1493
une	1470
film	1454
d	1364
qu	1345
elle	1307
Et	1254
dans	1198
tout	1193
je	1186
des	1121
...	...

12657 formes

Prêt

C:\Program Files\Lexico 3\Pantel.par

Texte analysé:

**<note=00><critique=10233>** Oh non ça moi j' ai passé l' âge , c' est trop mignon ça pour moi . Eh ben j' aurais dû aller le voir parce que j' adore Pierre Arditi et puis j' y suis pas allé .

**<critique=10442>** C' était le film à voir moi évidemment je l' ai pas vu .

**<critique=9279>** Ca je sais pas non plus , il paraît que c' est beau .

**<critique=8454>** Ah oui ça j' avoue humblement que je l' ai pas vu , hein . C' est peut - être rigolo parce que c' est des histoires plus ou moins de cul , et j' aurais dû y aller . Le Tchéky Karyo qui commence paraît - il à faire une carrière américaine . Il fait le French lover . Pour lui j' aurais dû y aller , mais on peut pas aller partout , hein . On serait tout le temps dans la rue , au cinéma , et tout ça , il faut bien rester chez soi .

**<critique=6539>** Non , ça non , hein , j' avoue que je les ai pas vues les Spice Girls !

**<critique=16220>** J' ai écrit un énorme papier sur elle Demi Moore et j' ai raté le film . Au lieu d' aller voir le film ( y avait qu' une seule projection ) je suis allée à la fête du film de Resnais , où il y avait personne . Je me suis plantée totalement . Il y avait ni Resnais , ni Azéma , ni personne . Alors j' ai pas vu le film , et j' ai pas vu Resnais non plus . Il paraît , ce que j' ai entendu dire , c' est que c' est très con , hein . Mais enfin . . .

**<critique=16874>** On m' a fait sauter mon avis dans la page cinéma de France Soir et c' était très gentil ce que je disais . Mocky est venu à France Soir , il m' a beaucoup fait rire . Il est formidable ! C' est un personnage , si ils étaient tous comme ça . . . C' est vrai qu' il ressemble un peu , pas pour le talent , pas pour le genre de film , mais pour l' originalité , il ressemble à Godard . Ce sont deux

# Analyses textuelles fines

The screenshot shows the Lexico3 software interface. The main window displays a text analysis of a document. A dialog box titled "Choix de la partition" is open, showing two options: "critique" and "note". The "note" option is selected and circled in red. The background text is highlighted in yellow, indicating the selected partition. The text includes phrases like "Oh non ça moi j' ai passé l' âge , c' est trop", "moi évidemment je l' ai pas vu .", "Non , ça non , hein , j' avoue que je les ai pas vues les Spice Girls !", and "J' ai écrit un énorme papier sur elle Demi Moore et j' ai raté le film .".

Navigation Rapport Dictionnaire

Sélectionnez une couleur : [ ]

Recherche : [ ]

Formes (ordre lexicométrique)	Fréquence
est	8461
de	3727
c	3525
il	3202
un	3116
le	3055
et	2866
la	2782
qui	2393
C	2165
pas	2097
a	2038
très	1781
les	1743
à	1733
l	1596
en	1521
que	1493
une	1470
film	1454
d	1364
qu	1345
elle	1307
Et	1254
dans	1198
tout	1193
je	1186
des	1121

12657 formes

Prêt

C:\Program Files\Lexico 3\Pantel.par



# Analyses textuelles fines

The screenshot displays the Lexico3 software interface. The main window is titled "Lexico3" and contains a menu bar (Fichier, Traitement, Fenêtre), a toolbar with various icons, and a navigation pane on the left. The navigation pane is set to "Dictionnaire" and shows a search field and a table of word forms and frequencies.

Formes (ordre lexicométrique)	Fréquence
est	8461
de	3727
c	3525
il	3202
un	3116
le	3055
et	2866
la	2782
qui	2393
C	2165
pas	2097
a	2038
très	1781
les	1743
à	1733
l	1596
en	1521
que	1493
une	1470
film	1454
d	1364
qu	1345
elle	1307
Et	1254
dans	1198
tout	1193
je	1186
des	1121

At the bottom of the navigation pane, it indicates "12657 formes".

The main window is divided into two panes. The top pane, titled "Graphique de ventilation pour la partition : note", is currently empty. The bottom pane, titled "TextPloreur", displays a text analysis of a paragraph. The text is as follows:

<note=00><critique=10233> Oh non ça moi j' ai passé l' âge , c' est trop mignon ça pour moi . Eh ben j' aurais dû aller le voir parce que j' adore Pierre Arditi et puis j' y suis pas allé .

<critique=10442> C' était le film à voir moi évidemment je l' ai pas vu .

<critique=9279> Ca je sais pas non plus , il paraît que c' est beau .

<critique=8454> Ah oui ça j' avoue humblement que je l' ai pas vu , hein . C' est peut - être rigolo parce que c' est des histoires plus ou moins de cul , et j' aurais dû y aller . Le Tchéky Karyo qui commence paraît - il à faire une carrière américaine . Il fait la French lover . Donc hi i' aurais dû y aller , mais on peut

The status bar at the bottom shows "Prêt" and the file path "C:\Program Files\Lexico 3\Pantel.par".

# Analyses textuelles fines

The screenshot displays the Lexico3 software interface. On the left, a table lists word forms and their frequencies. A red arrow points from the word 'et' in the table to a graph window on the right. The graph window, titled 'Graphique de ventilation pour la partition : note', shows a large empty area with the text 'glisser-déposer' (drag-and-drop) in red. Below the graph, there are radio buttons for 'Fréquence', 'relatives', and 'absolues', with 'relatives' selected. At the bottom, a 'TextPloreur' window shows a text snippet with highlighted words and their associated scores.

Formes (ordre lexicométrique)	Fréquence
est	8461
de	3727
c	3525
il	3202
un	3116
le	3055
et	2866
la	2782
qui	2393
C	2165
pas	2097
a	2038
très	1781
les	1743
à	1733
l	1596
en	1521
que	1493
une	1470
film	1454
d	1364
qu	1345
elle	1307
Et	1254
dans	1198
tout	1193
je	1186
des	1121

TextPloreur content:

<note=00><critique=10233> Oh non ça moi j ' ai passé l ' âge , c ' est trop mignon ça pour moi . Eh ben j ' aurais dû aller le voir parce que j ' adore Pierre Arditi et puis j ' y suis pas allé .

<critique=10442> C ' était le film à voir moi évidemment je l ' ai pas vu .

<critique=9279> Ca je sais pas non plus , il paraît que c ' est beau .

<critique=8454> Ah oui ça j ' avoue humblement que je l ' ai pas vu , hein . C ' est peut - être rigolo parce que c ' est des histoires plus ou moins de cul , et j ' aurais dû y aller . Le Tchéky Karyo qui commence paraît - il à faire une carrière américaine . Il fait la French lover . Douc hi i ' aurais dû aller , mais on peut

# Analyses textuelles fines

Lexico3

Fichier Traitement Fenêtre

Navigation Rapport Dictionnaire

Sélectionnez une couleur : [blue]

Recherche :

Formes (ordre lexicométrique)	Fréquence
sont	508
parce	473
au	464
sur	460
sa	458
ben	427
là	427
Eh	424
quand	419
aussi	410
suis	403
m	401
faire	360
n	335
dire	321
Je	317
Elle	297
voir	295
J	294
formidable	293
deux	284
ne	284
dit	281
petit	278
On	268
bon	268
coup	266
me	266

12657 formes

Prêt

C:\Program Files\Lexico 3\Pantel.par

Graphique de ventilation pour la partition : note

Fréquences relatives (X 10 000)

Fréquence  relatives  absolues Spécificités Zoom + Zoom - Effac

TextPloreur

<note=00><critique=10233> Oh non ça moi j' ai passé l' âge , c' est trop mignon ça pour moi . Eh ben j' aurais dû aller le voir parce que j' adore Pierre Ardit et puis j' y suis pas allé .

<critique=10442> C' était le film à voir moi évidemment je l' ai pas vu .

<critique=9279> Ca je sais pas non plus , il paraît que c' est beau .

<critique=8454> Ah oui ça j' avoue humblement que je l' ai pas vu , hein . C' est peut - être rigolo parce que c' est des histoires plus ou moins de cul , et j' aurais dû y aller . Le Tchéky Karyo qui commence paraît - il à faire une carrière américaine . Il fait la French lover . Doux hi j' aurais dû y aller , mais on peut

Principales c...

Feuille 1

# Analyses textuelles fines

Lexico3

Fichier Traitement Fenêtre

Navigation Rapport Dictionnaire

Sélectionnez une couleur : [bleu]

Recherche :

Formes (ordre lexicométrique)	Fréquence
jeune	195
génial	190
ont	189
cette	188
puis	186
vu	186
si	185
veut	182
fin	181
mon	180
alors	176
vie	173
nous	171
mal	169
peur	169
joue	167
où	165
Oh	163
ans	161
sans	161
Ah	160
petite	157
amour	155
films	155
voilà	154
enfin	153
aller	148
aux	142

12657 formes

Graphique de ventilation pour la partition : note

Spécificités

00 02 05 06 07 08 09 10 11 12 13 14 15 16 17 18

■ formidable  
■ amour

Fréquence  relatives  absolues  Spécificités Zoom + Zoom - Effac

TextPloreur

<note=00><critique=10233> Oh non ça moi j ' ai passé l ' âge , c ' est trop mignon ça pour moi . Eh ben j ' aurais dû aller le voir parce que j ' adore Pierre Arditi et puis j ' y suis pas allé .

<critique=10442> C ' était le film à voir moi évidemment je l ' ai pas vu .

<critique=9279> Ca je sais pas non plus , il paraît que c ' est beau .

<critique=8454> Ah oui ça j ' avoue humblement que je l ' ai pas vu , hein . C ' est peut - être rigolo parce que c ' est des histoires plus ou moins de cul , et j ' aurais dû y aller . Le Tchéky Karyo qui commence paraît - il à faire une carrière américaine . Il fait la French lover . Donc lui j ' aurais dû y aller , mais on peut

Prêt

C:\Program Files\Lexico 3\Pantel.par

# Analyses textuelles fines

Lexico3

Fichier Traitement Fenêtre

note

Navigation Rapport Dictionnaire

Sélectionnez une couleur : [bleu]

Recherche :

Formes (ordre lexicométrique)	Fréquence
jeune	195
génial	190
ont	189
cette	188
puis	186
vu	186
si	185
veut	182
fin	181
mon	180
alors	176
vie	173
nous	171
mal	169
peur	169
joue	167
où	165
Oh	163
ans	161
sans	161
Ah	160
petite	157
amour	155
films	155
voilà	154
enfin	153
aller	148
aux	142

12657 formes

Graphique de ventilation pour la partition : note

Catégorie	formidable	amour
00	-2	0
02	0	0
05	-2	0
06	-1	-1
07	-1	-1
08	-1	-1
09	-1	-1
10	0	0
11	-1	0
12	0	0
13	0	0
14	0	0
15	0	0
16	0	0
17	22	-4
18	0	-4

Fréquence  relatives  absolues Spécificités Zoom + Zoom - Effac

TextPloreur

<note=00><critique=10233> Oh non ça moi j ' ai passé l ' âge , c ' est trop mignon ça pour moi . Eh ben j ' aurais dû aller le voir parce que j ' adore Pierre Arditi et puis j ' y suis pas allé .

<critique=10442> C ' était le film à voir moi évidemment je l ' ai pas vu .

<critique=9279> Ca je sais pas non plus , il paraît que c ' est beau .

<critique=8454> Ah oui ça j ' avoue humblement que je l ' ai pas vu , hein . C ' est peut - être rigolo parce que c ' est des histoires plus ou moins de cul , et j ' aurais dû y aller . Le Tchéky Karyo qui commence paraît - il à faire une carrière américaine . Il fait la French lover . Donc lui j ' aurais dû y aller , mais on peut

Prêt

C:\Program Files\Lexico 3\Pantel.par

# Analyses textuelles fines

Lexico3 - [Principales caractéristiques de la partition : note]

Fichier Traitement Fenêtre

Navigation Rapport Dictionnaire

Sélectionnez une couleur :

Recherche :

Formes (ordre lexicométrique)	Fréquence
jeune	195
génial	190
ont	189
cette	188
puis	186
vu	186
si	185
veut	182
fin	181
mon	180
alors	176
vie	173
nous	171
mal	169
peur	169
joue	167
où	165
Oh	163
ans	161
sans	161
Ah	160
petite	157
amour	155
films	155
voilà	154
enfin	153
aller	148
aux	142

12657 formes

Princ. Caract. Lexicométriques

		Nombre d'occurrences:	163459	Nombre de formes:	12644	
		Nombre d'hapax:	6396	Fréquence maximale:	8461	
	Num	Partie	Occurenc	Formes	Hapax	Fmax
✓	1	00	2998	760	479	92
✓	2	02	1480	592	426	63
✓	3	05	3262	1027	687	171
✓	4	06	9344	2169	1387	450
✓	5	07	13212	2702	1685	631
✓	6	08	8759	1963	1266	388
✓	7	09	7128	1770	1146	293
✓	8	10	5253	1462	961	233
✓	9	11	8124	1979	1297	373
✓	10	12	10630	2286	1442	514
✓	11	13	6078	1596	1071	330
✓	12	14	3508	1004	675	199
✓	13	15	15564	2919	1832	858
✓	14	16	25833	4084	2478	1471
✓	15	17	40004	5332	2962	2257
✓	16	18	2282	699	443	138

Spécifs

Sp Diac.

AFC

CAH

Pareto

Ac

Sp. Tot

Feuille 1

Prêt

C:\Program Files\Lexico 3\Pantel.par

# Analyses textuelles fines

The screenshot shows the Lexico 3 software interface. A dialog box titled "Paramétrage de l'AFC" is open, allowing configuration of the AFC analysis. The dialog includes fields for "Titre de l'analyse" (set to "Pantel : note"), "NI" (1336), "NJ" (16), "NF" (15), "NI2" (0), "NJ2" (0), and "FreqMin" (10). There are "OK" and "Annuler" buttons at the bottom of the dialog.

In the background, a table displays the results of the analysis. The table has columns for "enc", "Formes", "Hapax", and "Fmax". The "enc" column contains a list of words, and the other columns contain their respective counts. The "enc" column is partially obscured by the dialog box.

enc	Formes	Hapax	Fmax
12644 8461			
760	479	92	
592	426	63	
1027	687	171	
2169	1387	450	
2702	1685	631	
1963	1266	388	
1770	1146	293	
1462	961	233	
1979	1297	373	
10630	2286	1442	514
6078	1596	1071	330
3508	1004	675	199
15564	2919	1832	858
25833	4084	2478	1471
40004	5332	2962	2257
2282	699	443	138

On the right side of the interface, there is a vertical toolbar with several buttons: "Spécifs", "Sp. Diac.", "AFC" (circled in red), "LAF", "Pareto", "Ac", and "Sp. Tot".

The status bar at the bottom shows "Prêt" and the file path "C:\Program Files\Lexico 3\Pantel.par".

# Analyses textuelles fines

The screenshot displays the Lexico 3 software interface. On the left, a table lists words and their frequencies. The main area shows a correspondence analysis plot with axes and numbered points. A red box highlights a cluster of points (02, 05, 06, 07, 08) in the bottom-right quadrant. A red circle highlights the 'Spécifs' button in the right-hand toolbar. Red text at the bottom of the plot area reads 'sélection puis clic sur Spécifs'.

**Table: Formes (ordre lexicométrique) vs Fréquence**

Formes (ordre lexicométrique)	Fréquence
jeune	195
génial	190
ont	189
cette	188
puis	186
vu	186
si	185
veut	182
fin	181
mon	180
alors	176
vie	173
nous	171
mal	169
peur	169
joue	167
où	165
Oh	163
ans	161
sans	161
Ah	160
petite	157
amour	155
films	155
voilà	154
enfin	153
aller	148
aux	142

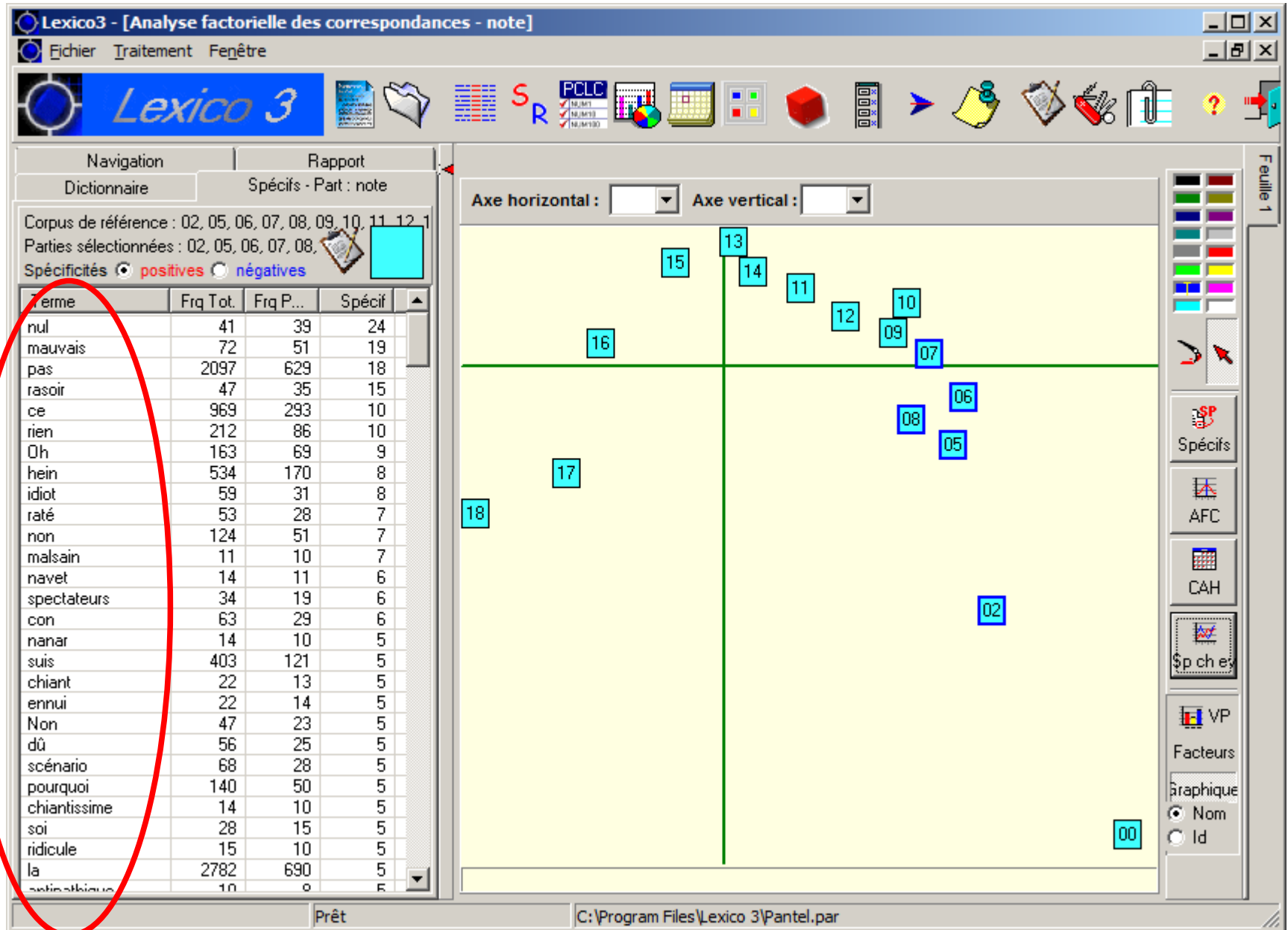
**Plot Labels:** 00, 02, 05, 06, 07, 08, 09, 10, 11, 12, 13, 14, 15, 16, 17, 18

**Toolbar Buttons:** Spécifs, AFC, CAH, Sp ch ex, VP, Facteurs, Graphique, Nom, Id

**Text:** sélection puis clic sur Spécifs



# Analyses textuelles fines



# Analyses textuelles fines

The screenshot shows the Lexico3 software interface. The title bar reads "Lexico3 - [Concordances]". The menu bar includes "Fichier", "Traitement", and "Fenêtre". The toolbar contains various icons, with a red circle highlighting the "S" icon (representing search or concordance). The main window is divided into several sections:

- Navigation:** Includes "Dictionnaire" and "Rapport".
- Forme:** A text input field.
- Tri:** A dropdown menu set to "Aucun".
- Regroupement:** A dropdown menu set to "<Aucun>".
- Type de documentation:** A dropdown menu set to "Concordance".
- Spécificités:** Radio buttons for "positives" (selected) and "négatives".
- Table:** A table with columns "Terme", "Frq Tot.", "Frq P...", and "Spécif".
- Text Area:** A large text area displaying concordance results for the word "rasoir".
- Footer:** "Nombre de contextes : 47" and "retour au texte" (highlighted in red).

Terme	Frq Tot.	Frq P...	Spécif
nul	41	39	24
mauvais	72	51	19
pas	2097	629	18
rasoir	47	35	15
ce	969	293	10
rien	212	86	10
Oh	163	69	9
hein	534	170	8
idiot	59	31	8
raté	53	28	7
non	124	51	7
malsain	11	10	7
navet	14	11	6
spectateurs	34	19	6
con	63	29	6
nanar	14	10	5
suis	403	121	5
chiant	22	13	5
ennui	22	14	5
Non	47	23	5
dû	56	25	5
scénario	68	28	5
pourquoi	140	50	5
chiantissime	14	10	5
soi	28	15	5
ridicule	15	10	5
la	2782	690	5
esthétique	10	0	5

Text area content (partial):  
nul . C ' est complètement nul , c ' est raser . Et les crit  
l ' ai écrit dans France Soir . C ' est raser . Mais là Jac  
isney on y va systématiquement . C ' est raser ! C ' est ins  
e n ' ira le voir . Ah non , il est trop raser ! C ' est pas  
possible . Et lui Théo Angelopoulos est raser , c ' est un  
s . 5 minutes après il dormait . C ' est raser , c ' est ras  
ès il dormait . C ' est raser , c ' est raser , c ' est len  
as - là . C ' est très mauvais . C ' est raser et c ' est ma  
n autre film d ' horreur et c ' est très raser . Un couple d  
Il reprend sa boutique de barbier et son raser . Il égorge l  
ans le monde . C ' est poussif , c ' est raser , et moi du c  
guerres , et puis tout ça . . . C ' est raser . C ' est idi  
rce Brosnan qui est James Bond . C ' est raser , c ' est hal  
hein . Mais le pire est arrivé . C ' est raser ! Mais pourqu  
dée à cette époque - là ? C ' est d ' un raser . Alors c ' e  
s qui le regarde . Non mais là , c ' est raser , hein , il f  
bon film . Vous voulez dire de l ' Opé - raser ? C ' est ton  
je lui ai demandé : Est - ce que c ' est raser ? Mais oui ,  
soir ? Mais oui , Monique , c ' est très raser , il me dit !  
à poil , et tout ça . Bon . C ' est très raser , c ' est pas  
abbé Pierre et Cécilia Sarkozy . C ' est raser , ce film , c  
ozy . C ' est raser , ce film , c ' est raser ! J ' étais t  
près un best - seller , et alors c ' est raser ! Je me le su  
de la peur qui ravira que lui . C ' est raser , hein . Enfi  
, on n ' y comprend rien et c ' est très raser . Matt Damon  
ça le rend malade . Le boxeur est super raser . C ' est enc