

JADT 2010 - 11/06/2010
Rome

***Utilisation de la visualisation
en nuage arboré
pour l'analyse littéraire***

Delphine Amstutz
(CELLF – Université Paris-Sorbonne Paris 4 / CNRS)
Philippe Gambette
(LIRMM – Université Montpellier 2 / CNRS)



Plan

- Le nuage arboré : occurrence + cooccurrence
- Des couleurs pour guider la lecture
- Processus de construction
- Implémentations
- Emplois du nuage arboré
- Illustration sur *Cinna* et *Othon*
- Perspectives

Plan

- Le nuage arboré : occurrence + cooccurrence
- Des couleurs pour guider la lecture
- Processus de construction
- Implémentations
- Emplois du nuage arboré
- Illustration sur *Cinna* et *Othon*
- Perspectives

Nuages de tags

- Construits depuis un ensemble de tags
- Taille de police liée à la fréquence



Ce qui est habituellement cité comme
le premier nuage de tags, dans
Microserfs de D. Coupland,
HarperCollins, Toronto, 1995

Extraire l'information sémantique d'un texte

- analyse arborée Brunet (Hyperbase)
- graphe de cooccurrence Brunet (Hyperbase)
- graphe sémantique Grimmer (Wordmapper)
- lexicogramme récursif Martinez (Coocs)
- désambiguïisation lexicale Véronis (Hyperlex)
- réseau Phrasenet Viegas et al. (IBM Many Eyes)
- projection géodésique Viprey (Astartex)

Extraire l'information sémantique d'un texte

- analyse arborée
- graphe de cooccurrence
- graphe sémantique
- lexicogramme récursif
- désambiguïisation lexicale
- réseau Phrasenet
- projection géodésique

Brunet (Hyperbase)

Brunet (Hyperbase)

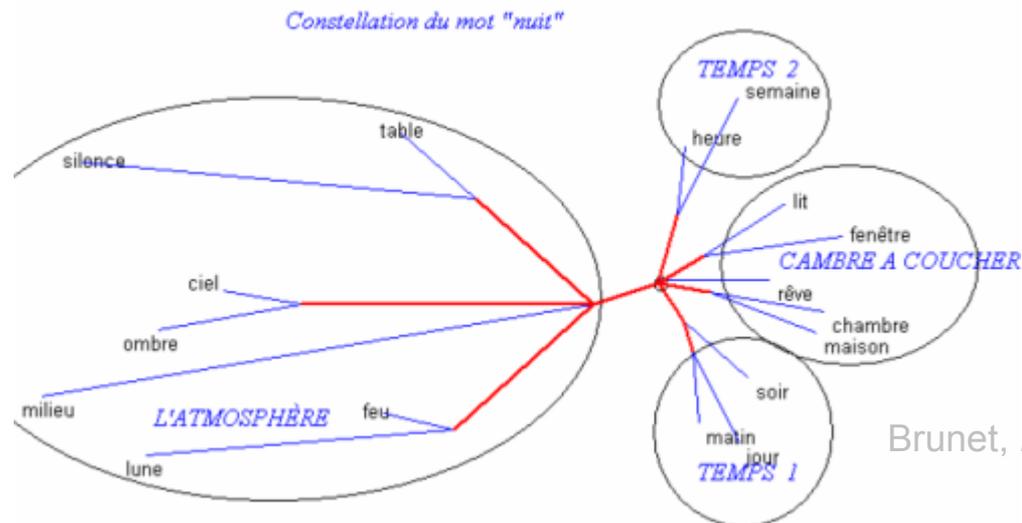
Grimmer (Wordmapper)

Martinez (Coocs)

Véronis (Hyperlex)

Viegas et al. (IBM Many Eyes)

Viprey (Astartex)



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

Extraire l'information sémantique d'un texte

- analyse arborée
- graphe de cooccurrence
- graphe sémantique
- lexicogramme récursif
- désambiguïisation lexicale
- réseau Phrasenet
- projection géodésique

Brunet (Hyperbase)

Brunet (Hyperbase)

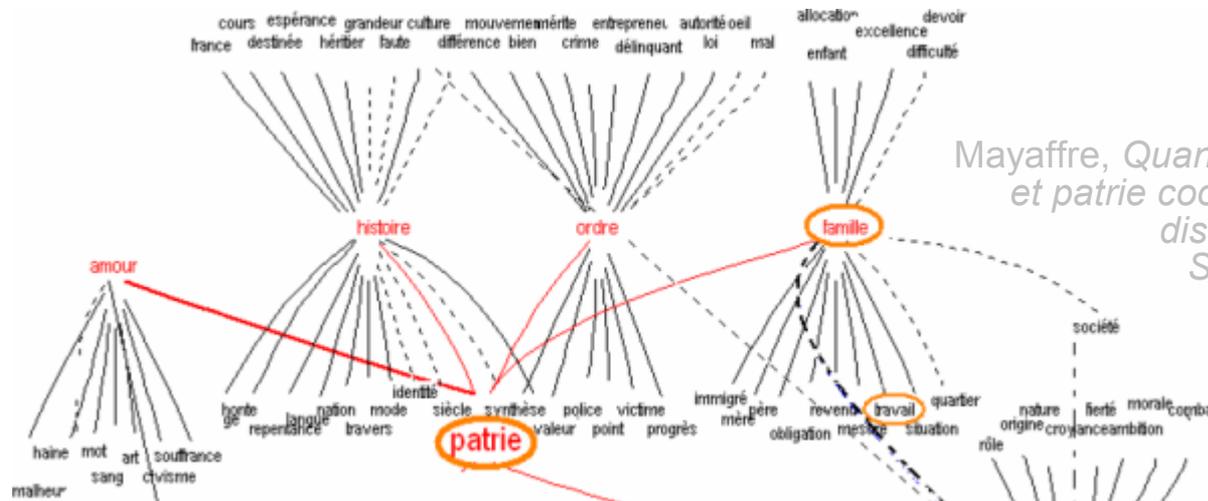
Grimmer (Wordmapper)

Martinez (Coocs)

Véronis (Hyperlex)

Viegas et al. (IBM Many Eyes)

Viprey (Astartex)



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

Extraire l'information sémantique d'un texte

- analyse arborée
- graphe de cooccurrence
- graphe sémantique
- lexicogramme récursif
- désambiguïisation lexicale
- réseau Phrasenet
- projection géodésique

Brunet (Hyperbase)

Brunet (Hyperbase)

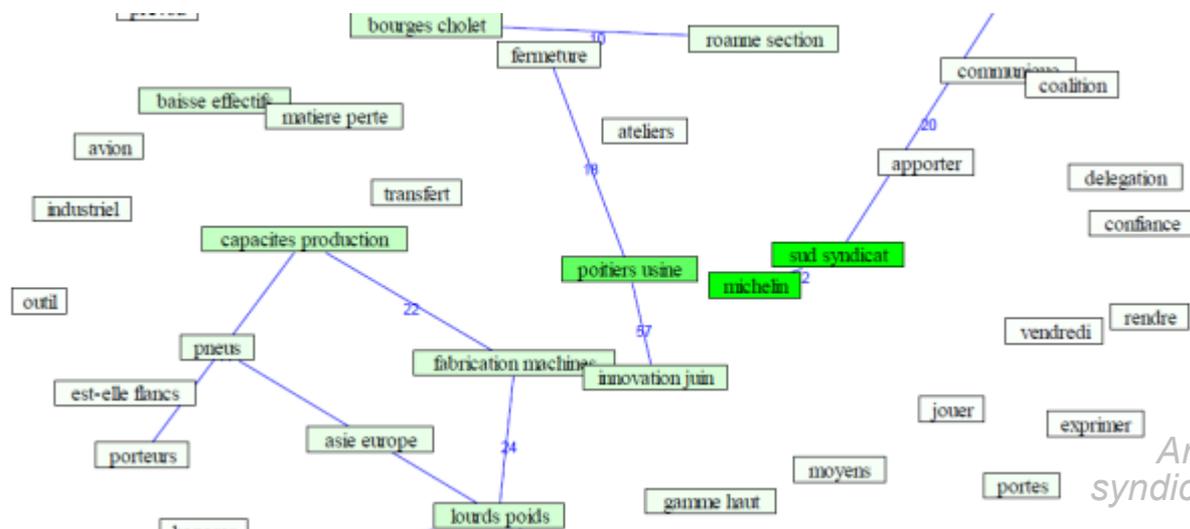
Grimmer (Wordmapper)

Martinez (Coocs)

Véronis (Hyperlex)

Viegas et al. (IBM Many Eyes)

Viprey (Astartex)



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

Extraire l'information sémantique d'un texte

- analyse arborée
- graphe de cooccurrence
- graphe sémantique
- lexicogramme récursif
- désambiguïisation lexicale
- réseau Phrasenet
- projection géodésique

Brunet (Hyperbase)

Brunet (Hyperbase)

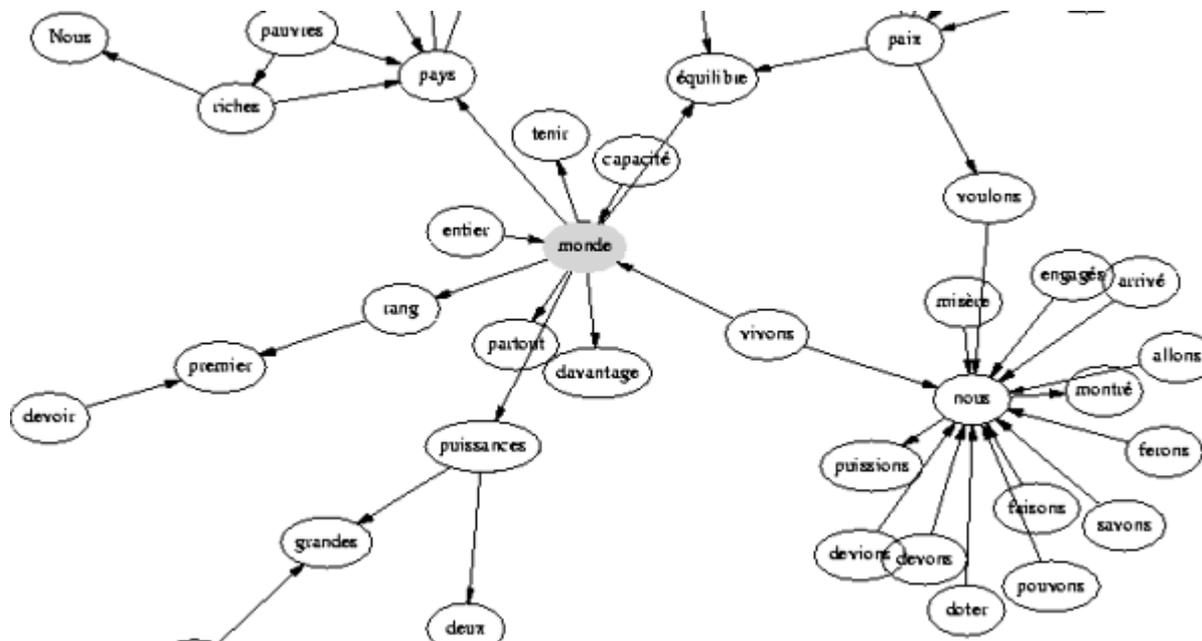
Grimmer (Wordmapper)

Martinez (Coocs)

Véronis (Hyperlex)

Viegas et al. (IBM Many Eyes)

Viprey (Astartex)



Leblanc, Martinez
*L'analyse contrastive des
réseaux de cooccurrence*
JADT 2006.

Extraire l'information sémantique d'un texte

- analyse arborée
- graphe de cooccurrence
- graphe sémantique
- lexicogramme récursif
- désambiguïisation lexicale
- réseau Phrasenet
- projection géodésique

Brunet (Hyperbase)

Brunet (Hyperbase)

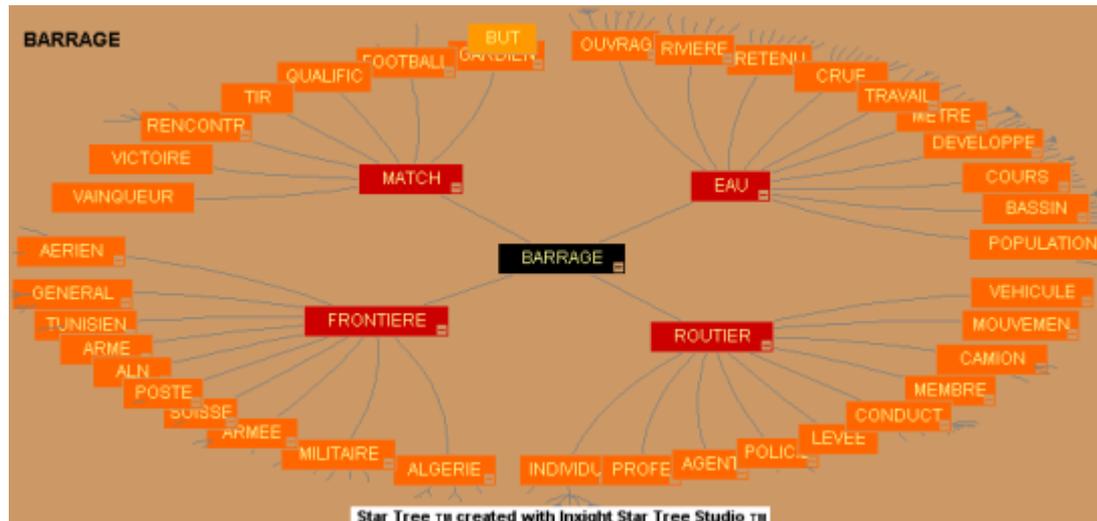
Grimmer (Wordmapper)

Martinez (Coocs)

Véronis (Hyperlex)

Viegas et al. (IBM Many Eyes)

Viprey (Astartex)



Désambiguïisation du mot
"barrage".

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

Plan

- Le nuage arboré : occurrence + cooccurrence
- **Des couleurs pour guider la lecture**
- Processus de construction
- Implémentations
- Emplois du nuage arboré
- Illustration sur *Cinna* et *Othon*
- Perspectives

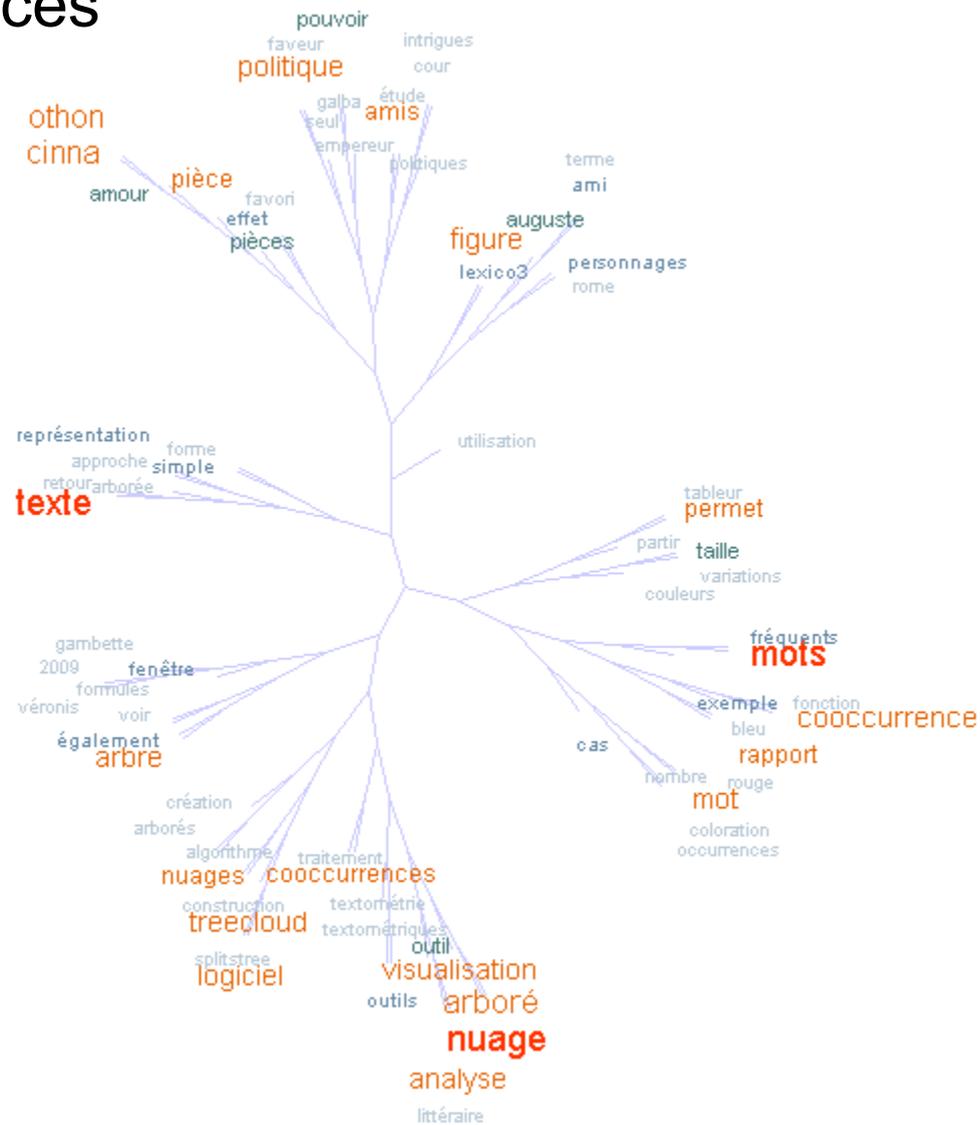
Des couleurs pour guider la lecture

- coloration selon les fréquences
- coloration chronologique
- coloration de la dispersion
- coloration ciblée sur un mot
- coloration grammaticale

Des couleurs pour guider la lecture

- coloration selon les fréquences
- coloration chronologique
- coloration de la dispersion
- coloration ciblée sur un mot
- coloration grammaticale

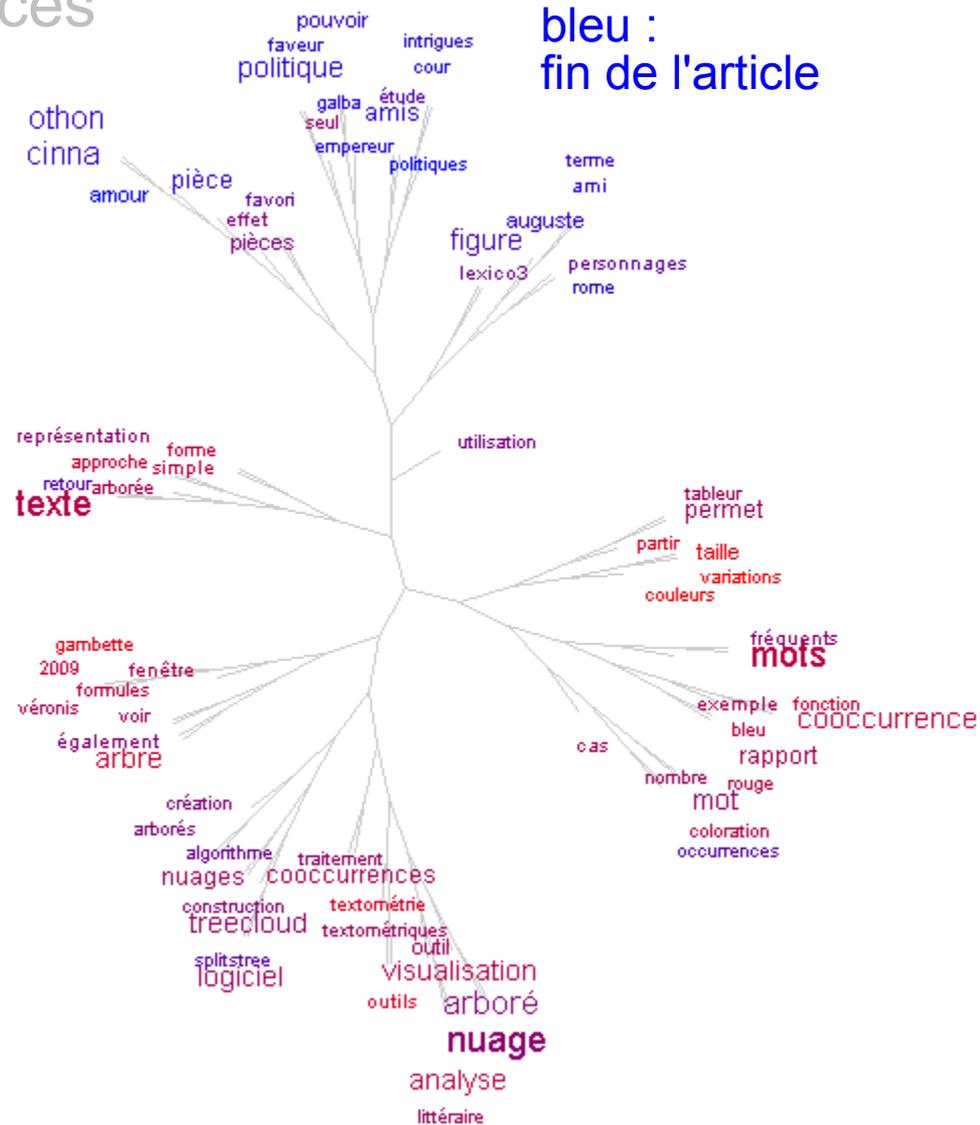
Nuage arboré des mots apparaissant 5 fois ou plus dans cet article JADT, distance Liddell, fenêtre de 20 mots, coloration Yahoo



Des couleurs pour guider la lecture

- coloration selon les fréquences
- coloration chronologique
- coloration de la dispersion
- coloration ciblée sur un mot
- coloration grammaticale

rouge :
début de l'article
bleu :
fin de l'article

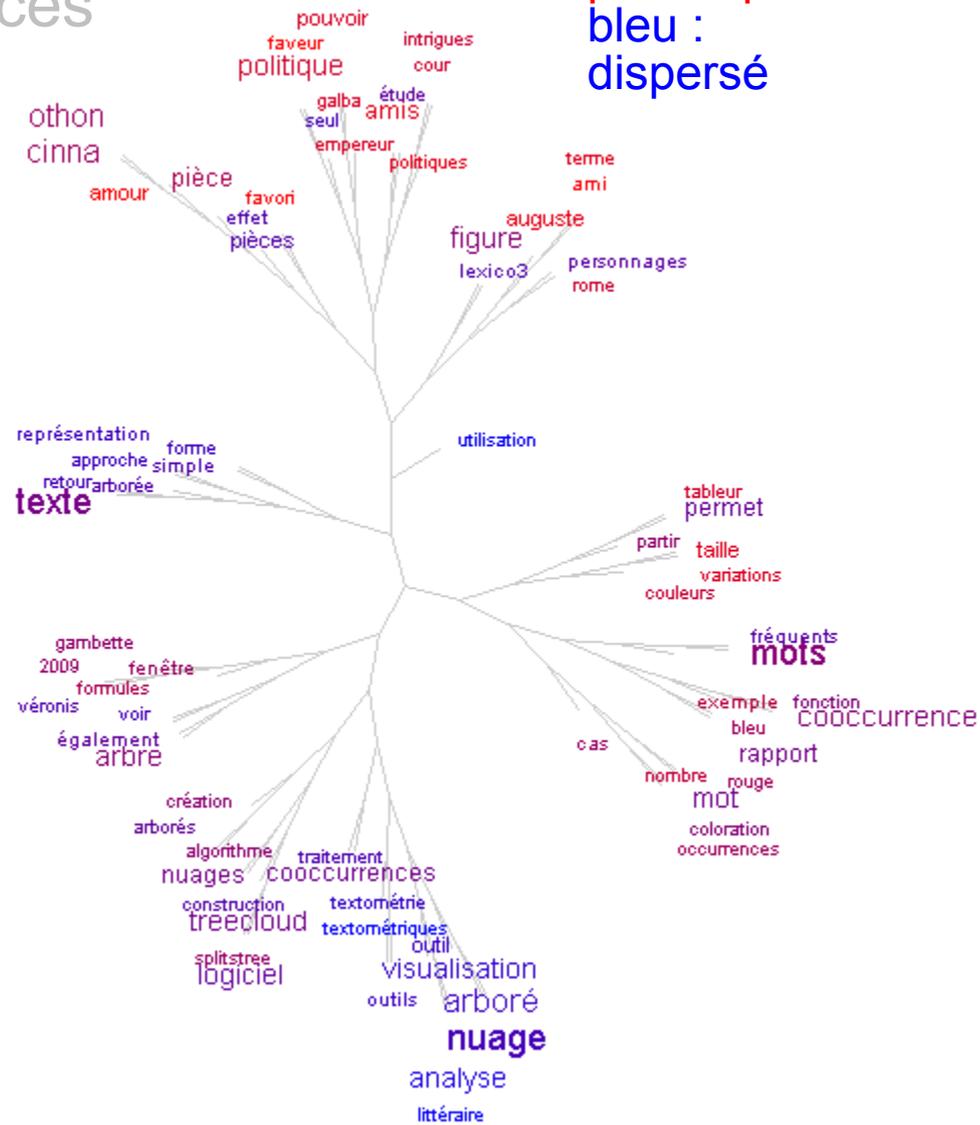


Nuage arboré des mots
apparaissant 5 fois ou
plus dans cet article
JADT, distance Liddell,
fenêtre de 20 mots,
coloration chronologique

Des couleurs pour guider la lecture

- coloration selon les fréquences
- coloration chronologique
- coloration de la dispersion
- coloration ciblée sur un mot
- coloration grammaticale

rouge :
peu dispersé
bleu :
dispersé



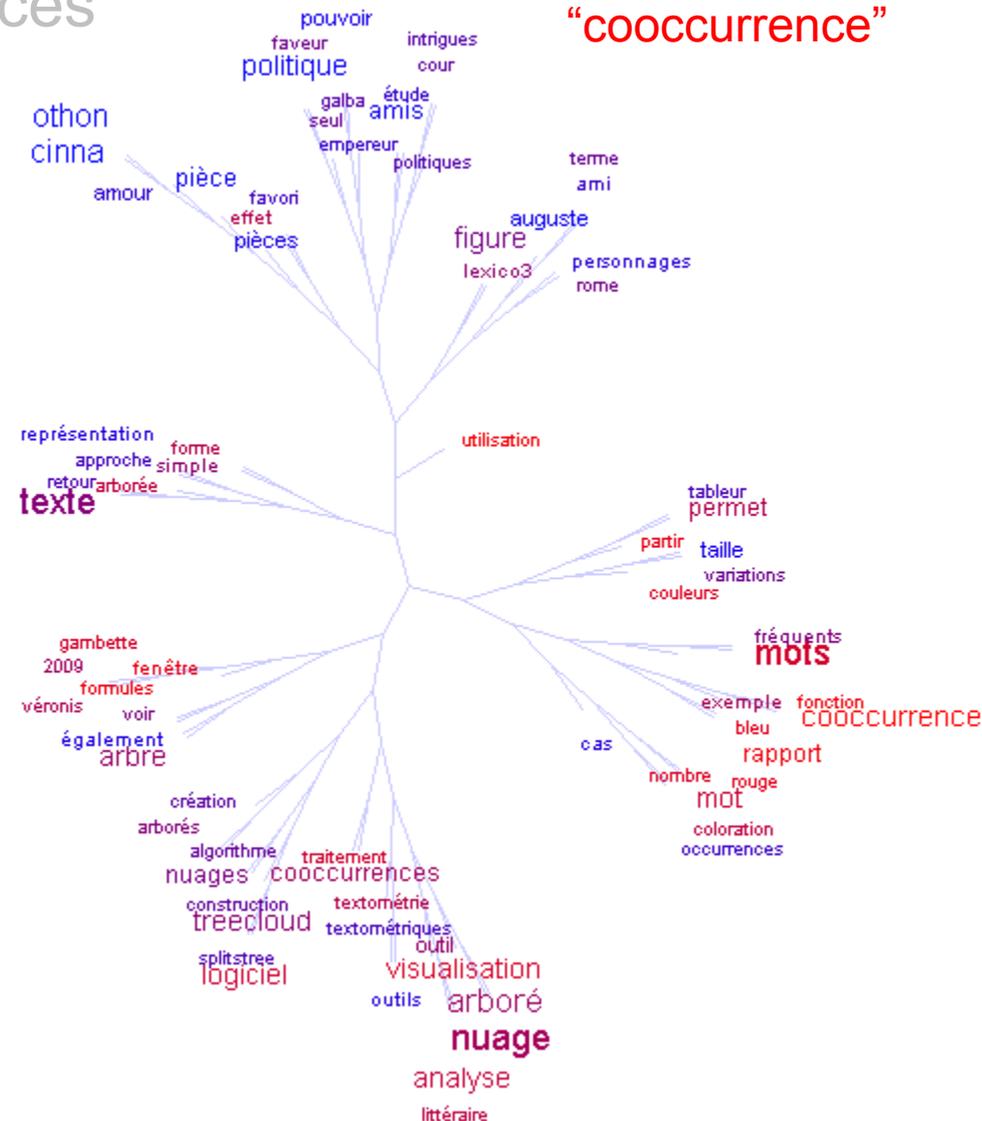
Nuage arboré des mots
apparaissant 5 fois ou
plus dans cet article
JADT, distance Liddell,
fenêtre de 20 mots,
coloration dispersion

Des couleurs pour guider la lecture

- coloration selon les fréquences
- coloration chronologique
- coloration de la dispersion
- coloration ciblée sur un mot
- coloration grammaticale

Nuage arboré des mots apparaissant 5 fois ou plus dans cet article JADT, distance Liddell, fenêtre de 20 mots, coloration ciblée sur le mot “cooccurrence”

rouge :
cooccurents de
“cooccurrence”



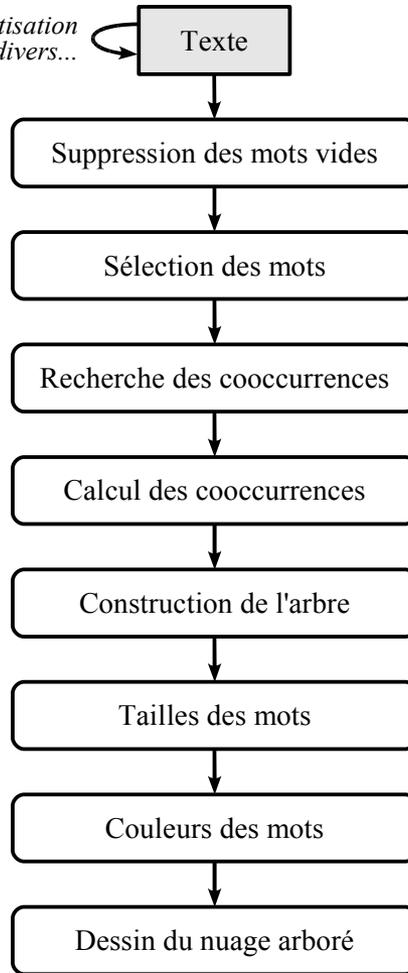
Plan

- Le nuage arboré : occurrence + cooccurrence
- Des couleurs pour guider la lecture
- **Processus de construction**
- Implémentations
- Emplois du nuage arboré
- Illustration sur *Cinna* et *Othon*
- Perspectives

Processus de construction

Import/export

*Concordance d'un mot, lemmatisation
ou remplacements divers...*

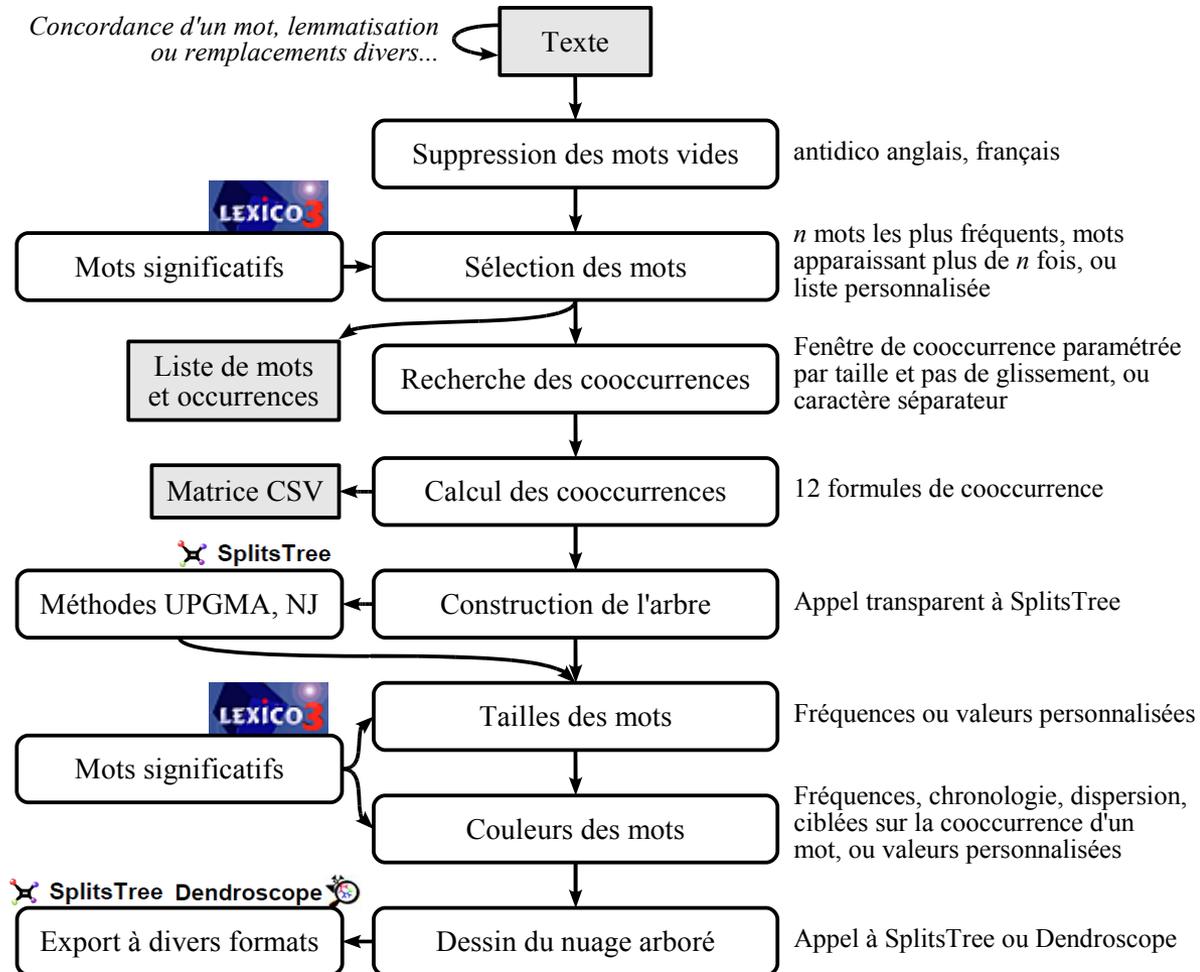


Processus de construction

Import/export

Proposé dans TreeCloud

Concordance d'un mot, lemmatisation
ou remplacements divers...



Construction – les cooccurrences

De nombreuses formules de distance sémantique utilisent la cooccurrence.

Texte

fenêtre
glissante S

→
*Pas de
glissement s*

┌
└
largeur w

matrices de cooccurrence

O_{11} , O_{12} , O_{21} , O_{22}

	$v \in S$	$v \notin S$
$u \in S$	O_{11}	O_{12}
$u \notin S$	O_{21}	O_{22}



matrice de dissimilarité
sémantique

*chi squared, mutual
information, liddel, dice,
jaccard, gmean, hyperlex,
minimum sensitivity, odds
ratio, zscore, log likelihood,
poisson-stirling...*

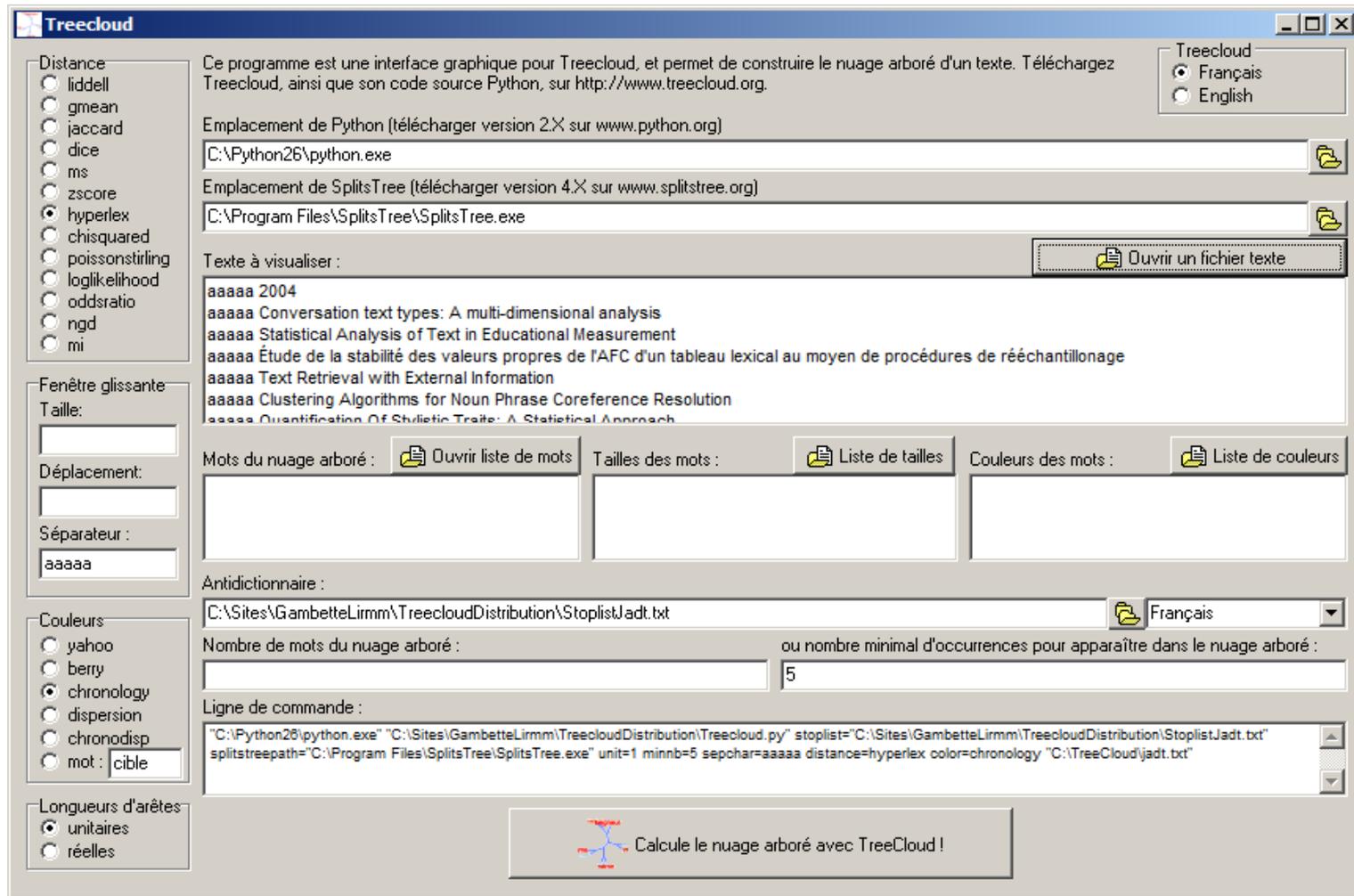
Evert, *Statistics of words cooccurrences*, thèse, 2005
Gambette, *User manual for TreeCloud*, 2009

Plan

- Le nuage arboré : occurrence + cooccurrence
- Des couleurs pour guider la lecture
- Processus de construction
- **Implémentations**
- Emplois du nuage arboré
- Illustration sur *Cinna* et *Othon*
- Perspectives

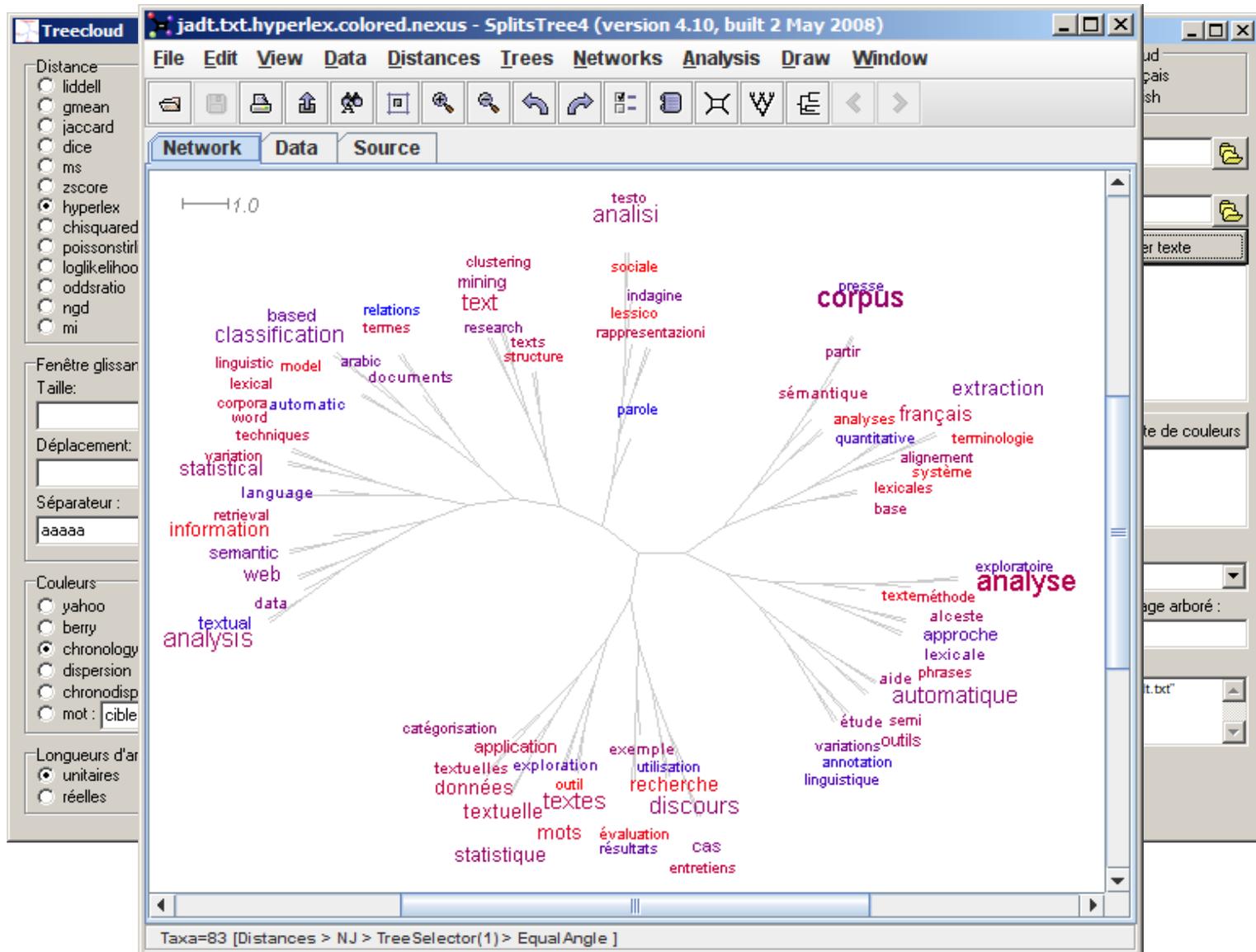
Implémentations

Logiciel libre TreeCloud (Python/Delphi) + SplitsTree (Java)



Implémentations

Logiciel libre TreeCloud (Python/Delphi) + SplitsTree (Java)



Interface web



Create! Downloads Gallery Credits FAQ
Créer! Téléchargements Galerie A propos FAQ

This website helps you to generate **tree clouds** from a text, that is word clouds where the words are arranged on a tree which reflects their semantic proximity inside the text. The first tree cloud appeared on [Jean Véronis's blog](#) in December 2007, you can now [create your own with this website](#), or [with the TreeCloud software](#).

Create your own tree cloud online!

Ce site web vous permet de générer des **nuages arborés** à partir d'un texte, c'est à dire des nuages de mots disposés autour d'un arbre qui indique leur proximité dans le texte. Le premier nuage arboré est apparu sur le [blog de Jean Véronis](#) en décembre 2007, vous pouvez maintenant [créer les vôtres avec ce site web](#), ou [avec le logiciel TreeCloud](#).

Créez vos propres nuages arborés en ligne !

Documents :

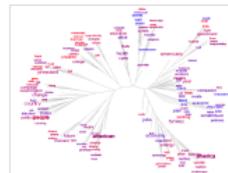


If you use TreeCloud or this website, please cite www.treecloud.org or:

Philippe Gambette et Jean Véronis: *Visualising a Text with a Tree Cloud*, In Locarek-Junge H. and Weihs C., editors, *Classification as a Tool of Research, Proc. of IFC'S'09 (11th Conference of the International Federation of Classification Societies)*, to appear, 2010 ([supplementary material](#)).

Pour des exemples d'utilisation de la visualisation en nuage arboré, vous pouvez lire :

Delphine Amstutz et Philippe Gambette: *Utilisation de la visualisation en nuage arboré pour l'analyse littéraire*, *Proc. of IADT'10 (10th International Conference on statistical analysis of textual data)*, à paraître, 2010 ([matériel supplémentaire](#)).



www.treecloud.org

Interface basée sur le logiciel libre NuageArboré de Jean-Charles Bontemps, en C, CGI/Python, et JavaScript.

<http://sourceforge.net/projects/nuagearbor/>

Interface web

www.treecloud.org

 [Create!](#) [Downloads](#) [Gallery](#) [Credits](#) [FAQ](#)
[Créer!](#) [Téléchargements](#) [Galerie](#) [A propos](#) [FAQ](#)

This website helps you to generate tree clouds. The words are arranged on a tree which reflects their relationships. The first tree cloud appeared on [Jean Véronis's website](#). You can [create your own with this website](#), or [with the TreeCloud software](#).

Create your own tree cloud online

Ce site web vous permet de générer des nuages de mots disposés autour d'un arbre. Le premier nuage arboré est apparu sur le site de [Jean Véronis](#). Vous pouvez maintenant [créer les vôtres avec ce site](#).

Créez vos propres nuages arborés

Documents :



If you use TreeCloud or this website, please contact [Philippe Gambette et Jean Véronis](#): [Visualizing Classification as a Tool of Research, Proc. of the 10th International Conference on Intelligent Systems in Management Science \(ISMS'10\)](#), to appear, 2010 ([supplementary material](#)).

Pour des exemples d'utilisation de la visualisation, voir [Delphine Amstutz et Philippe Gambette](#): [Utilisation de la visualisation de données pour l'analyse de données textuelles](#), [JADT'10 \(10th International Conference on Intelligent Systems in Management Science\)](#), 2010 ([supplémentaire](#)).

© 2007-2010 - Jean Véronis

 [Créer!](#) [Téléchargements](#) [Galerie](#) [A propos](#) [FAQ](#)

Créez vos propres nuages arborés !

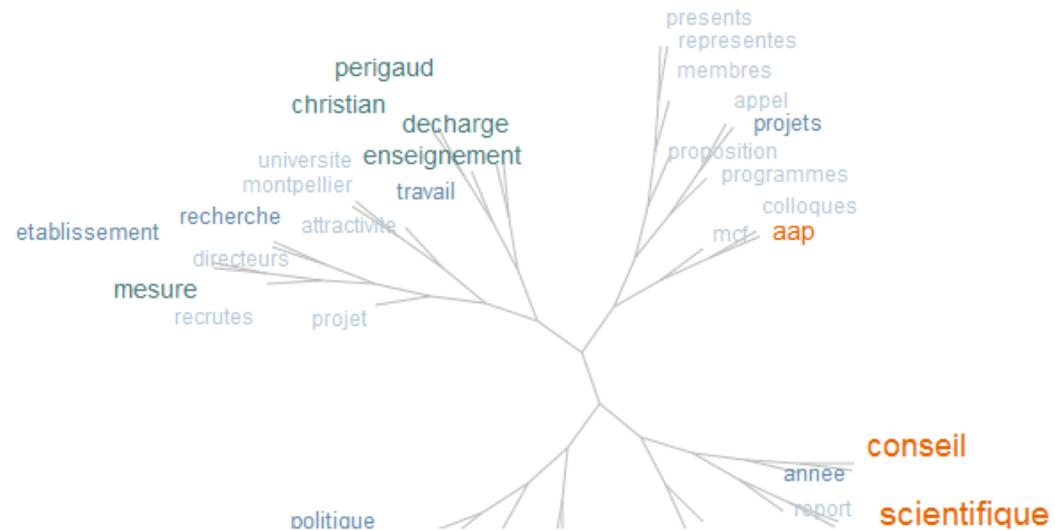
Collez votre texte dans le cadre ci-dessous, puis cliquez sur *Envoyer* ! Attention, l'utilisateur suivant verra votre texte quand il se connectera au site, si vous ne voulez pas faire apparaître vos textes, installez plutôt [TreeCloud](#) sur votre machine.

Texte :

conseil
scientifique
projet
proces
verbal

Envoyer

Vous pouvez déplacer les étiquettes par cliquer-glisser, l'étiquette reprend sa place d'origine lors d'un nouveau clic. L'infobulle indique le nombre d'occurrences du mot.



Plan

- Le nuage arboré : occurrence + cooccurrence
- Des couleurs pour guider la lecture
- Processus de construction
- Implémentations
- **Emplois du nuage arboré**
- Illustration sur *Cinna* et *Othon*
- Perspectives

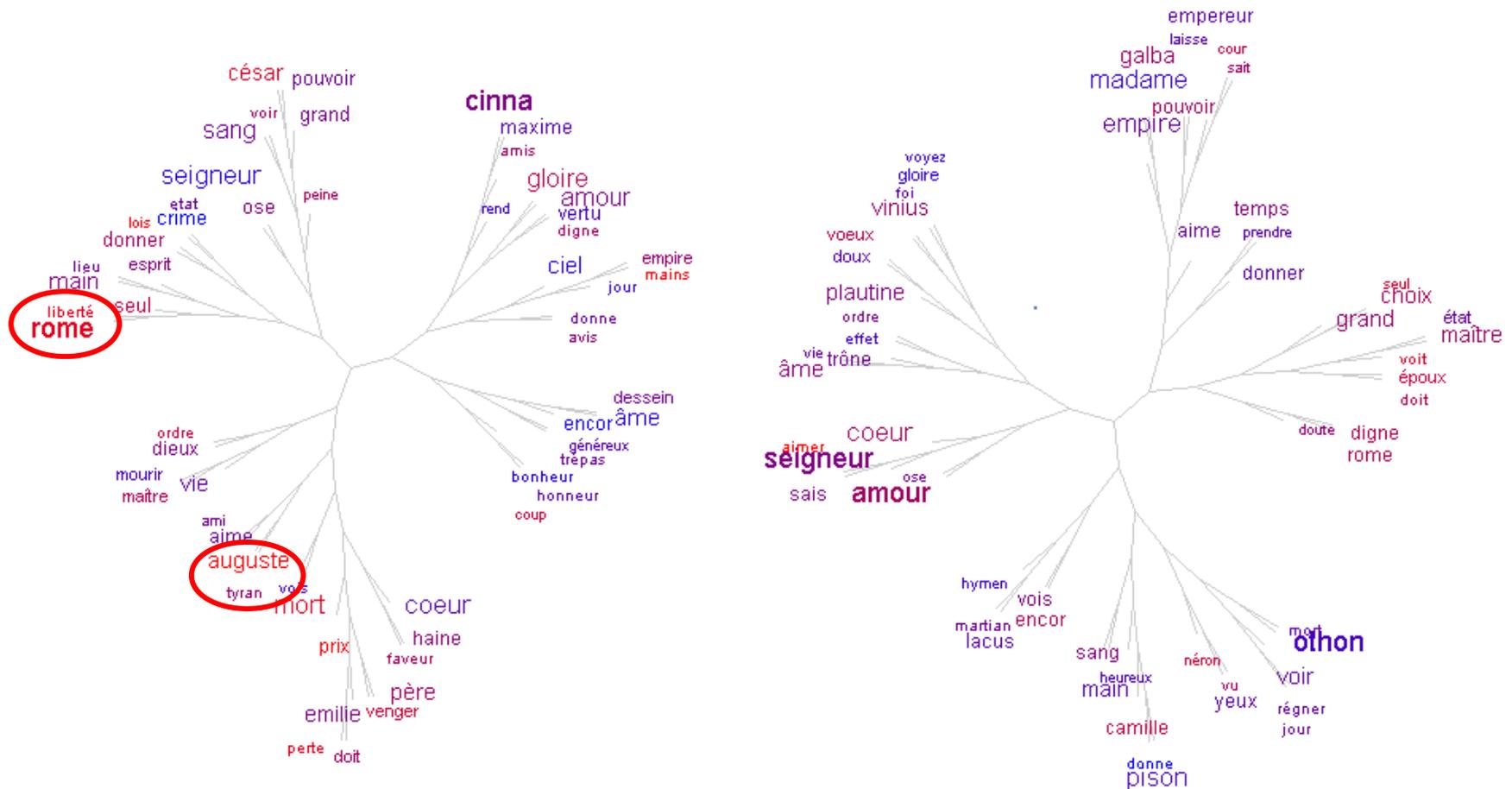
Emplois du nuage arboré

- susciter, formaliser et étayer des **hypothèses de travail**
- **comparer des textes** selon leur représentation arborée
- hiérarchiser l'utilisation d'**autres outils textométriques**
- représenter les **résultats de l'analyse**

Plan

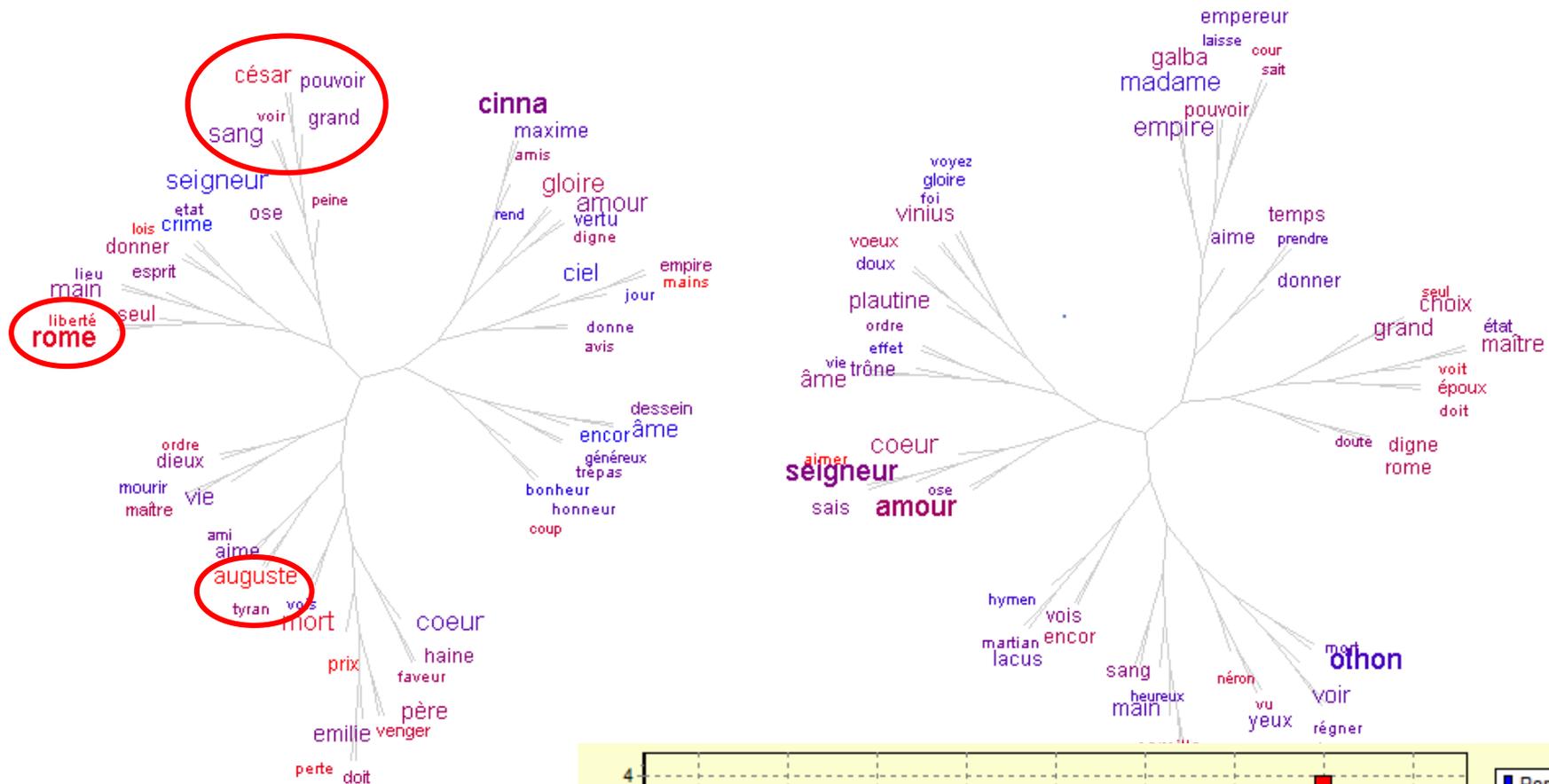
- Le nuage arboré : occurrence + cooccurrence
- Des couleurs pour guider la lecture
- Processus de construction
- Implémentations
- Emplois du nuage arboré
- **Illustration sur *Cinna* et *Othon***
- Perspectives

Illustration sur *Cinna* et *Othon*



Nuages arborés globaux des 60 mots les plus fréquents dans Cinna et Othon (distance Liddell, fenêtre de largeur 20), colorés chronologiquement (rouge au début, bleu à la fin)

Illustration sur *Cinna* et *Othon*



Spécificités d'emploi de « Rome », « liberté » et « empire » chez les différents personnages de Cinna dans Lexico3.

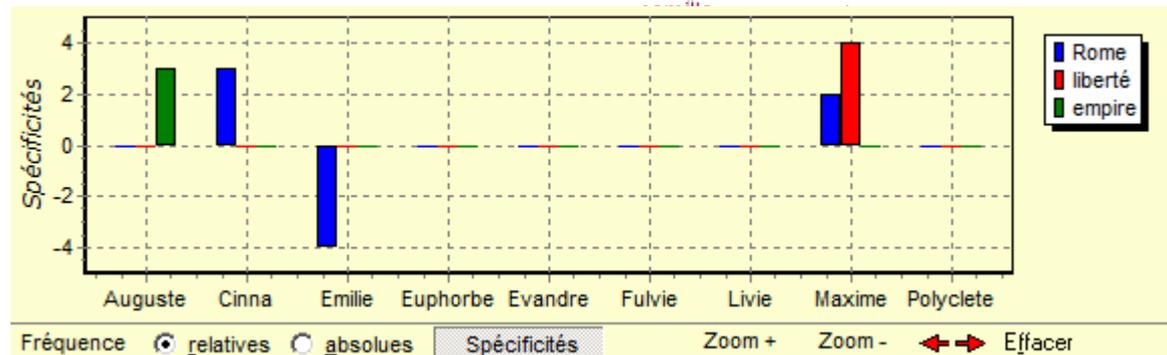
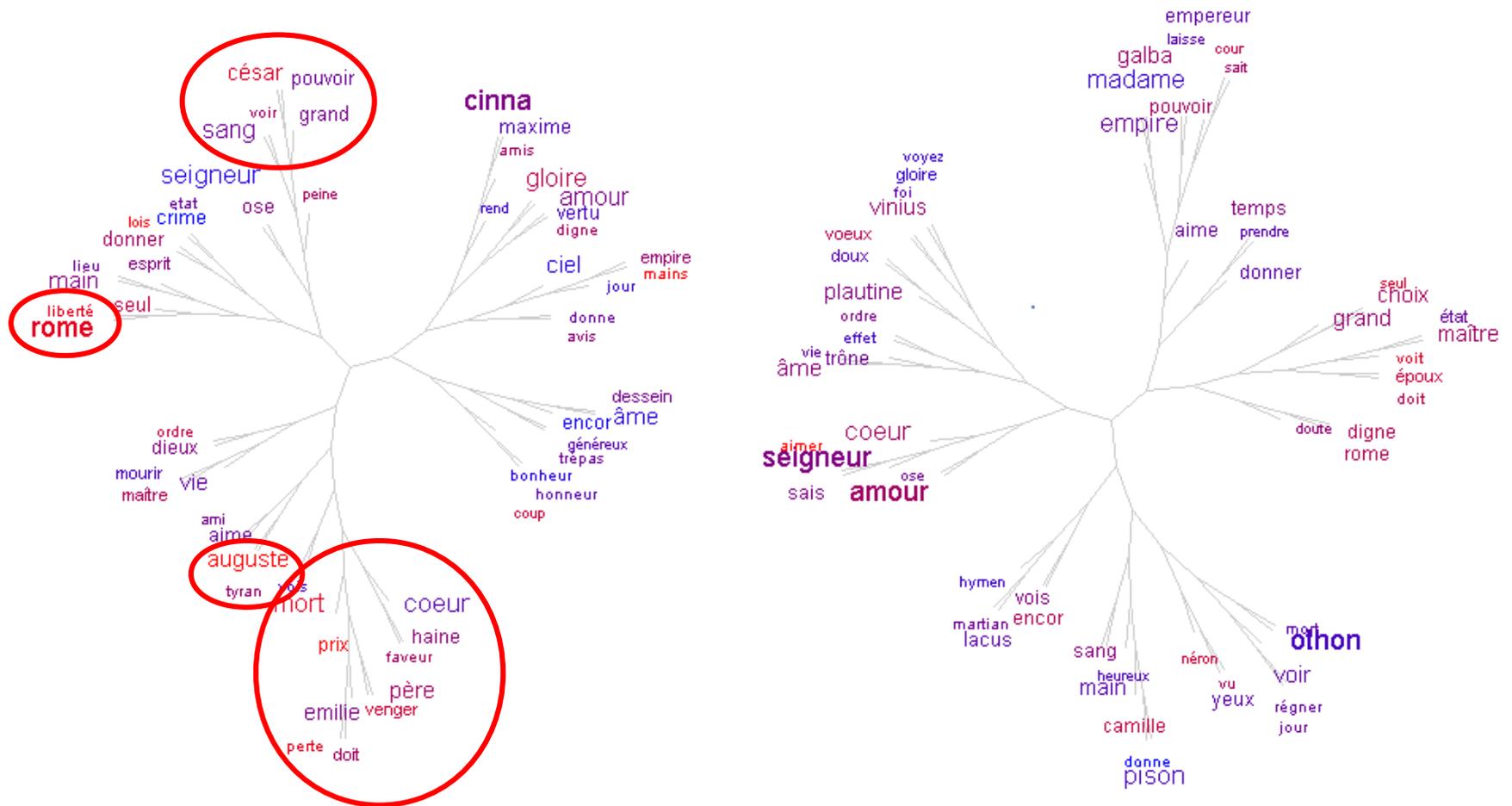
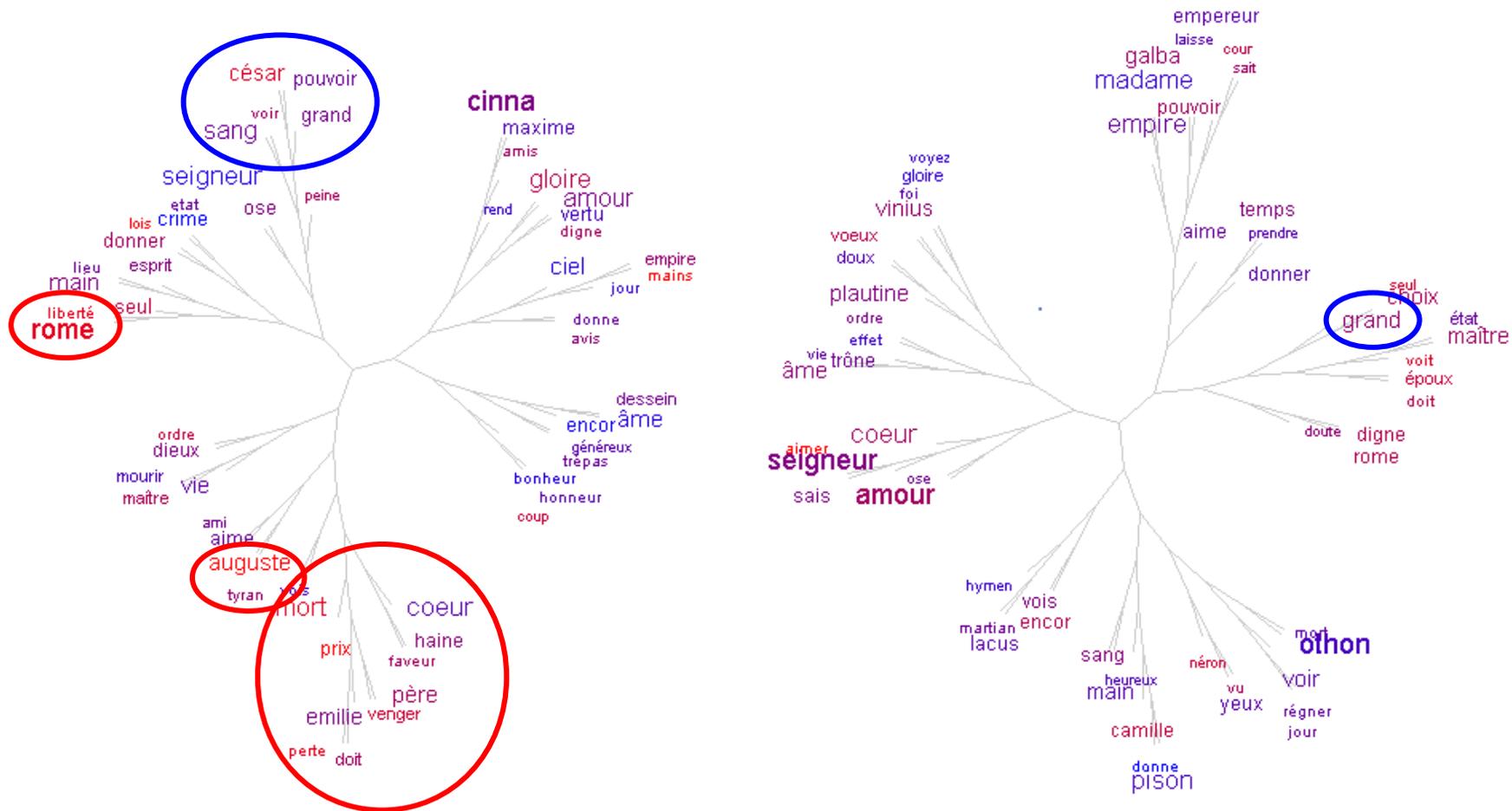


Illustration sur *Cinna* et *Othon*



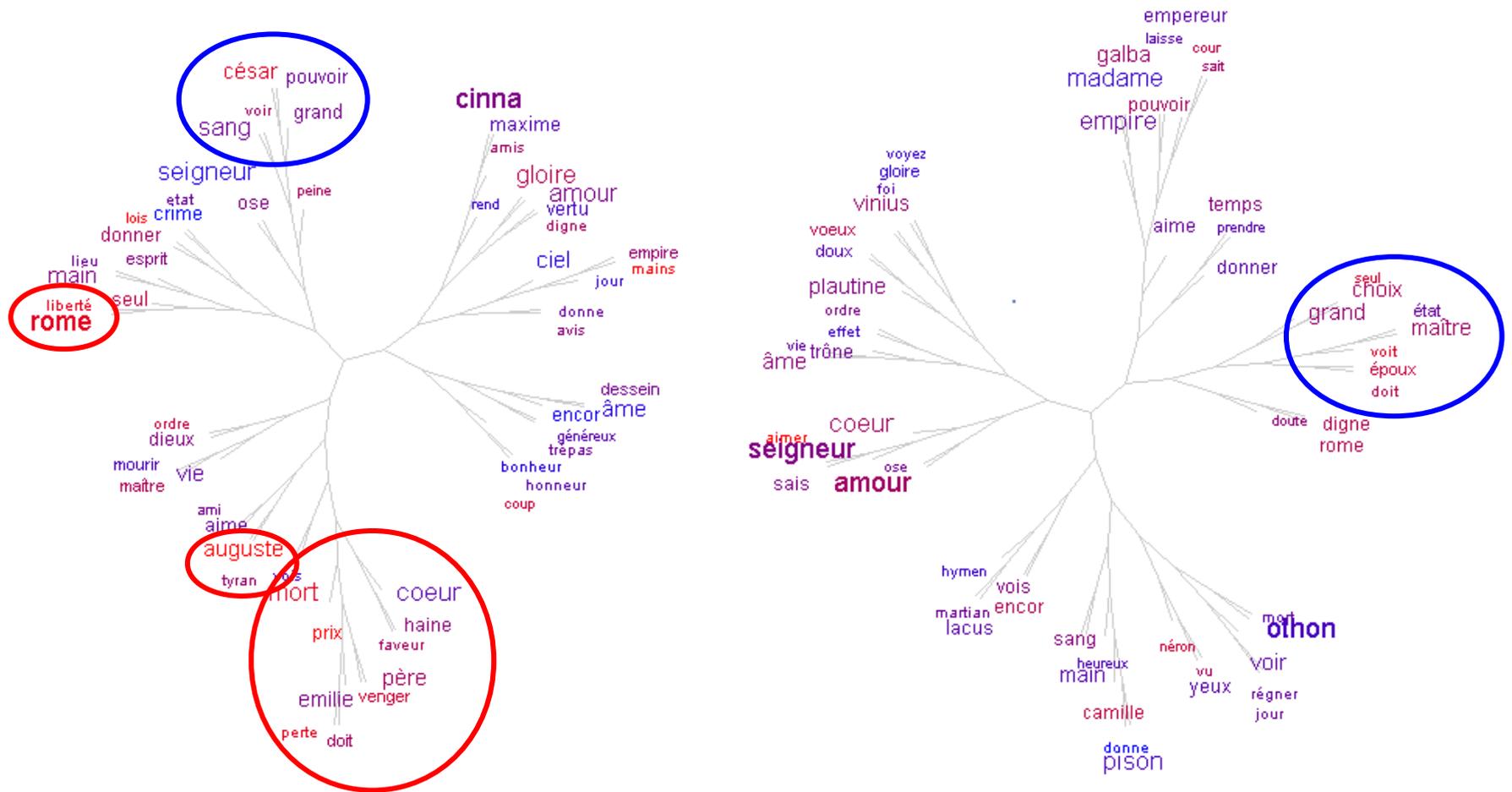
Nuages arborés globaux des 60 mots les plus fréquents dans Cinna et Othon (distance Liddell, fenêtre de largeur 20), colorés chronologiquement (rouge au début, bleu à la fin)

Illustration sur *Cinna* et *Othon*



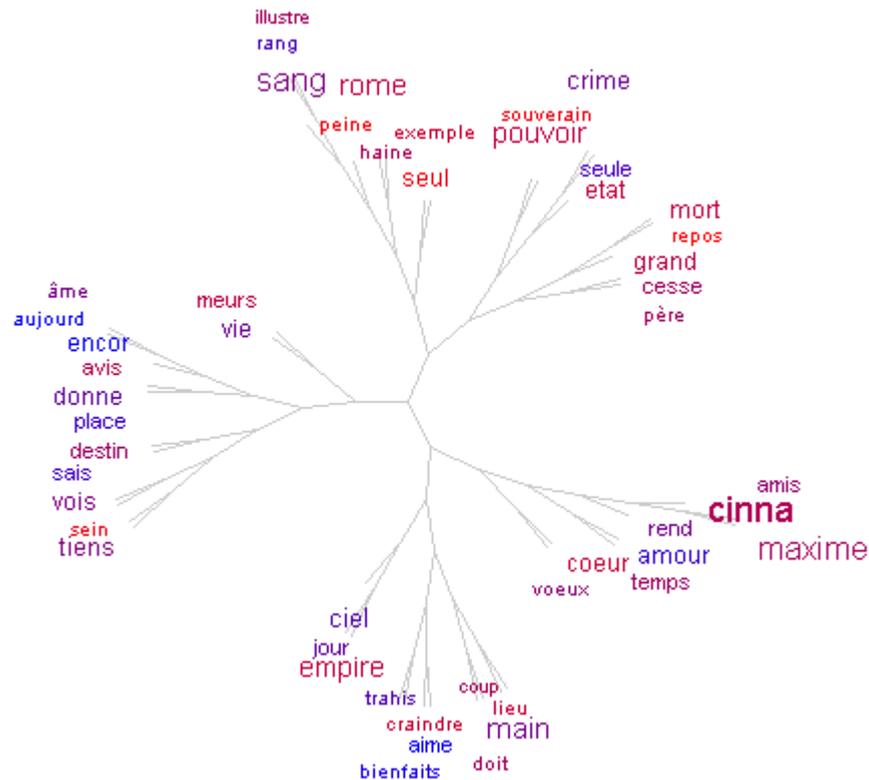
Nuages arborés globaux des 60 mots les plus fréquents dans Cinna et Othon (distance Liddell, fenêtre de largeur 20), colorés chronologiquement (rouge au début, bleu à la fin)

Illustration sur *Cinna* et *Othon*



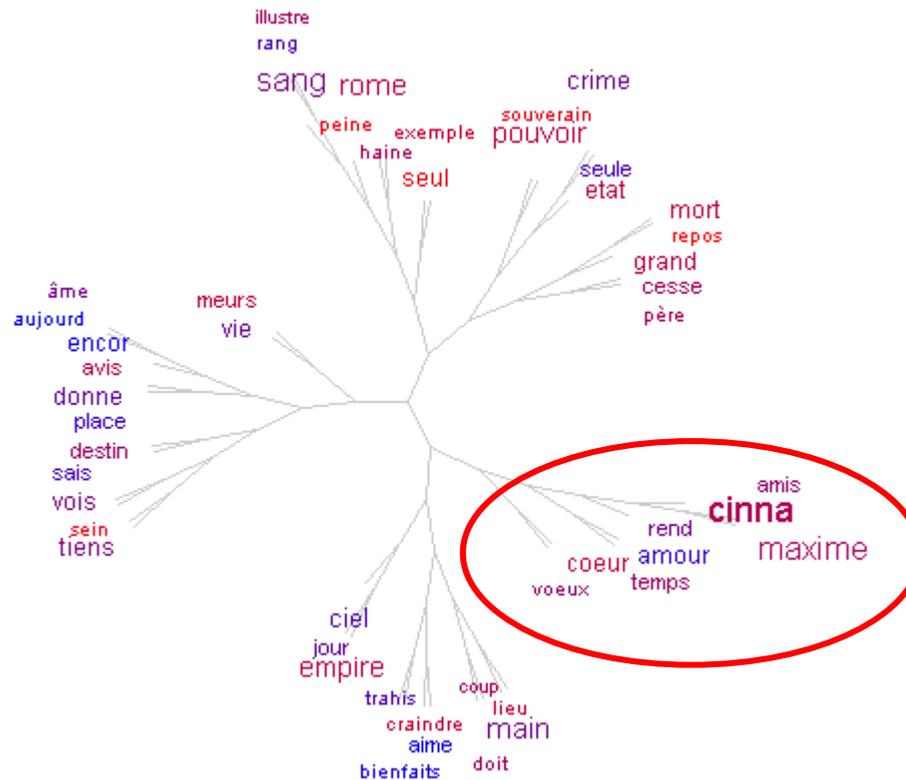
Nuages arborés globaux des 60 mots les plus fréquents dans Cinna et Othon (distance Liddell, fenêtre de largeur 20), colorés chronologiquement (rouge au début, bleu à la fin)

Illustration sur *Cinna* et *Othon*



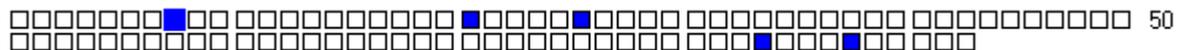
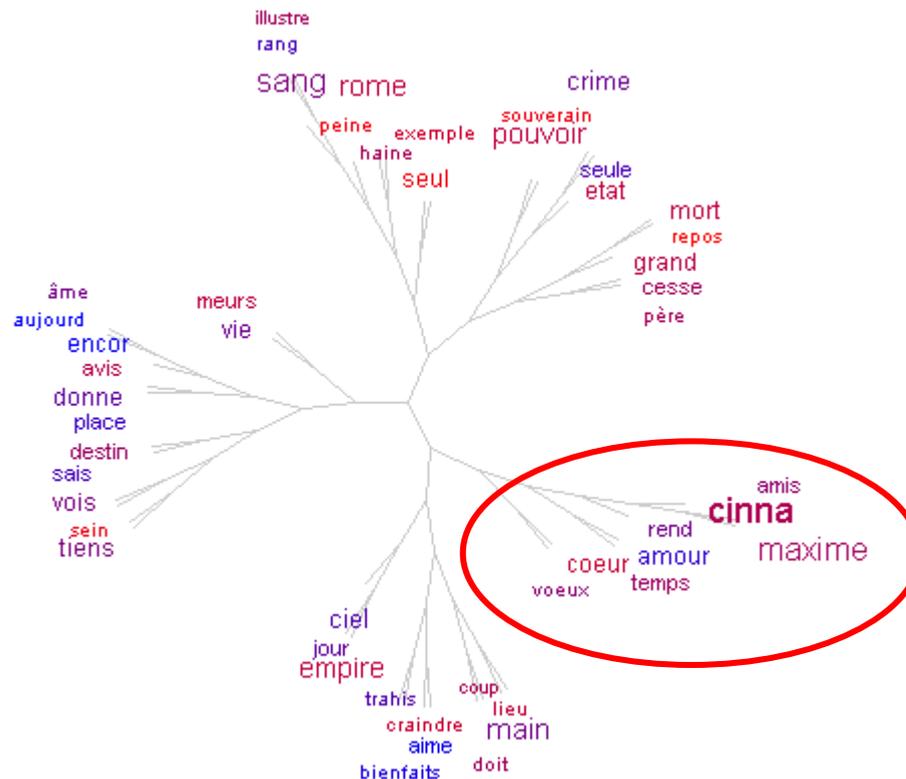
Nuage arboré des 50 mots les plus fréquents des paroles d'Auguste dans Cinna

Illustration sur *Cinna* et *Othon*



Nuage arboré des 50 mots les plus fréquents des paroles d'Auguste dans Cinna

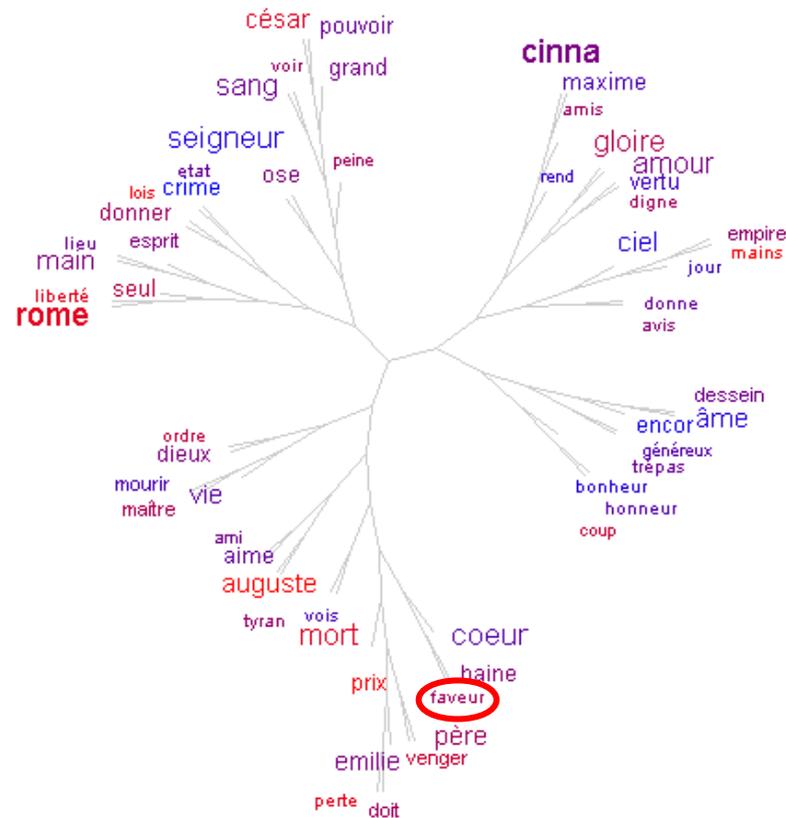
Illustration sur *Cinna* et *Othon*



Carte des sections Lexico3 et contextes de « amis » dans les paroles d'Auguste dans *Cinna*.

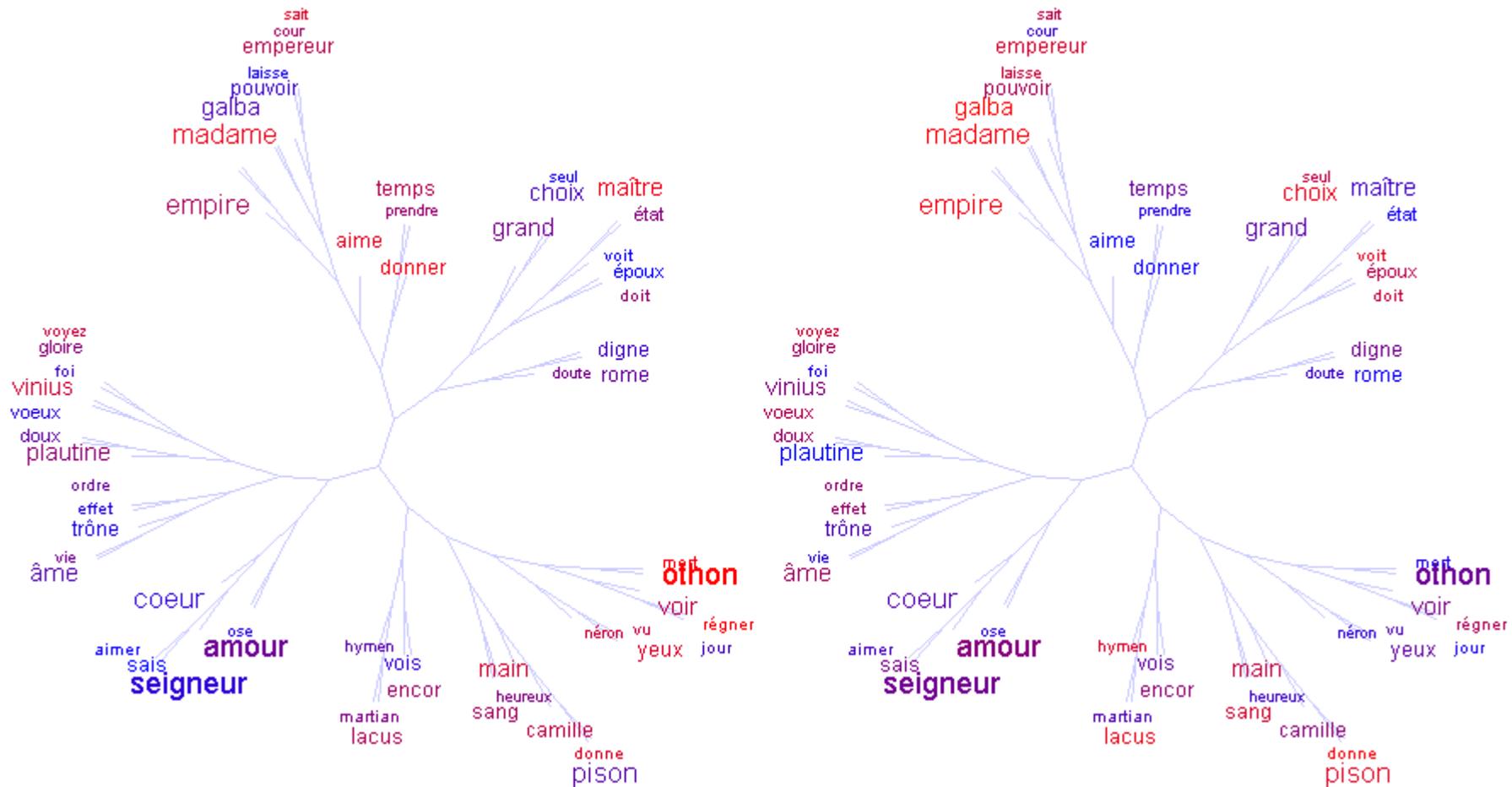
1. Voilà, mes chers **amis**, ce qui me met en peine.
2. Quoi ! mes plus chers **amis** ! quoi ! Cinna ! quoi ! Maxime !
3. Reprenez le pouvoir que vous m'avez commis, Si donnant des sujets il ôte les **amis**
4. Soyons **amis**, Cinna, c'est moi qui t'en convie
5. Il nous a trahis tous ; mais ce qu'il a commis Vous conserve innocents, et me rend mes **amis**.

Illustration sur *Cinna* et *Othon*



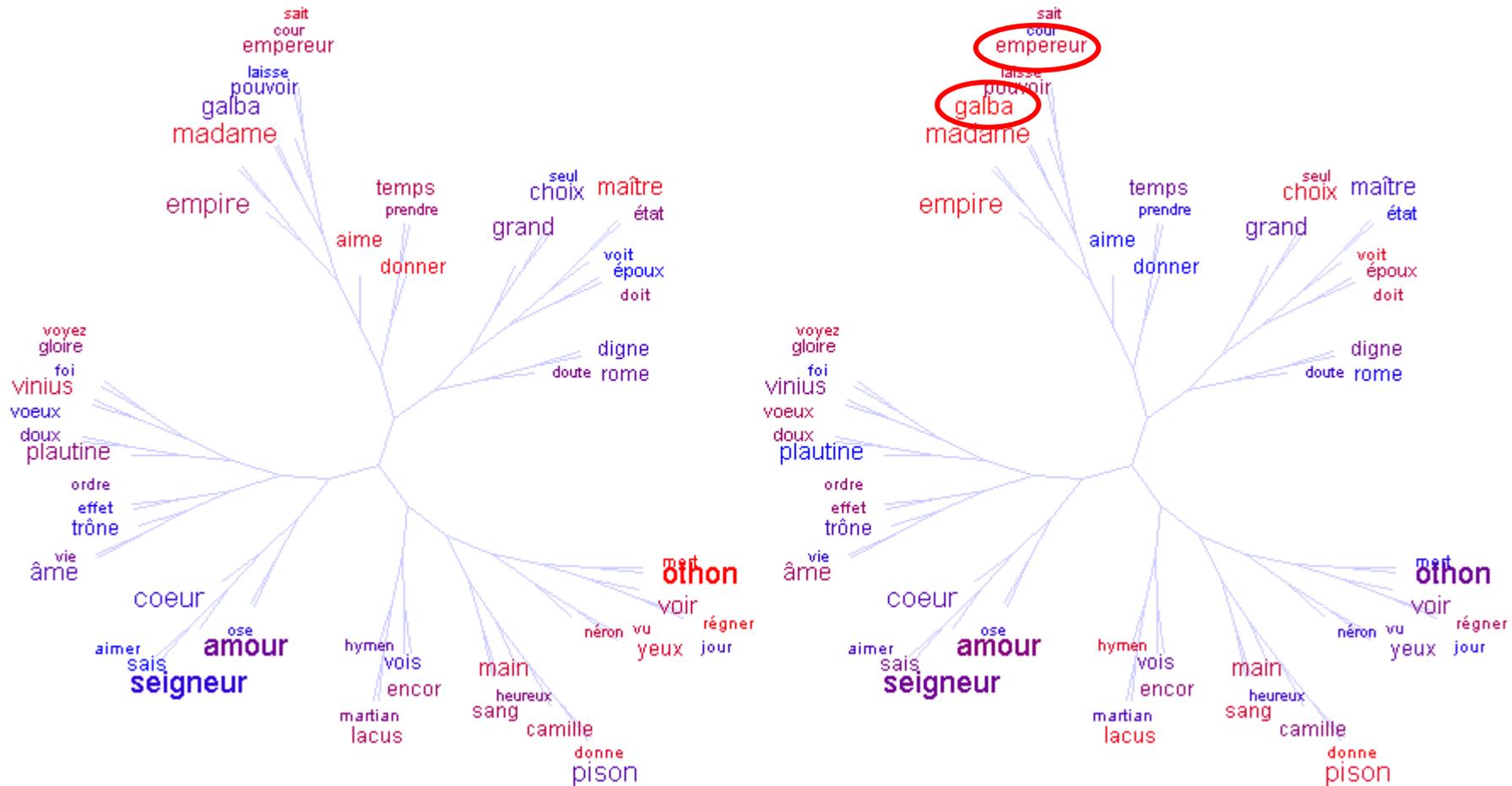
Nuages arborés globaux des 60 mots les plus fréquents dans Cinna et Othon (distance Liddell, fenêtre de largeur 20), colorés chronologiquement (rouge au début, bleu à la fin)

Illustration sur *Cinna* et *Othon*



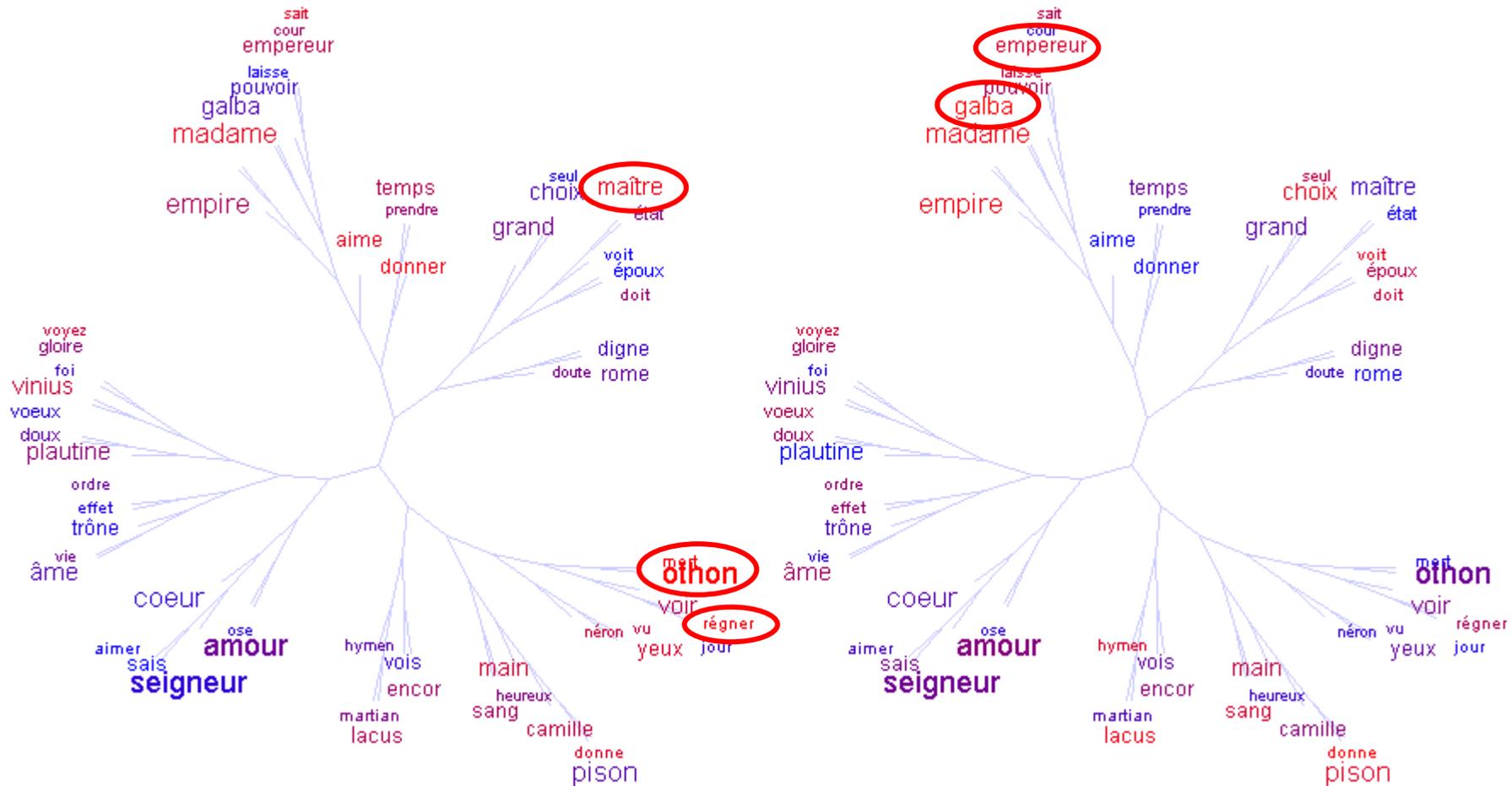
Nuage arboré des 30 mots les plus fréquents de la pièce Othon, coloré à gauche par rapport aux cooccurrences avec « Othon », à droite par rapport à celles avec « Galba »

Illustration sur *Cinna* et *Othon*



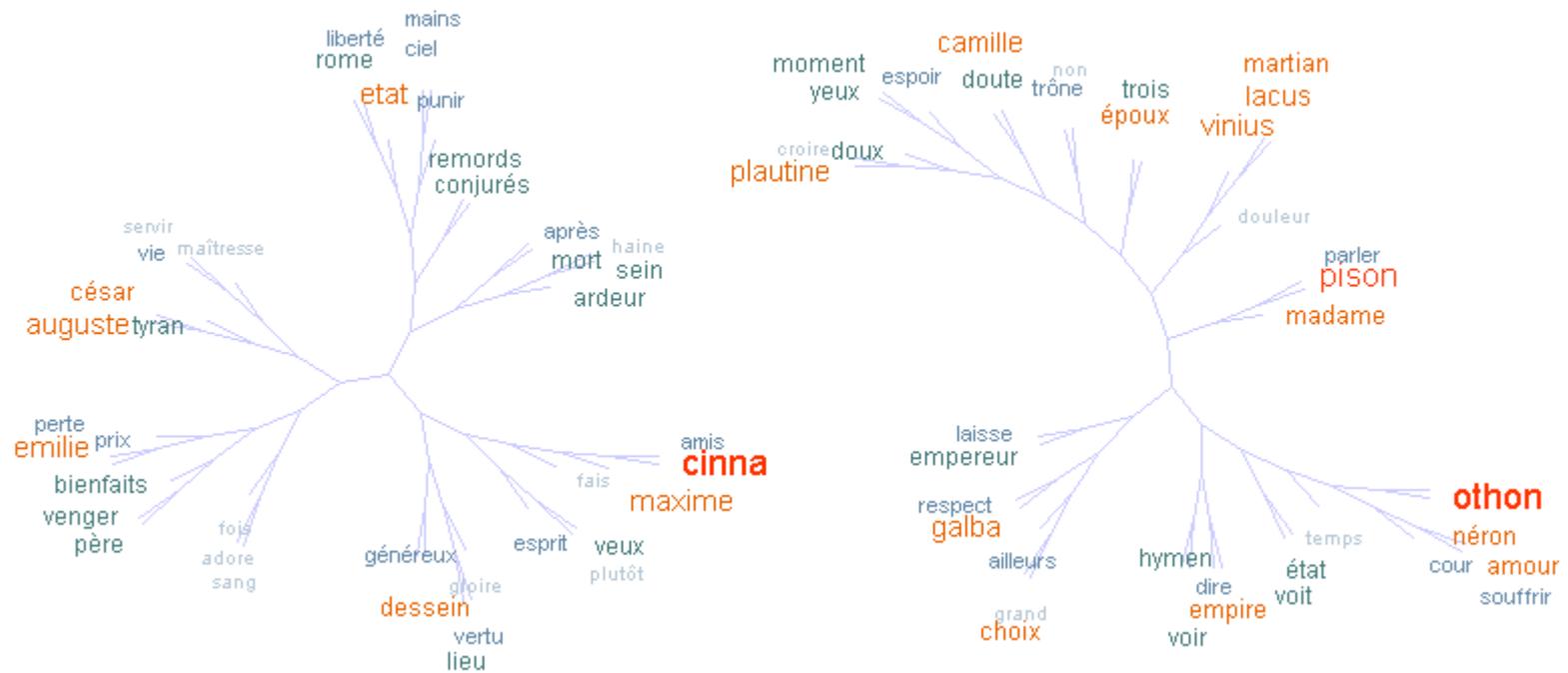
Nuage arboré des 30 mots les plus fréquents de la pièce Othon, coloré à gauche par rapport aux cooccurrences avec « Othon », à droite par rapport à celles avec « Galba »

Illustration sur *Cinna* et *Othon*



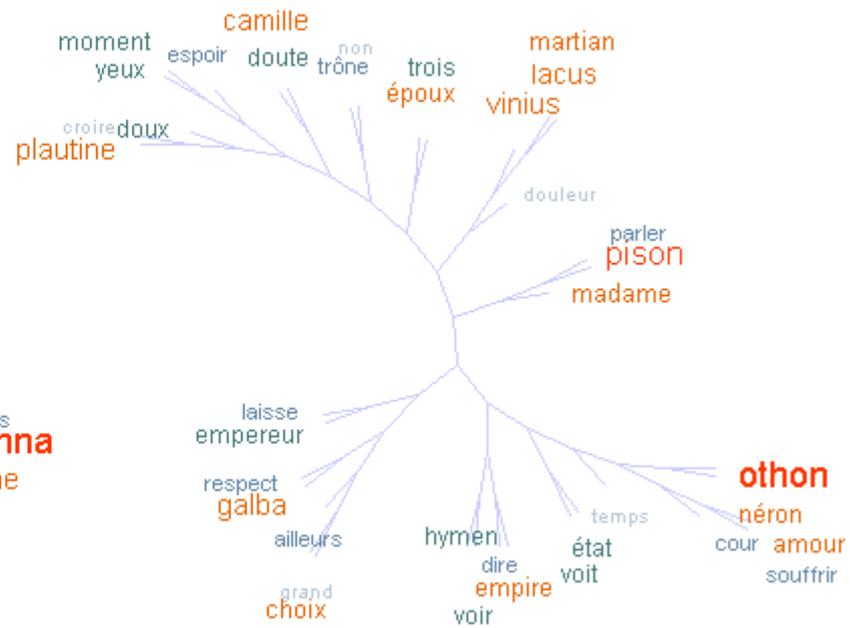
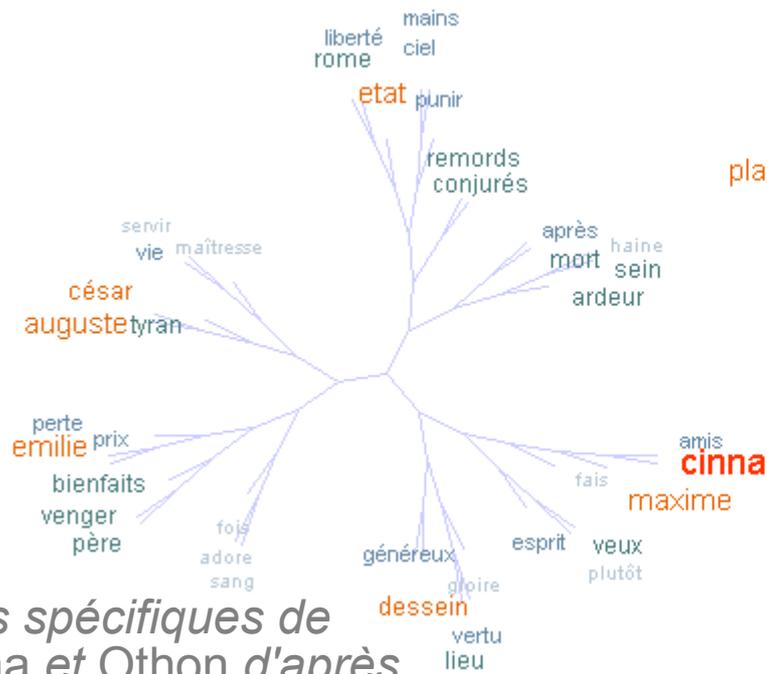
Nuage arboré des 30 mots les plus fréquents de la pièce Othon, coloré à gauche par rapport aux cooccurrences avec « Othon », à droite par rapport à celles avec « Galba »

Illustration sur *Cinna* et *Othon*



Nuages arborés des mots spécifiques de Cinna et Othon, dimensionnés et colorés d'après leur spécificité calculée dans Lexico3.

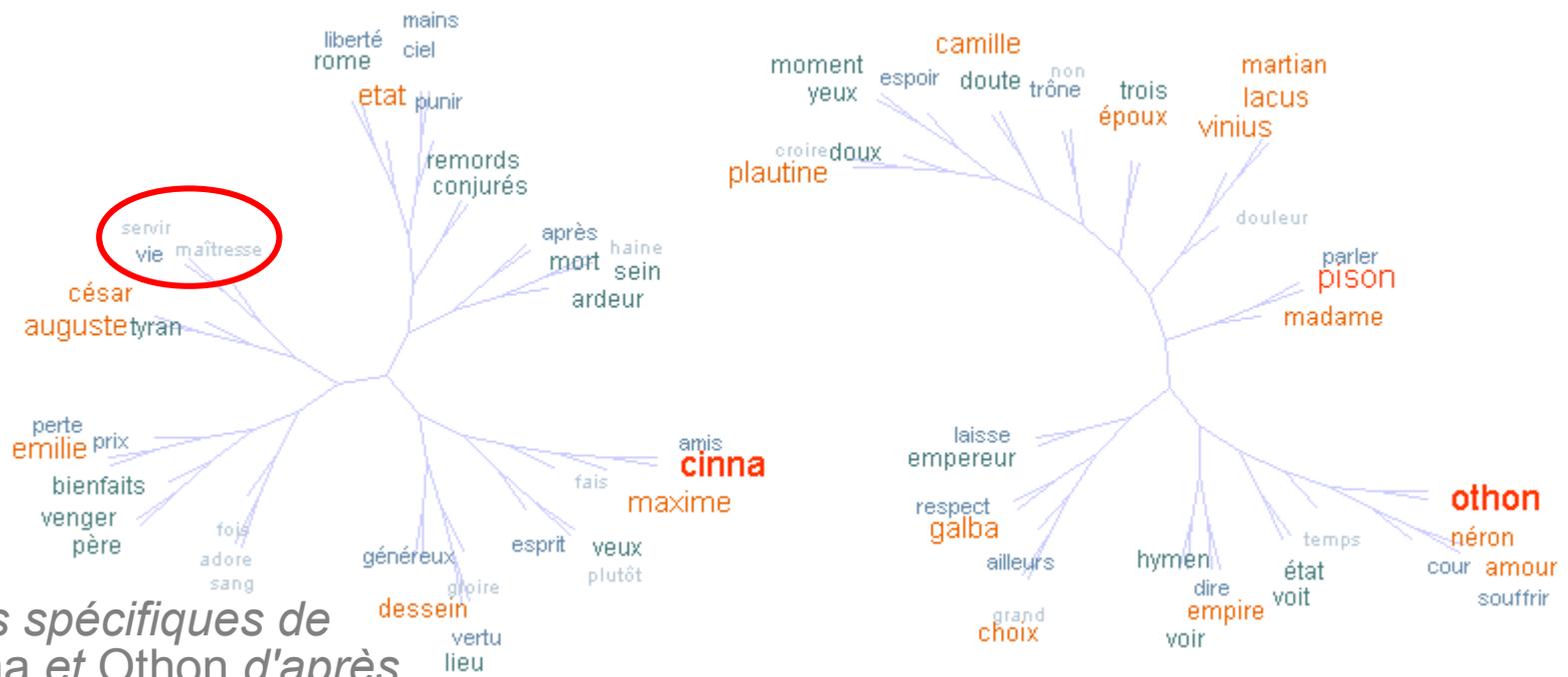
Illustration sur *Cinna* et *Othon*



mots spécifiques de Cinna et Othon d'après Lexico3

	<i>Cinna</i>	<i>Othon</i>
Lieu du pouvoir et objet de la confrontation entre les personnages	Rome (« liberté »)	Empire (« trône »)
Souverain en place	tyran	Empereur
Membres du corps politique	amis	maîtres / seigneurs
Moyens au service de la cause politique	gloire	amour matrimonial (« amour », « hymen », « choix »)
Caractérisation de la pièce	Pièce de FONDATION	Pièce de SUCCESSION DYNASTIQUE

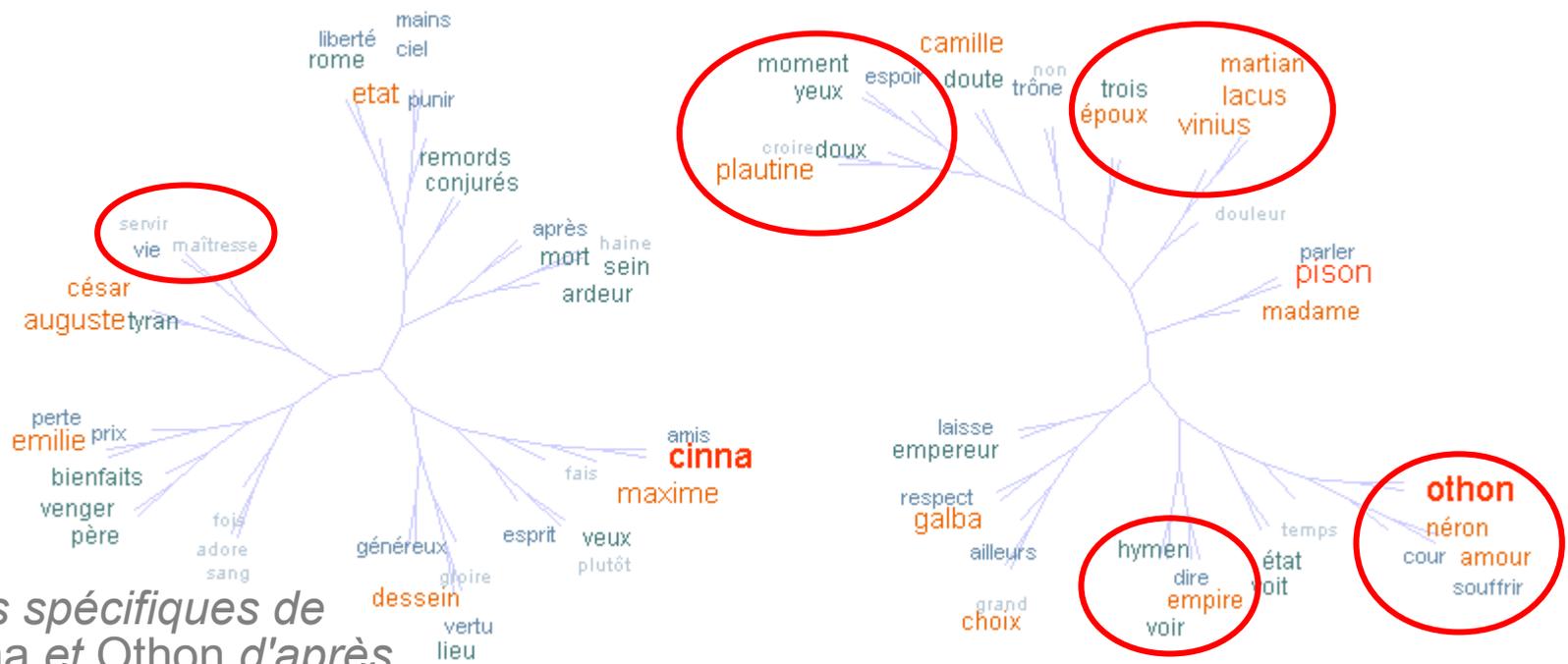
Illustration sur *Cinna* et *Othon*



mots spécifiques de Cinna et Othon d'après Lexico3

	<i>Cinna</i>	<i>Othon</i>
Lieu du pouvoir et objet de la confrontation entre les personnages	Rome (« liberté »)	Empire (« trône »)
Souverain en place	tyran	Empereur
Membres du corps politique	amis	maîtres / seigneurs
Moyens au service de la cause politique	gloire	amour matrimonial (« amour », « hymen », « choix »)
Caractérisation de la pièce	Pièce de FONDATION	Pièce de SUCCESSION DYNASTIQUE

Illustration sur *Cinna* et *Othon*



mots spécifiques de Cinna et Othon d'après Lexico3

	<i>Cinna</i>	<i>Othon</i>
Lieu du pouvoir et objet de la confrontation entre les personnages	Rome (« liberté »)	Empire (« trône »)
Souverain en place	tyran	Empereur
Membres du corps politique	amis	maîtres / seigneurs
Moyens au service de la cause politique	gloire	amour matrimonial (« amour », « hymen », « choix »)
Caractérisation de la pièce	Pièce de FONDATION	Pièce de SUCCESSION DYNASTIQUE

Plan

- Le nuage arboré : occurrence + cooccurrence
- Des couleurs pour guider la lecture
- Processus de construction
- Implémentations
- Emplois du nuage arboré
- Illustration sur *Cinna* et *Othon*
- Perspectives

Perspectives

- intégration de la visualisation en nuages arborés
 - dans les outils de textométrie existants
 - par des interfaces d'import/export adaptées
 - pour faciliter le retour au texte
- amélioration des méthodes de construction
 - transformée de Farris pour le calcul des distances
 - algorithme de Luong pour le calcul de l'arbre

