

Séminaire “Les méthodes quantitatives d’analyse des textes politiques”
20/11/2013 – Lille (MESHS, CERAPS)

Nuages arborés et analyse textuelle de corpus politiques avec TreeCloud

Philippe Gambette

LIGM
Université Paris-Est
Marne-la-Vallée



Plan

- Nuages arborés, intérêts et limites
- Construction des nuages arborés
- Options de coloration
- Utilisation des nuages arborés
- Prétraitements du texte utilisé
- Perspectives

Démo de TreeCloud

Firefox

TreeCloud - Nuage arboré - Build your own t... +

www.treedcloud.org/

Google

TreeCloud.org
Nuage Arboré

Create! Downloads Gallery Credits FAQ
Créer! Téléchargements Gallerie 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 :

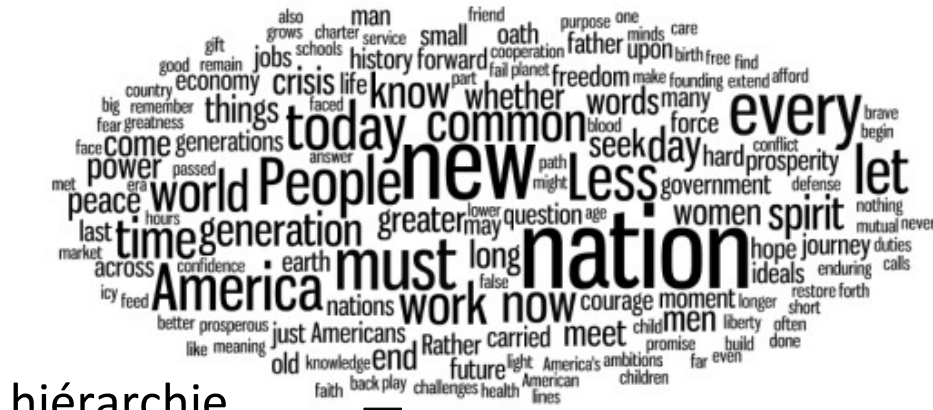


ABP

Plan

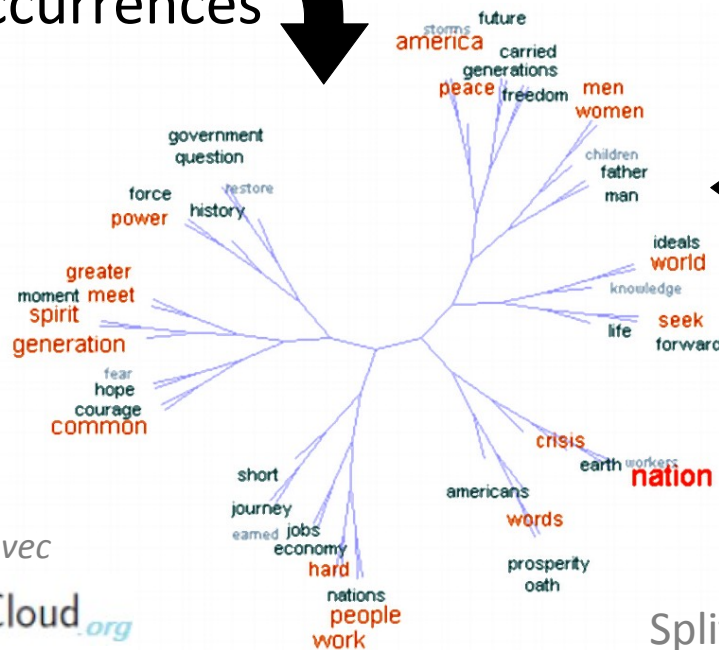
- Nuages arborés, intérêts et limites
- Construction des nuages arborés
- Options de coloration
- Utilisation des nuages arborés
- Prétraitements du texte utilisé
- Perspectives

Nuage arboré, une information double

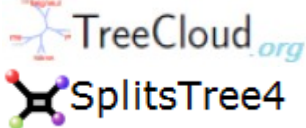


hiérarchie
des mots

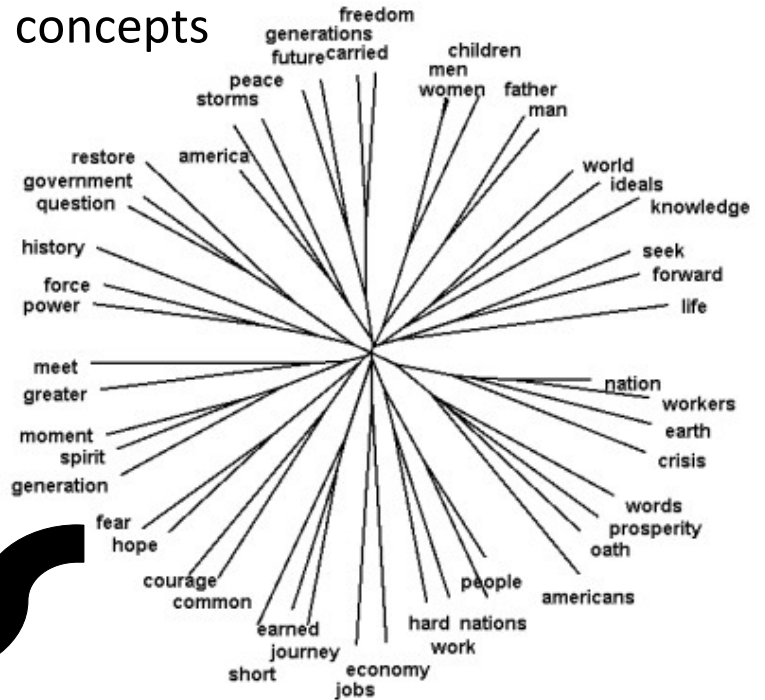
occurrences



construit avec



hiérarchie des
concepts



cooccurrences


Discours inaugural de Barack Obama

SplitsTree : Huson & Bryant, *Bioinformatics*, 2006

TreeCloud : Gambette & Véronis, *IFCS'09*

Intérêts de la visualisation arborée

- **Quantité d'information :**

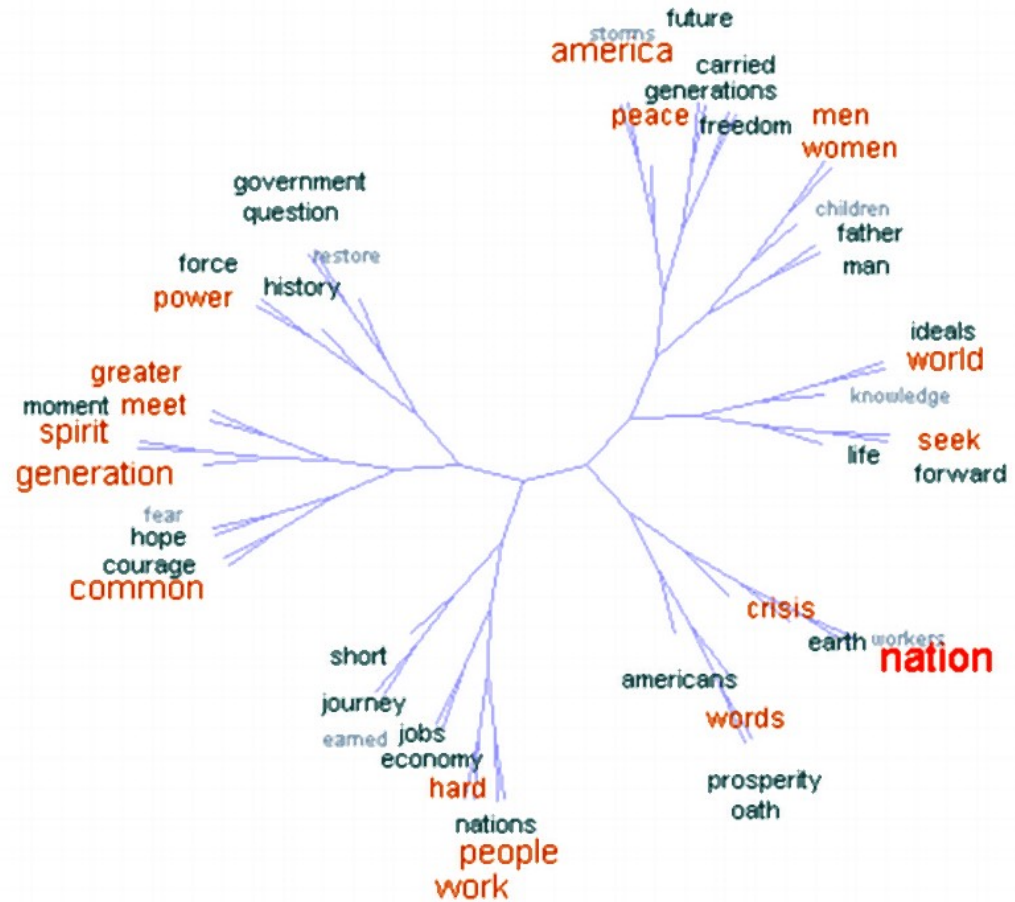
- nombre de sacs de mots imbriqués linéaire (\neq réseaux, AFC) 

- **Qualité d'information :**

- prise en compte d'une information globale pour le rapprochement de mots dans l'arbre (\neq réseaux)

- **Lisibilité :**

- dessin normalisé de l'arbre par méthode radiale (\neq réseaux)
- placement des étiquettes sans chevauchement (\neq AFC)



Intérêts de la visualisation arborée

- **Quantité d'information :**

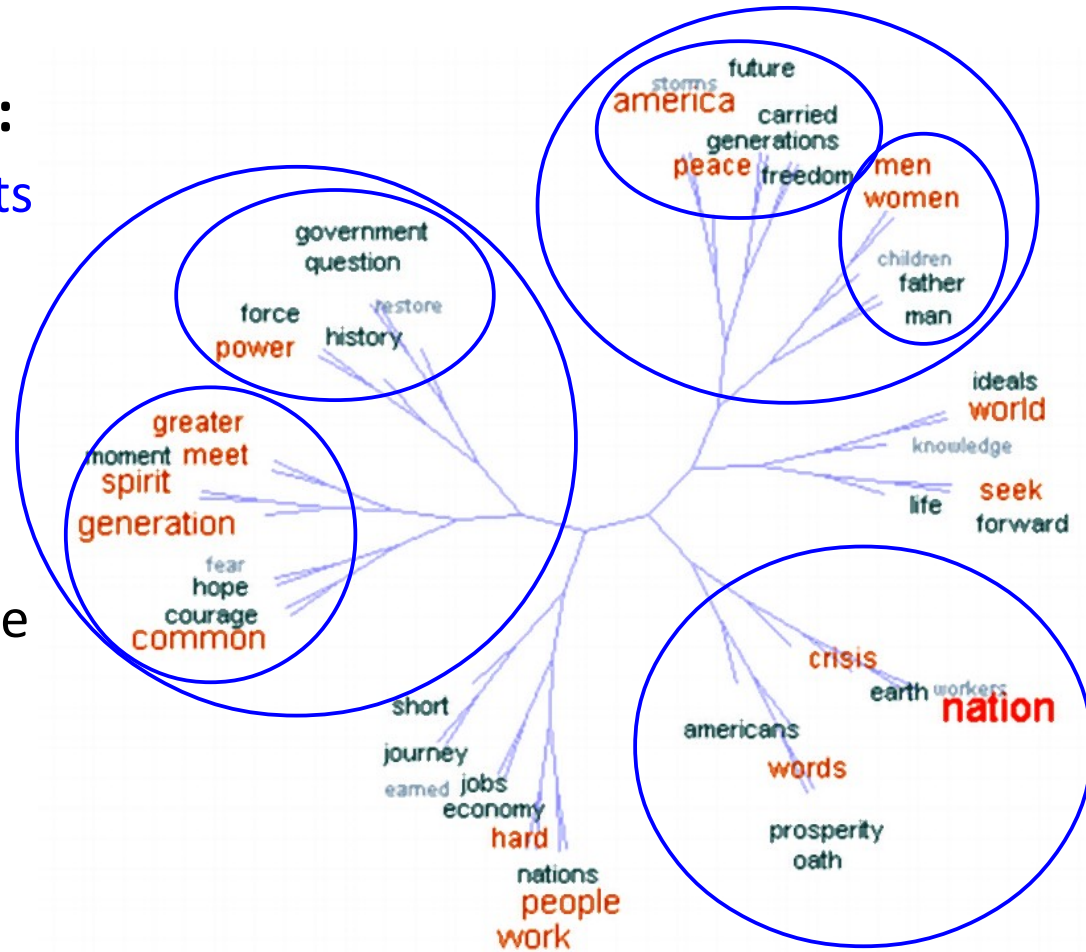
- nombre de **sacs de mots imbriqués** linéaire (\neq réseaux, AFC)

- **Qualité d'information :**

- prise en compte d'une information globale pour le rapprochement de mots dans l'arbre (\neq réseaux)

- **Lisibilité :**

- dessin normalisé de l'arbre par méthode radiale (\neq réseaux)
- placement des étiquettes sans chevauchement (\neq AFC)



Intérêts de la visualisation arborée

- **Quantité d'information :**

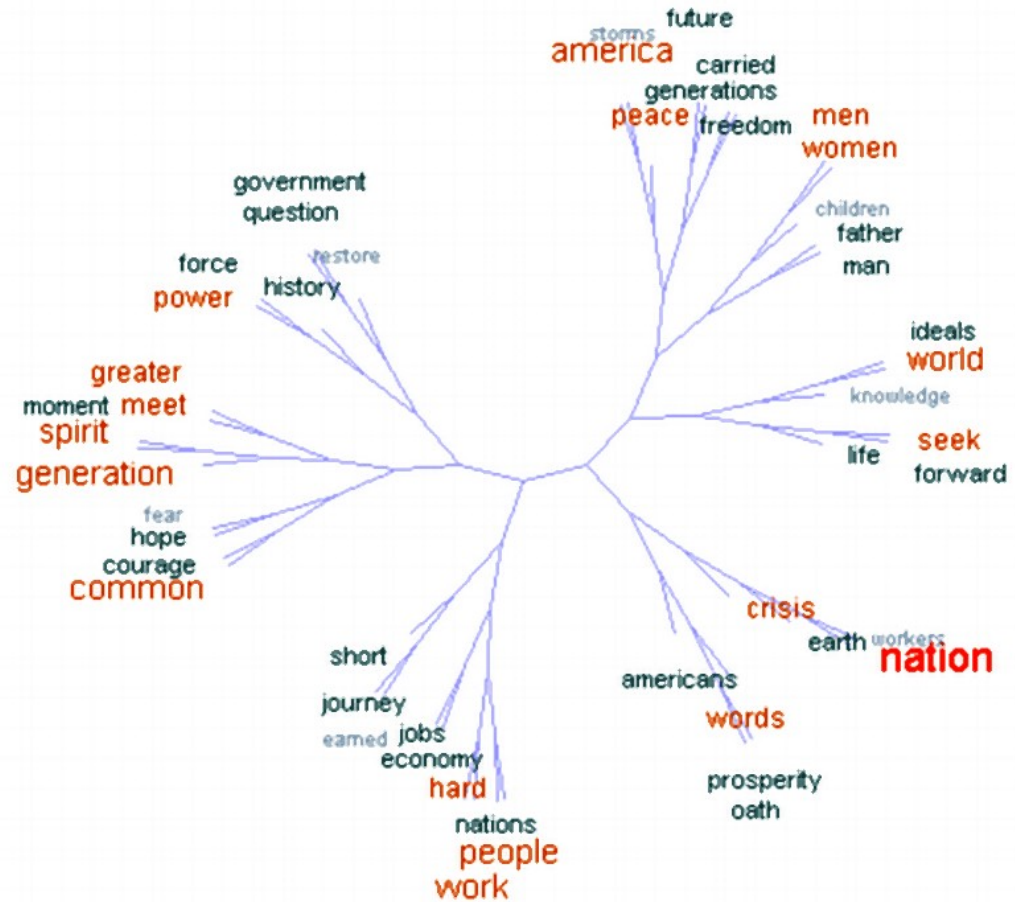
- nombre de sacs de mots imbriqués linéaire (\neq réseaux, AFC)

- **Qualité d'information :**

- prise en compte d'une information globale pour le rapprochement de mots dans l'arbre (\neq réseaux)

- **Lisibilité :**

- dessin normalisé de l'arbre par méthode radiale (\neq réseaux)
- placement des étiquettes sans chevauchement (\neq AFC)



Limites de la visualisation arborée

- **Quantité d'information :**

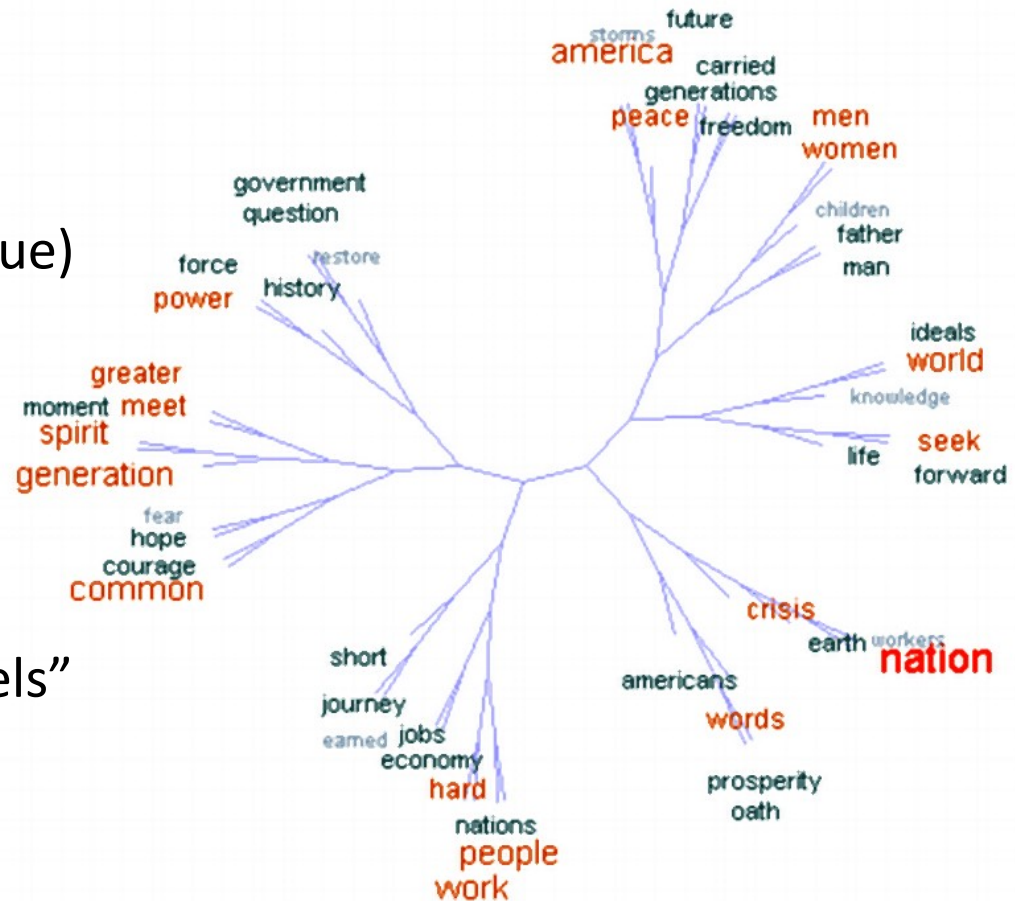
- nombre de mots linéaire en la largeur du dessin (≠ AFC, réseaux : quadratique)

- **Qualité d'information :**

- artefacts de la méthode : pas d'arbre "parfait"
- rapprochements "artificiels"
- instabilité (≠ AFC)

- **Lisibilité :**

- placement des étiquettes compliqué
- problème des longueurs de branches
- attention aux mauvaises interprétations



Limites de la visualisation arborée

- **Quantité d'information :**

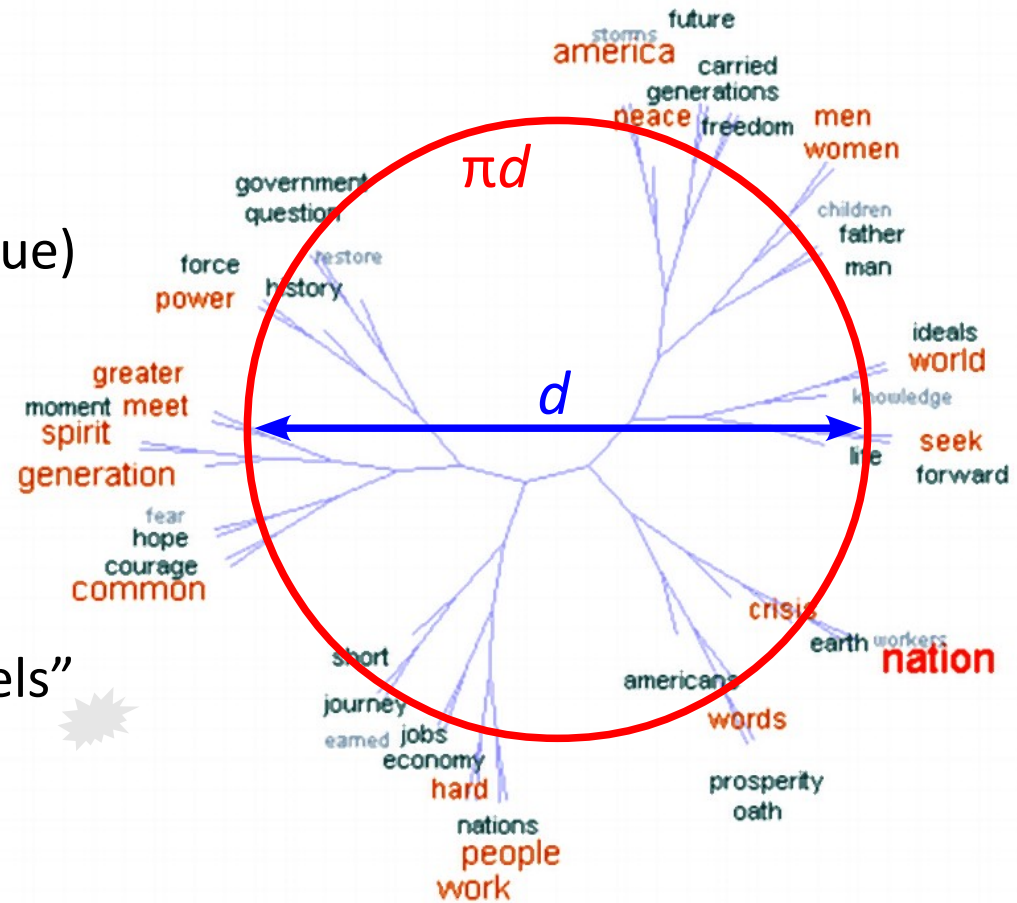
- nombre de mots **linéaire**
en la **largeur du dessin**
(\neq AFC, réseaux : quadratique)

- **Qualité d'information :**

- artefacts de la méthode : pas d'arbre "parfait"
- rapprochements "artificiels"
- instabilité (\neq AFC)

- **Lisibilité :**

- placement des étiquettes compliqué
- problème des longueurs de branches
- attention aux mauvaises interprétations



Limites de la visualisation arborée

- **Quantité d'information :**

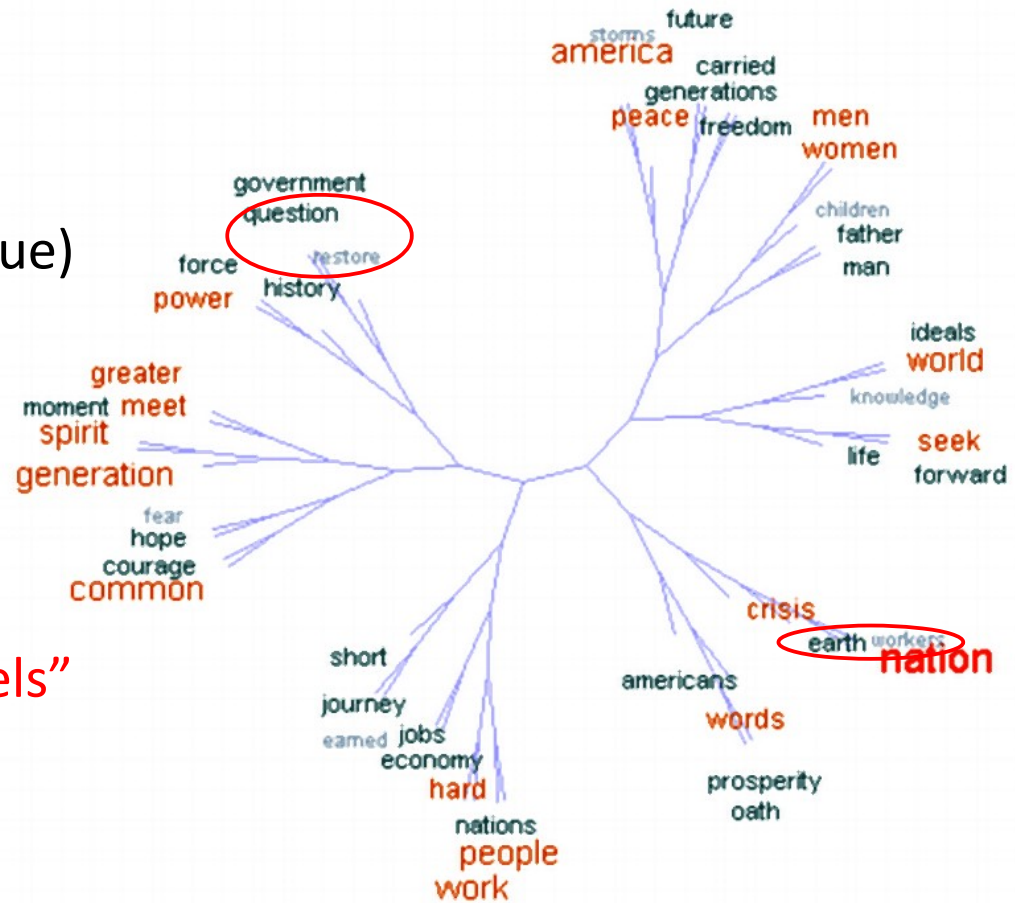
- nombre de mots linéaire en la largeur du dessin (≠ AFC, réseaux : quadratique)

- **Qualité d'information :**

- artefacts de la méthode : pas d'arbre "parfait"
- rapprochements "artificiels"
- instabilité (≠ AFC)

- **Lisibilité :**

- placement des étiquettes compliqué
- problème des longueurs de branches
- attention aux mauvaises interprétations



Limites de la visualisation arborée


- **Quantité d'information :**

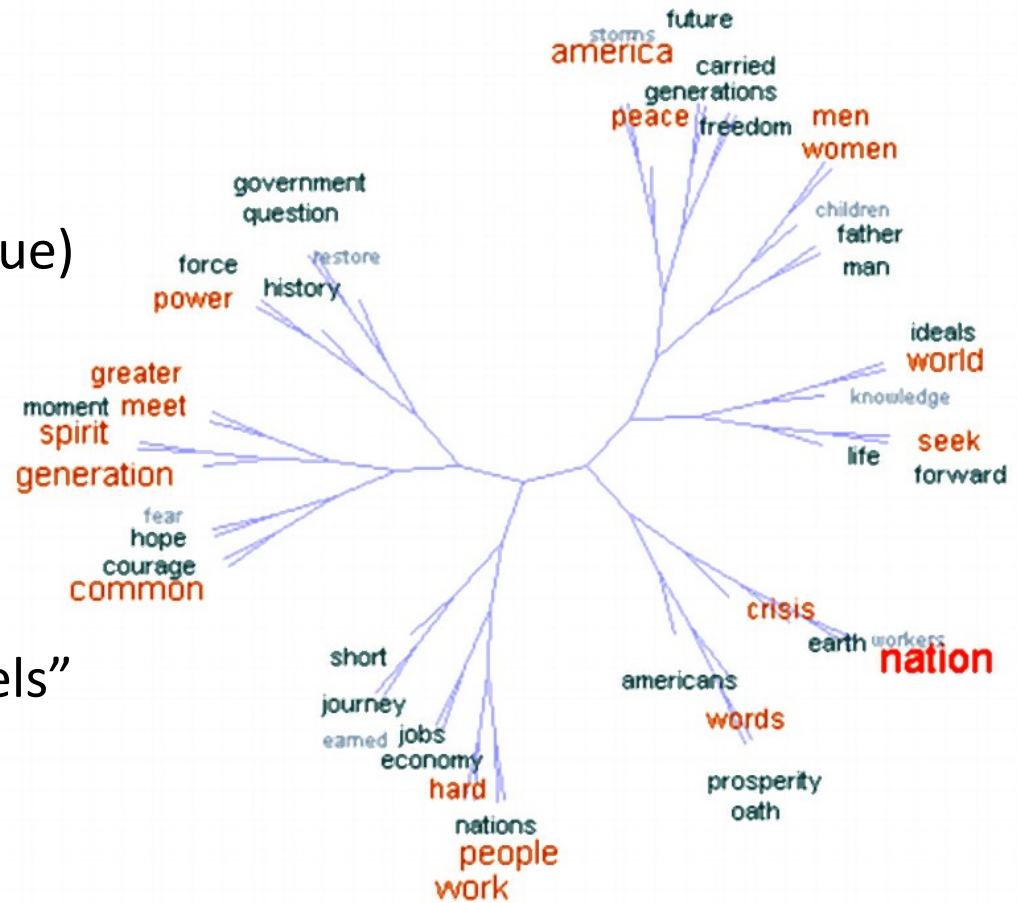
- nombre de mots linéaire en la largeur du dessin (≠ AFC, réseaux : quadratique)

- **Qualité d'information :**

- artefacts de la méthode : pas d'arbre "parfait"
- rapprochements "artificiels"
- instabilité (≠ AFC)

- **Lisibilité :**

- placement des étiquettes compliqué
- problème des longueurs de branches 
- attention aux mauvaises interprétations



Limites de la visualisation arborée

- **Quantité d'information :**

- nombre de mots linéaire en la largeur du dessin (\neq AFC, réseaux : quadratique)

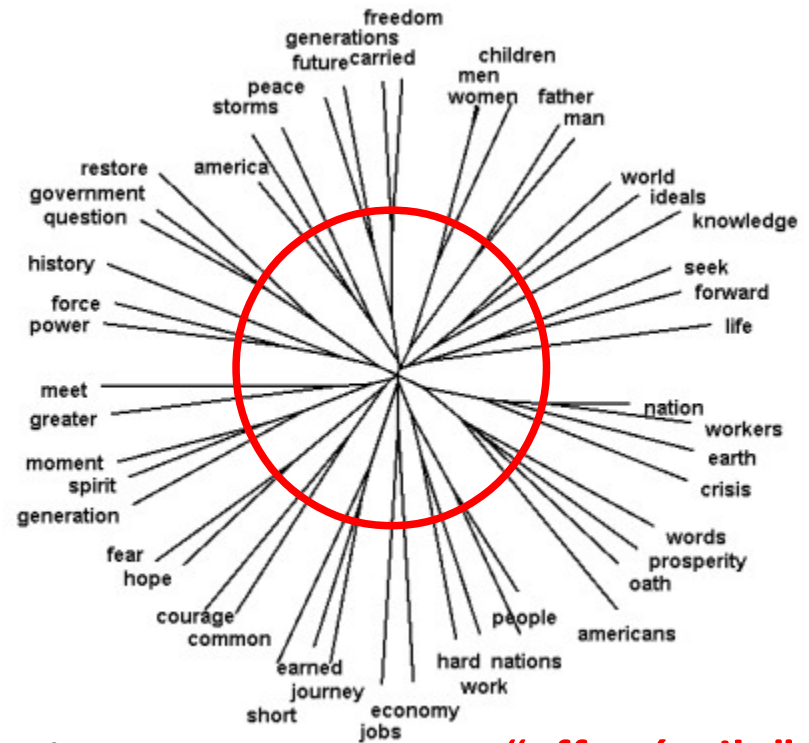
- **Qualité d'information :**

- artefacts de la méthode : pas d'arbre "parfait"
- rapprochements "artificiels"
- instabilité (\neq AFC)

- **Lisibilité :**

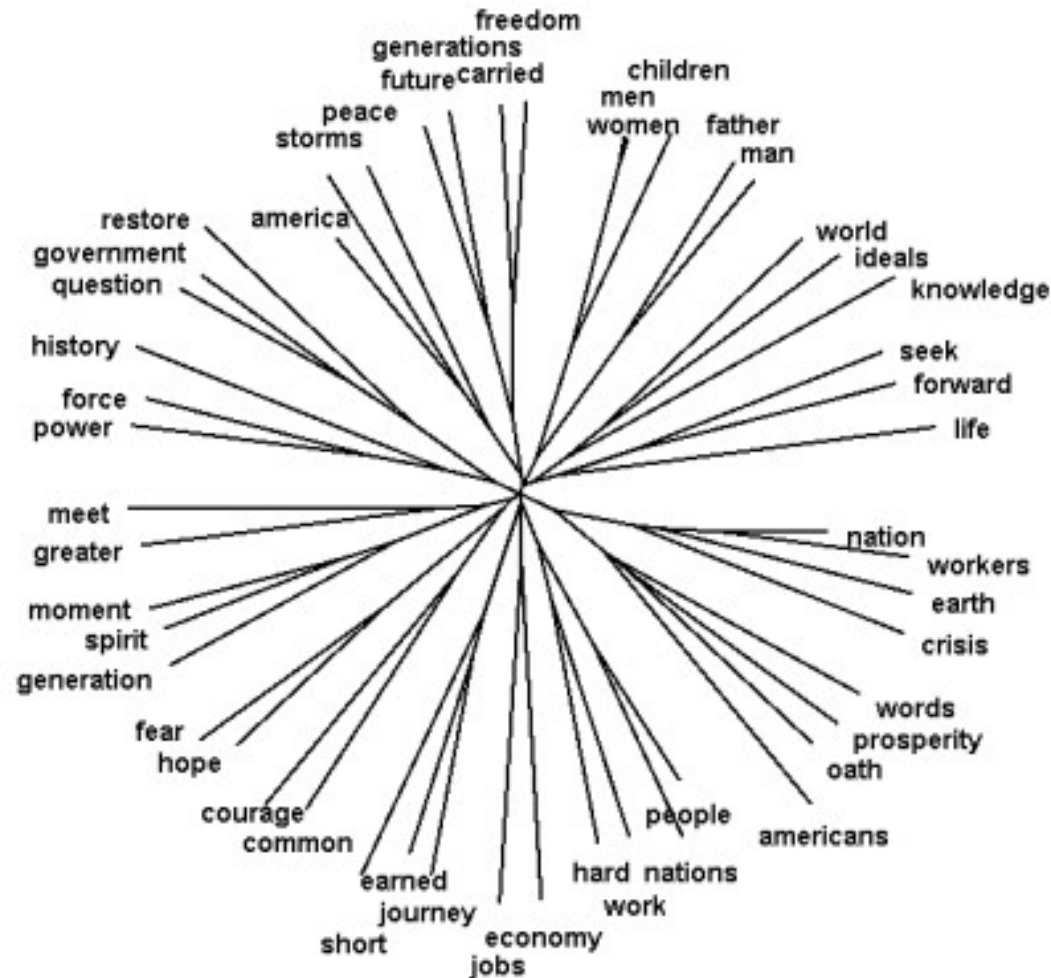
- placement des étiquettes compliqué
- **problème des longueurs de branches**
- attention aux mauvaises interprétations

Gambette, Gala, & Nasr, Longueur de branches et arbres de mots, *Corpus* 11:129-146, 2012.



**“effet étoile”,
centre illisible**

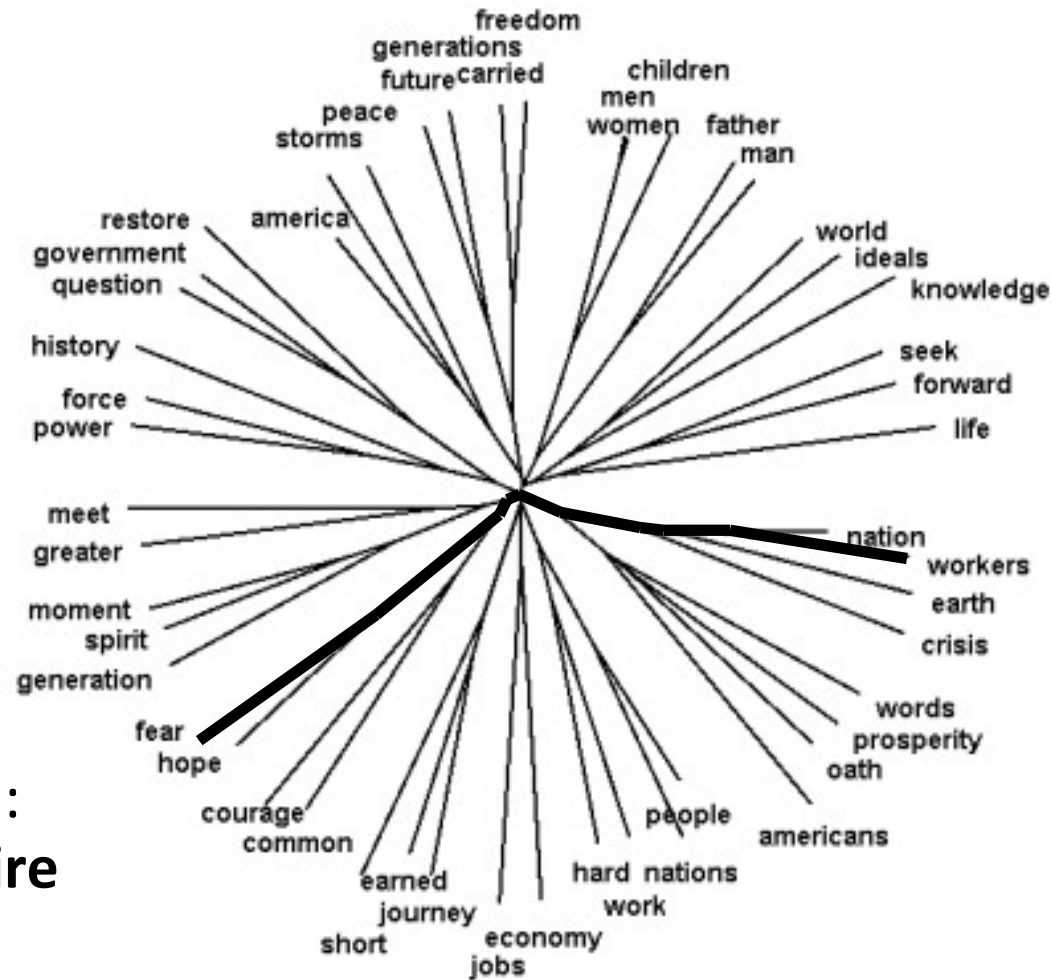
Interprétation réelle



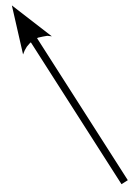
Principe de construction de l'arbre :

Les **distances** dans l'arbre entre deux mots reflètent *au mieux* le **degré de cooccurrence** entre ces deux mots

Interprétation réelle

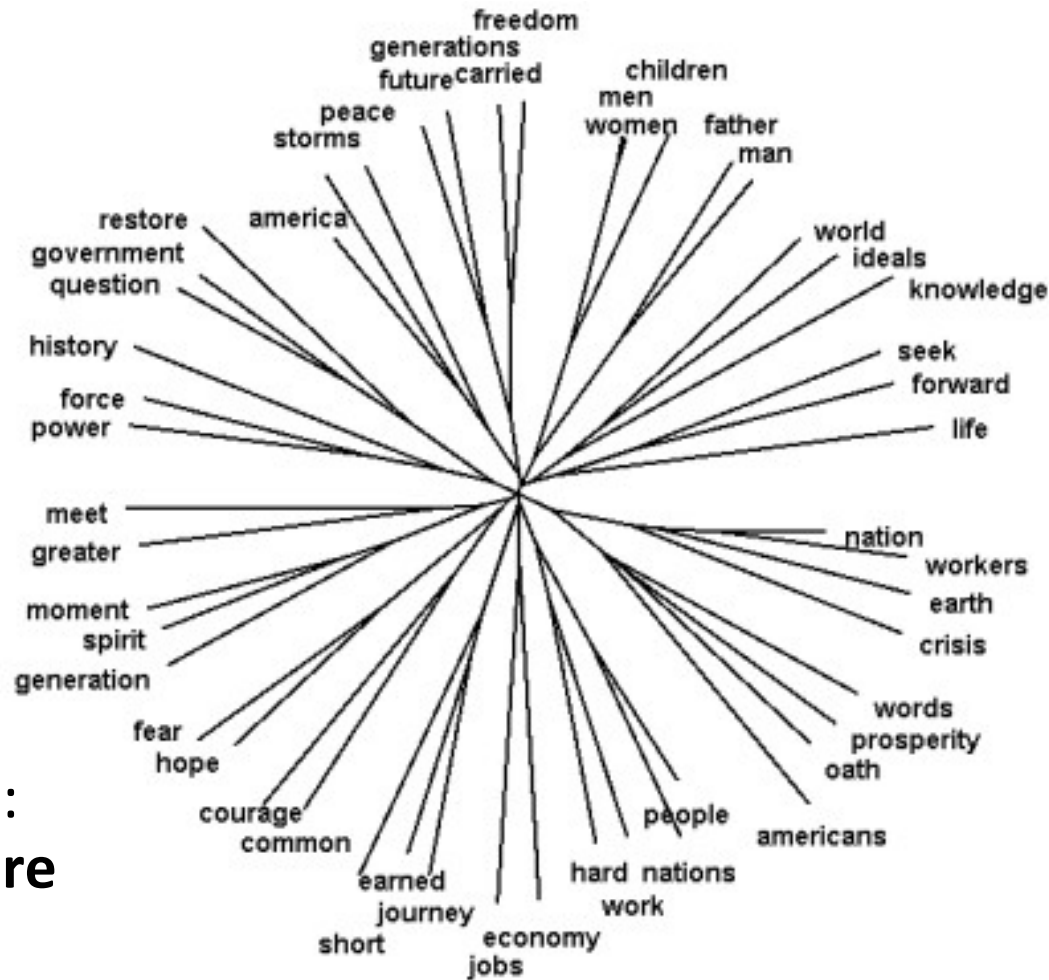


Problème 1 :
difficiles à lire



Les **distances** dans l'arbre entre deux mots reflètent *au mieux* le **degré de cooccurrence** entre ces deux mots

Interprétation réelle

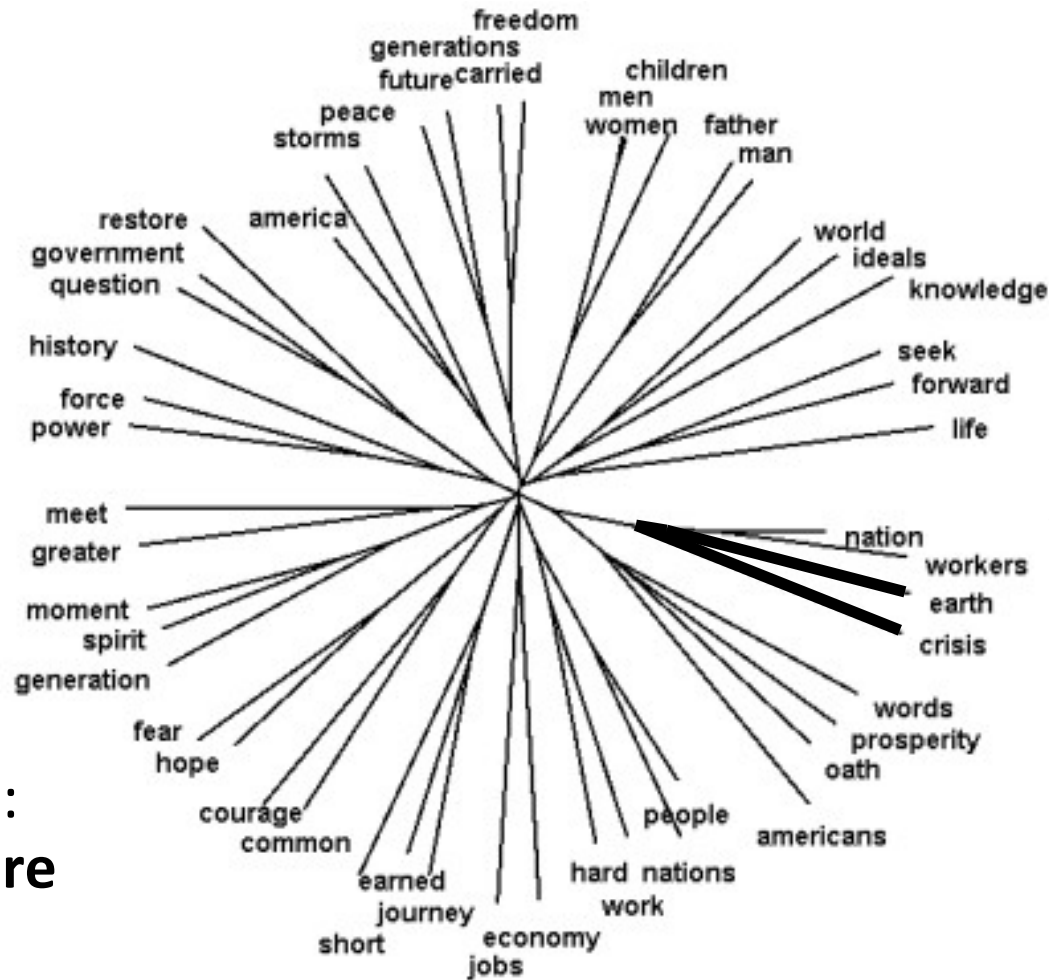


Problème 1 :
difficiles à lire

Problème 2 :
peu fiables

Les **distances dans l'arbre** entre deux mots reflètent *au mieux* le **degré de cooccurrence** entre ces deux mots

Interprétation réelle



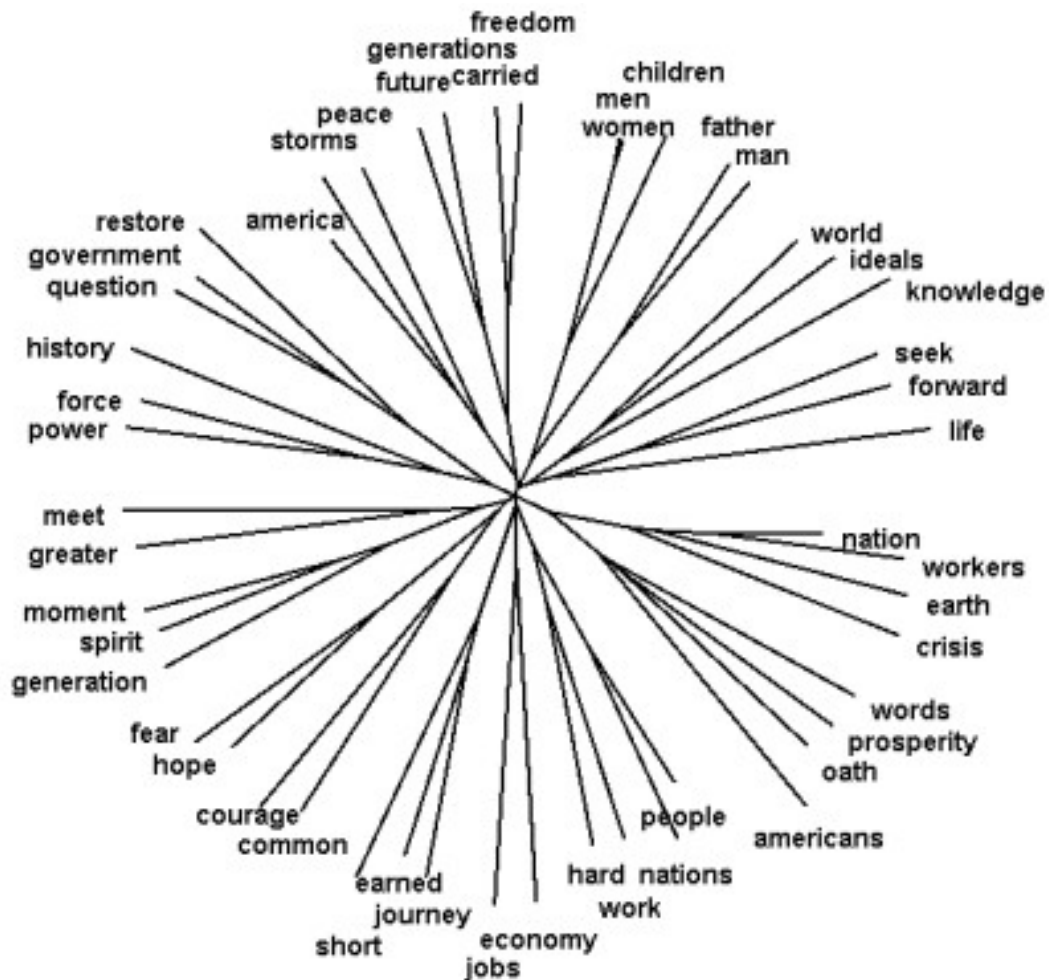
Optimisation **globale**, pas de garanties locales de qualité

Problème 1 : **difficiles à lire**

Problème 2 : **peu fiables**

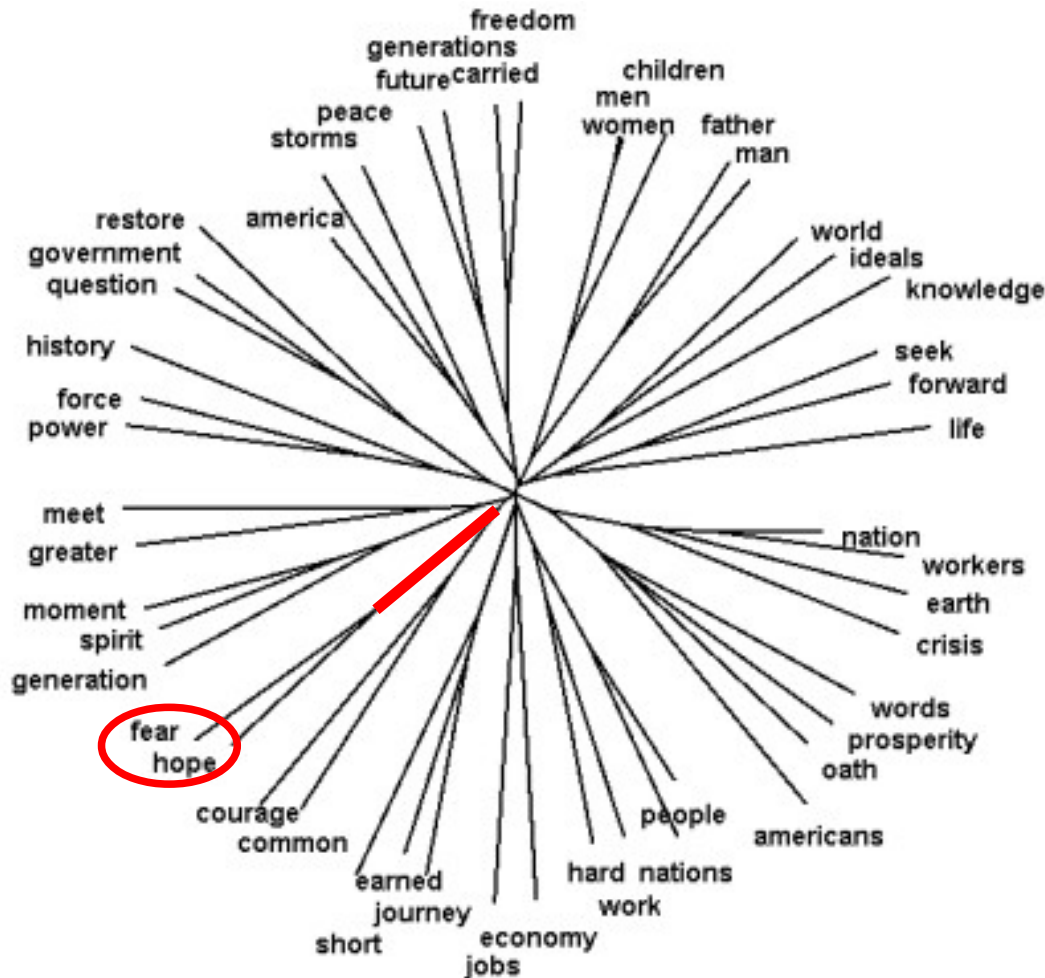
Les **distances dans l'arbre** entre deux mots reflètent *au mieux* le **degré de cooccurrence** entre ces deux mots

Interprétation pratique



arbre de distances
utilisé comme
classification

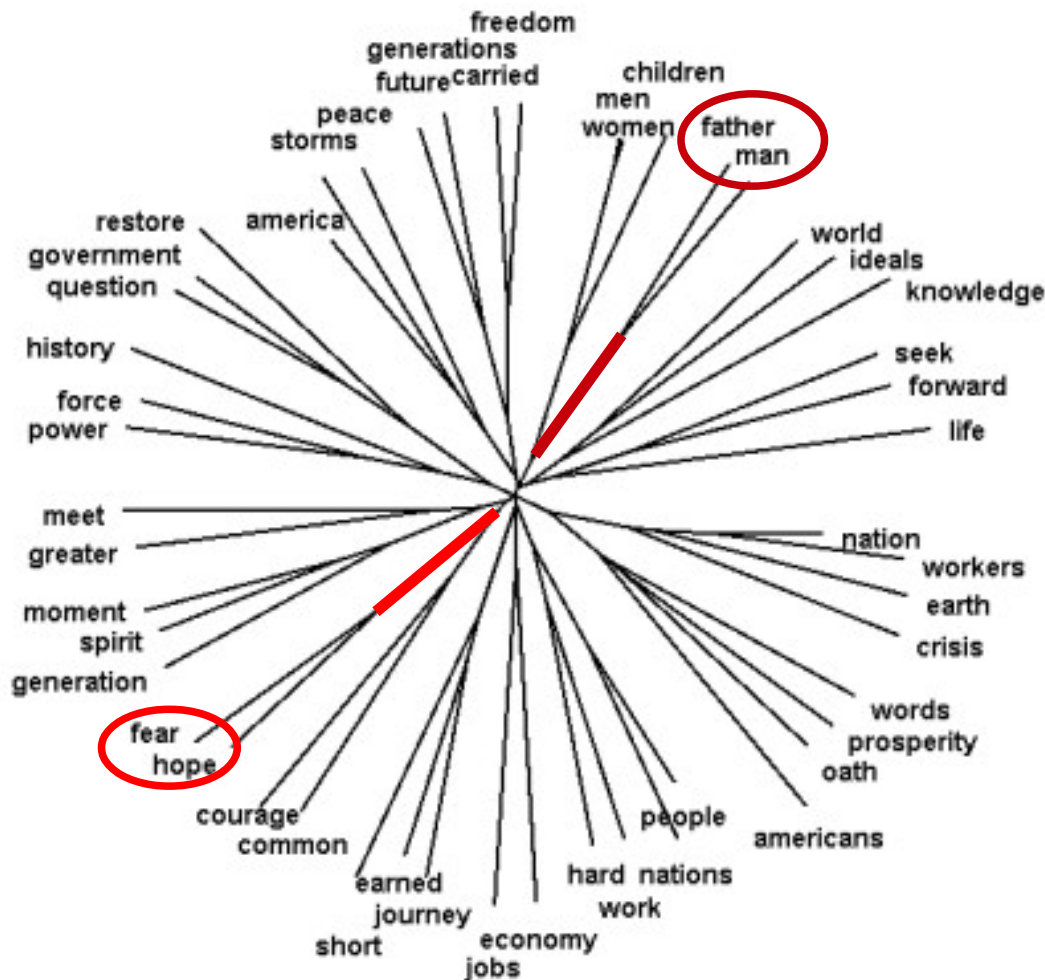
Interprétation pratique



arbre de distances
utilisé comme
classification

Les mots d'un **même sous-arbre** bien séparé du reste de l'arbre
constituent une classe de mots

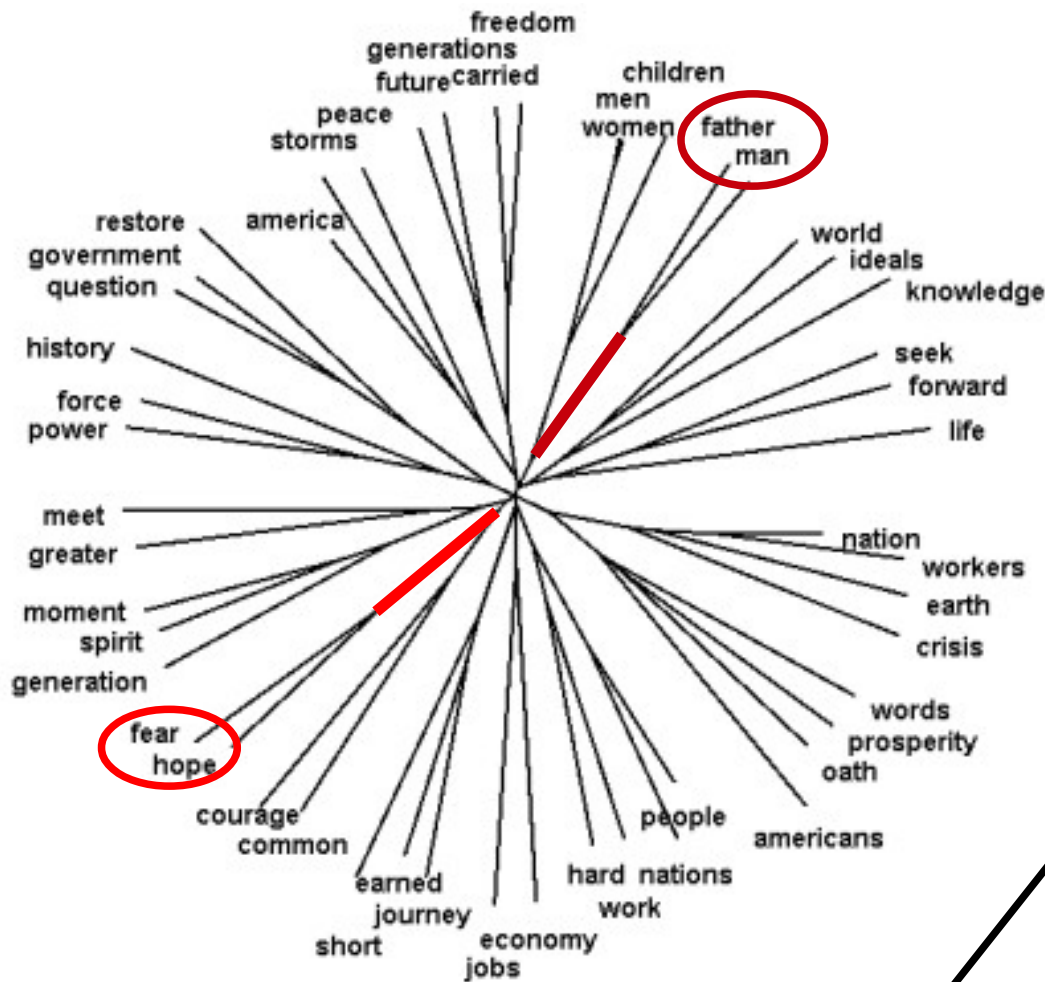
Interprétation pratique



arbre de distances
utilisé comme
classification

Les mots d'un **même sous-arbre** bien séparé du reste de l'arbre
constituent une classe de mots

Interprétation pratique



arbre de distances
utilisé comme
classification

Problème : **toujours
peu lisible** (longueur
des arêtes externes)
et **peu fiable**

Les mots d'un **même sous-arbre** bien séparé du reste de l'arbre
constituent une classe de mots

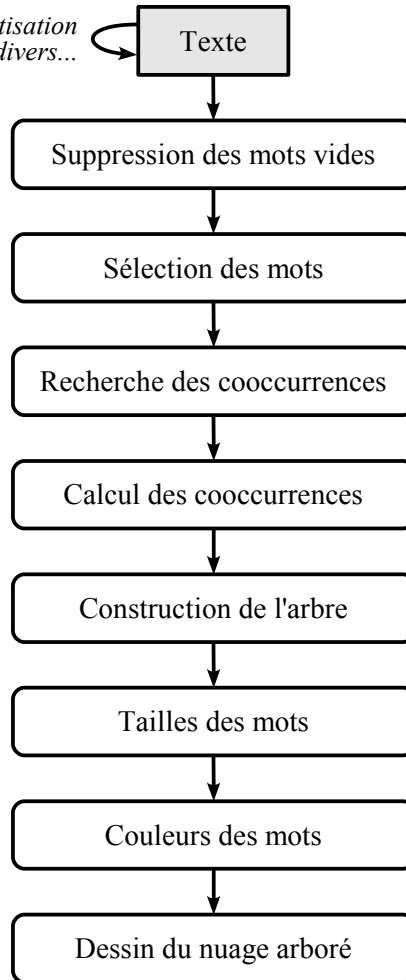
Plan

- Nuages arborés, intérêts et limites
- **Construction des nuages arborés**
- Options de coloration
- Utilisation des nuages arborés
- Prétraitements du texte utilisé
- Perspectives

Processus de construction

Import/export

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



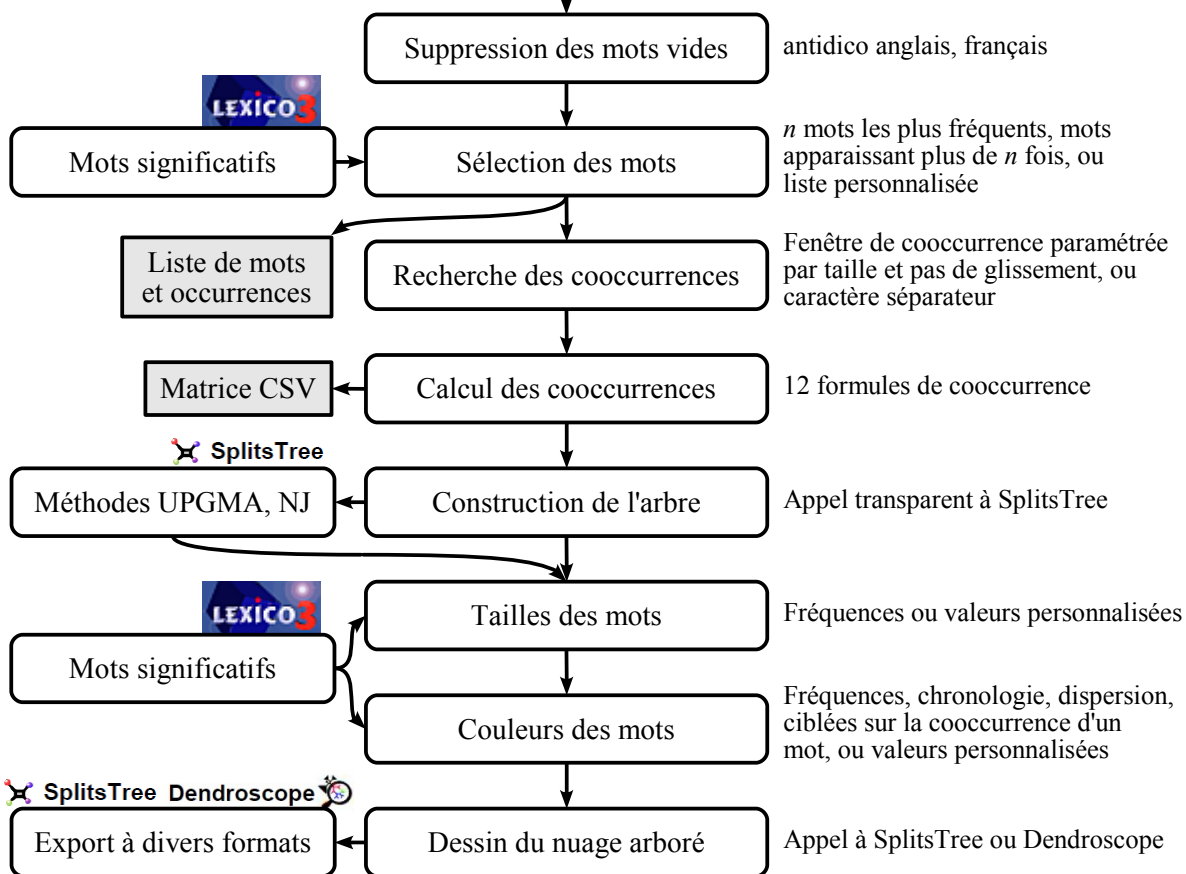
Processus de construction

Import/export

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

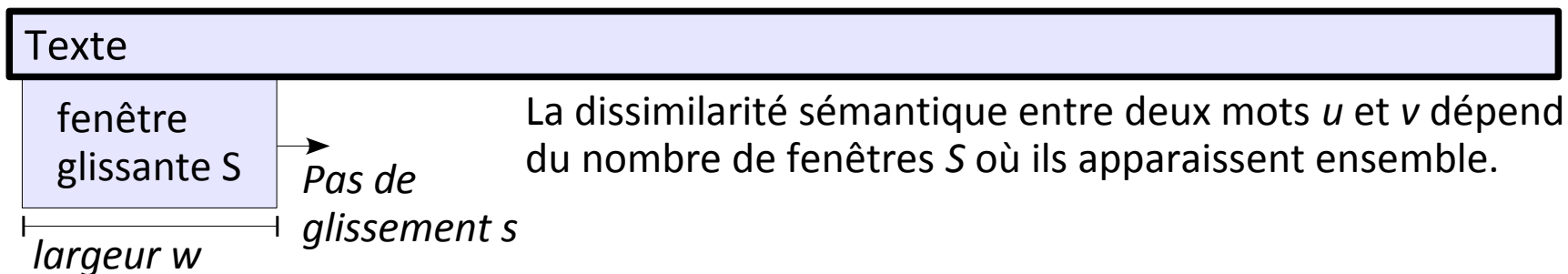
Texte

Proposé dans TreeCloud



Calcul des scores de cooccurrence

Calcul de la matrice de distance entre mots



matrices de cooccurrence

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

Pour 2 mots u et v	$v \in S$	$v \notin S$
$u \in S$	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...

Calcul des distances de cooccurrence

Les formules statistiques fournissent un score de similarité.

Comment obtenir des dissimilarités, dans l'intervalle $[0,1]$?

$$\text{dissimilarité} = 1 - \text{similarité normalisée sur } [0,1]$$

Normalisation des scores de similarité sur $[0,1]$:

- normalisation linéaire pour les matrices positives
- normalisation affines pour les matrices contenant des valeurs négatives, afin d'obtenir des distances dans l'intervalle $[a,1]$ ($a=0.1$)

Construction de l'arbre

Plusieurs méthodes pour construire un arbre à partir d'une matrice de distances (classification hiérarchique) :

- Neighbor-Joining

Saitou & Nei, 1987



- Variantes d'Addtree

Barthélemy & Luong, 1987

- Heuristique des quadruplets

Cilibrasi & Vitanyi, 2007

Décoration de l'arbre

Tailles des mots :

- calculées directement à partir des **fréquences**
(avec un log!)
- calculées à partir des **rangs des fréquences**
(distribution exponentielle)
- **score de spécificité** par rapport à un corpus de référence
(TF-IDF, écart réduit...)

Dessin de l'arbre

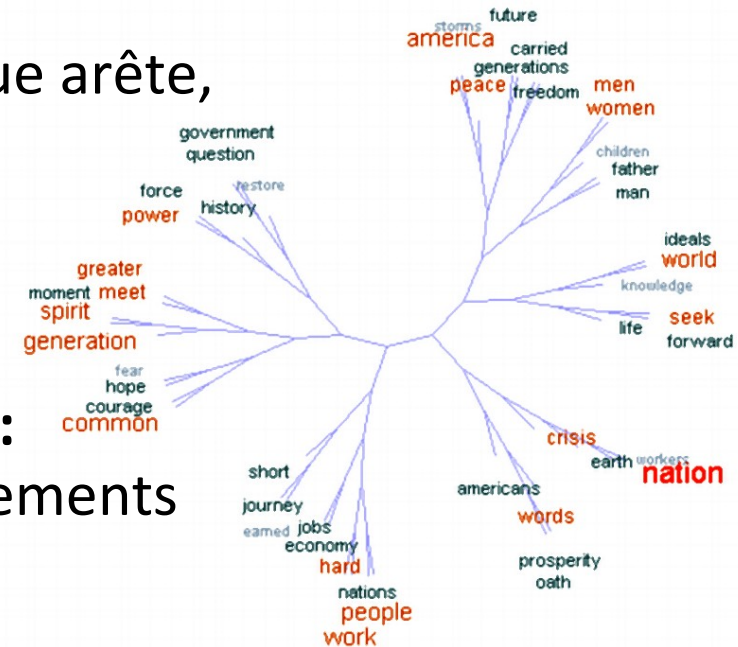
Algorithme “equal angle” :

- montant pour calculer l'angle de chaque arête, en partant des feuilles
- descendant pour placer chaque arête en partant d'un sommet interne

Placement automatique des étiquettes :

→ heuristique pour éviter les chevauchements

Question des longueurs d'arêtes ?



Dessin de l'arbre

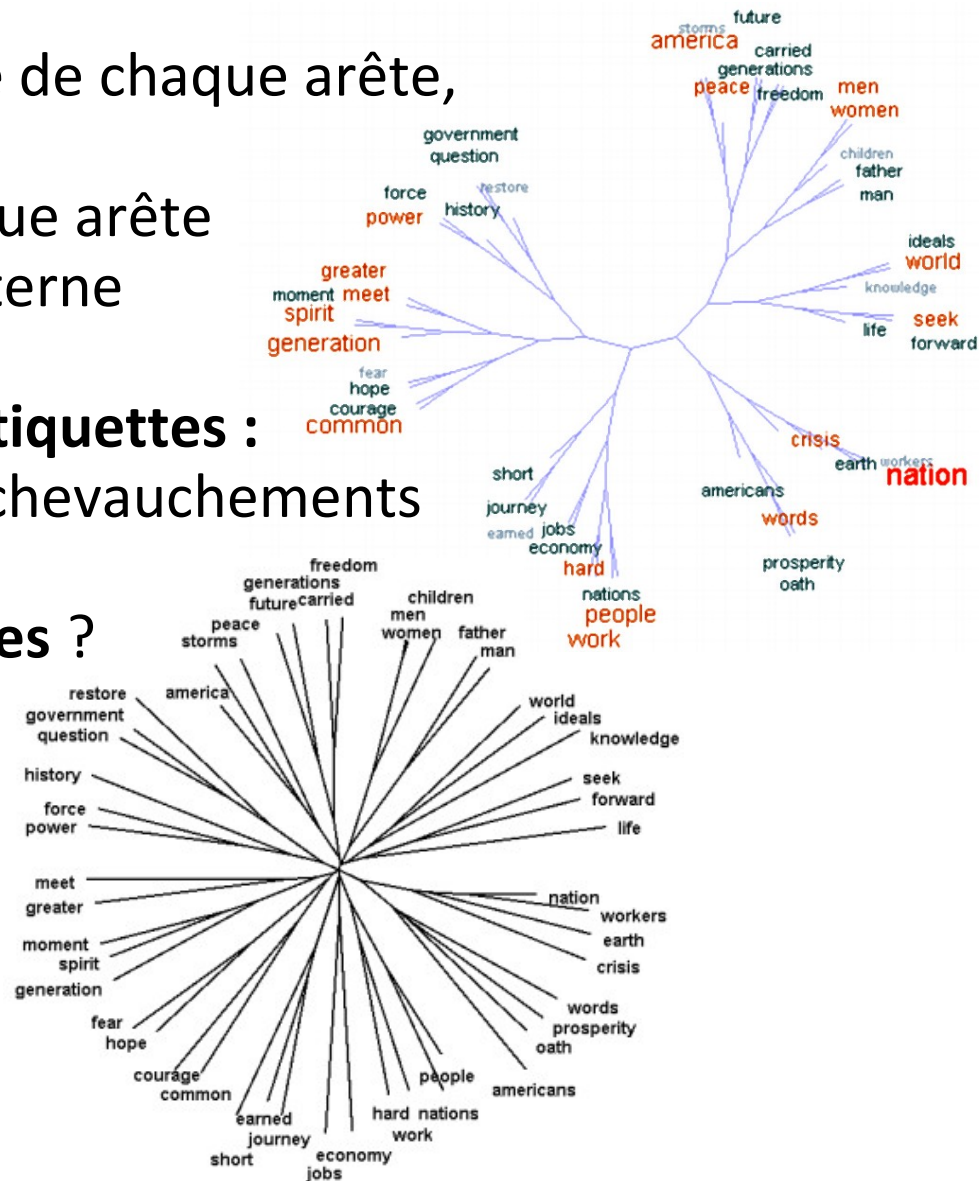
Algorithme "equal angle" :

- montant pour calculer l'angle de chaque arête, en partant des feuilles
- descendant pour placer chaque arête en partant d'un sommet interne

Placement automatique des étiquettes :

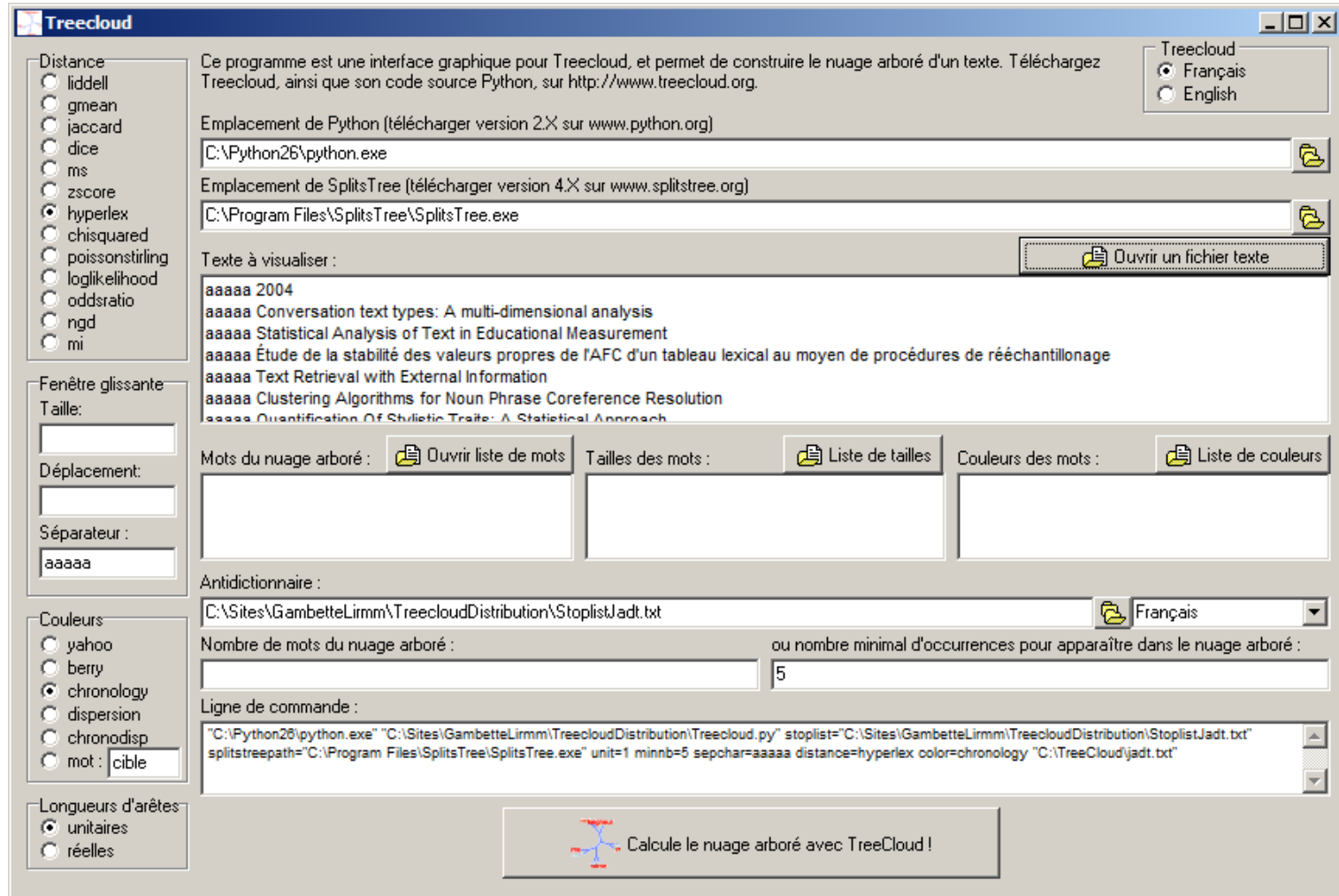
→ heuristique pour éviter les chevauchements

Question des longueurs d'arêtes ?



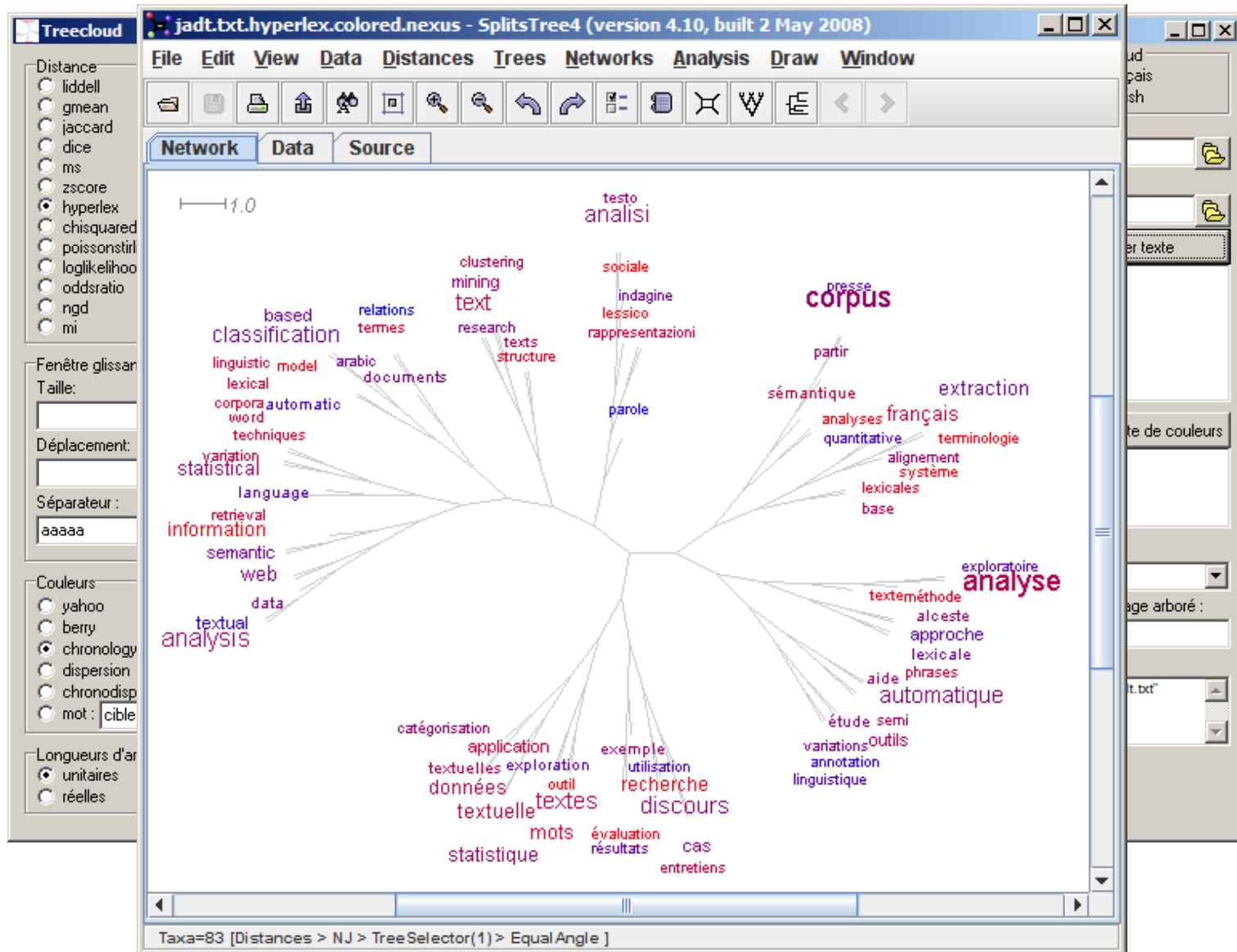
Implémentations

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](#)).



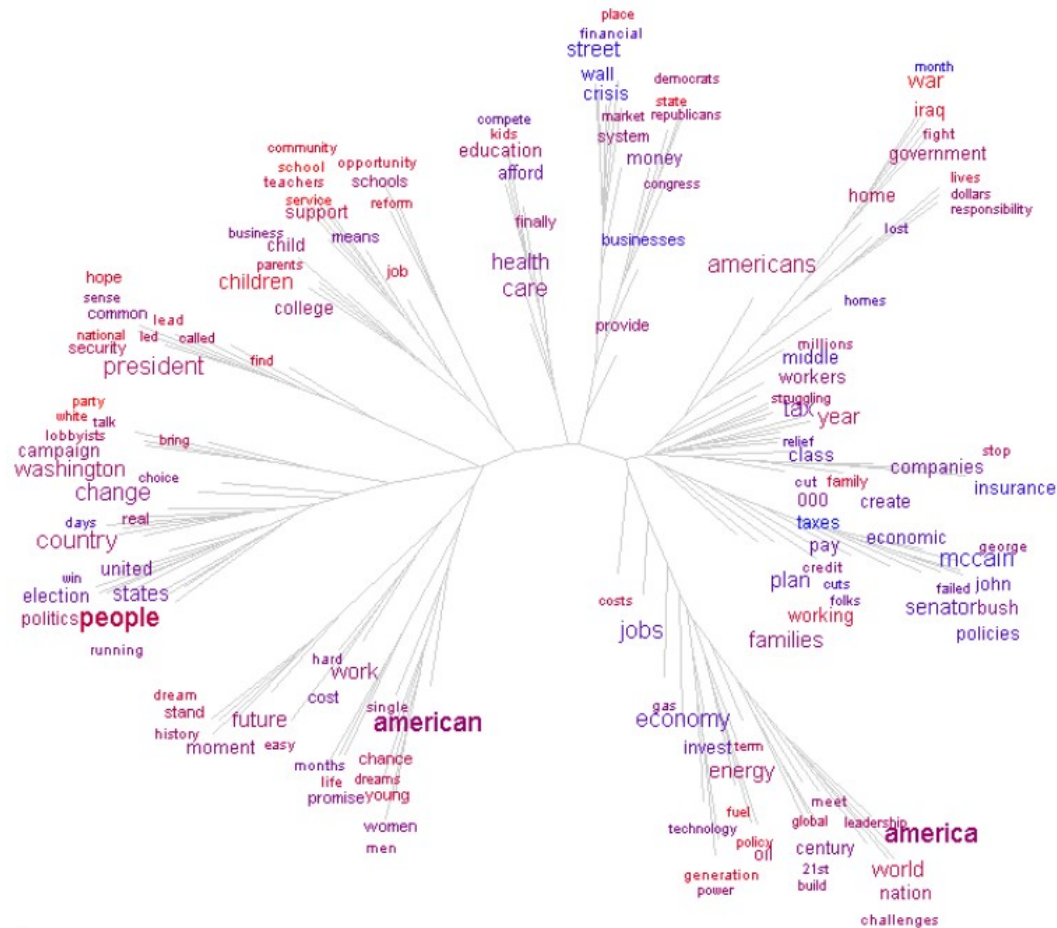
www.treecloud.org

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

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

Temps d'exécution

Limites sur la taille du corpus pour utiliser TreeCloud ?



30 secondes pour la construction du nuage arboré de l'ensemble des discours de campagne de Barack Obama (>300 000 mots)

Plan

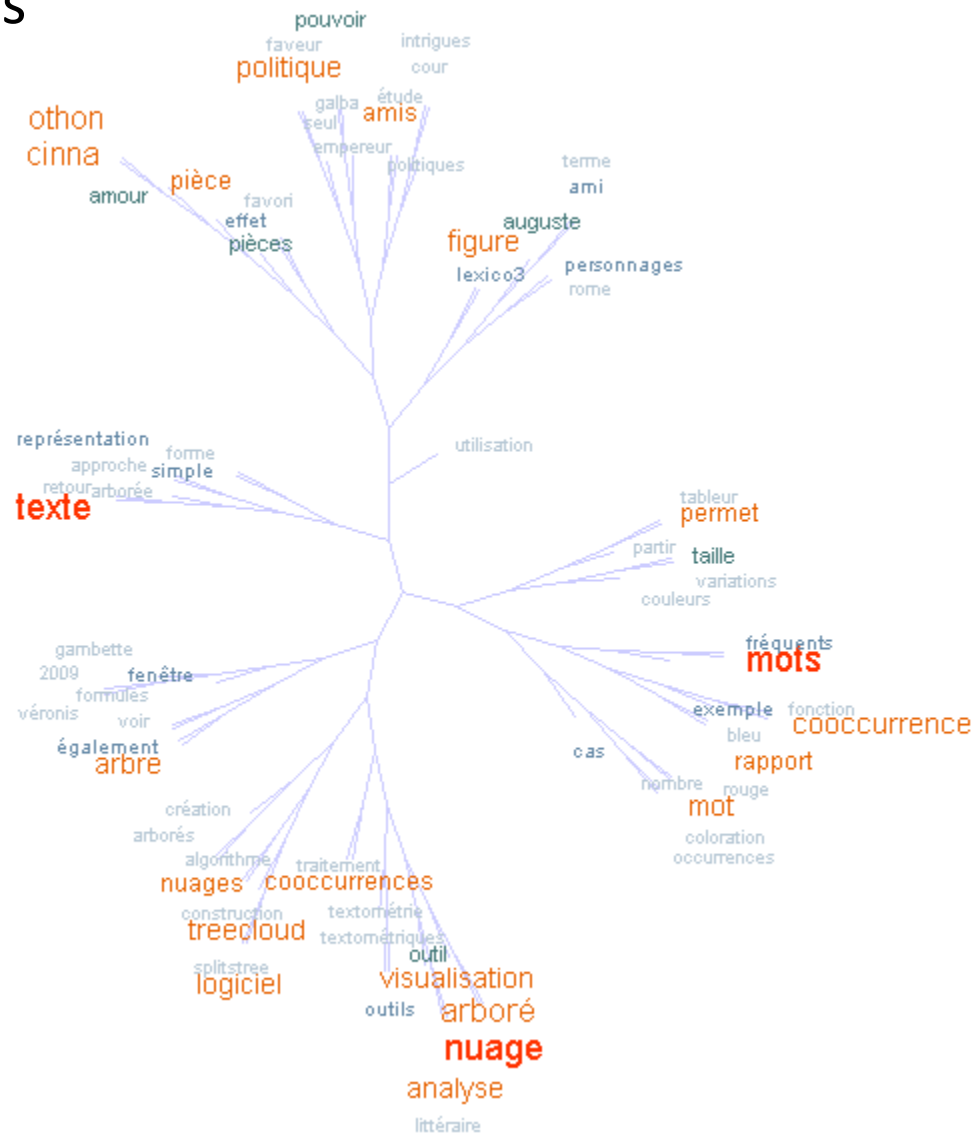
- Nuages arborés, intérêts et limites
- Construction des nuages arborés
- Options de coloration
- Utilisation des nuages arborés
- Prétraitements du texte utilisé
- Perspectives

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

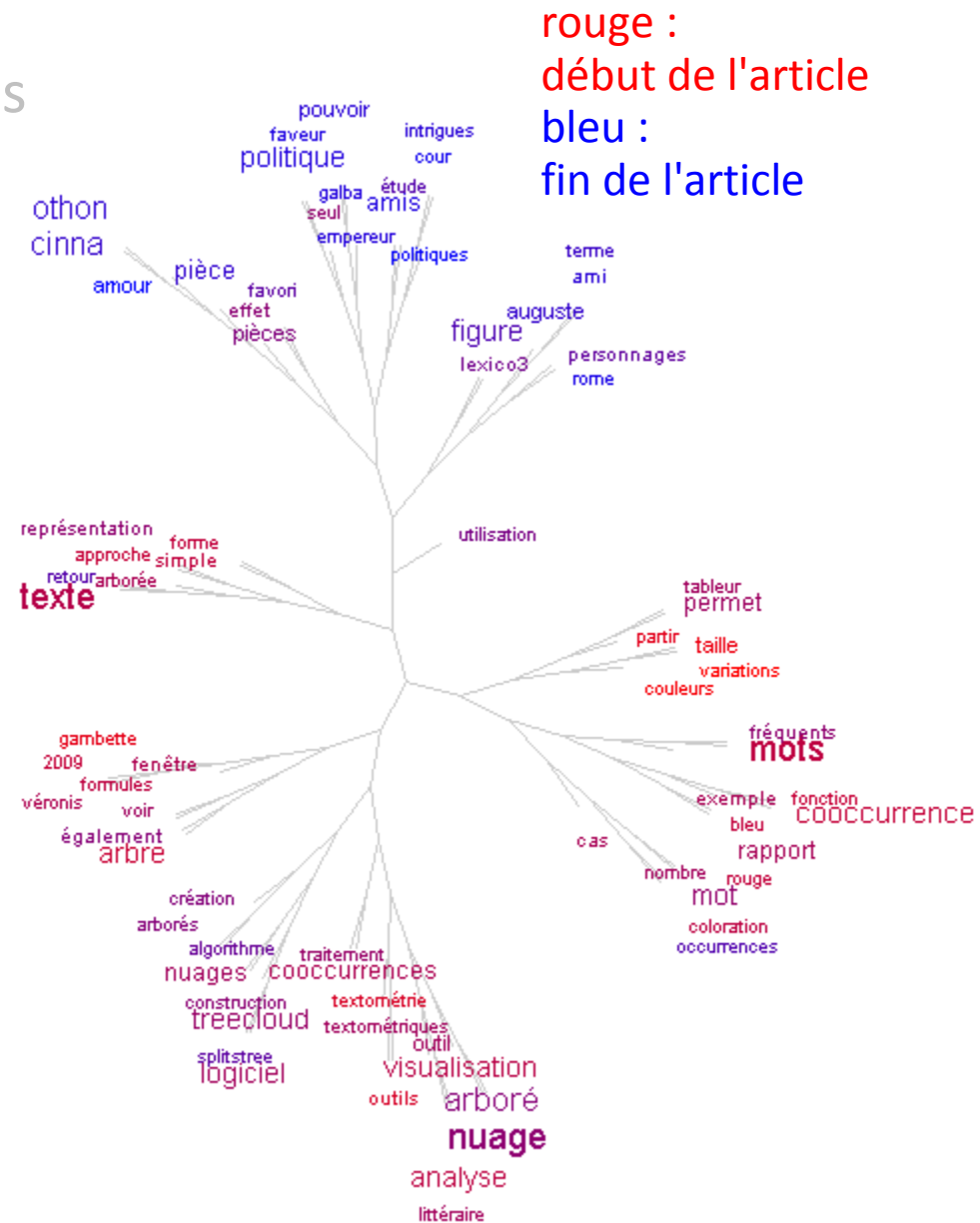


Nuage arboré des mots apparaissant 5 fois ou plus dans l'article d'Amstutz & Gambette, JADT 2010, distance Liddell, fenêtre de 20 mots, coloration Yahoo

Des couleurs pour guider la lecture

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

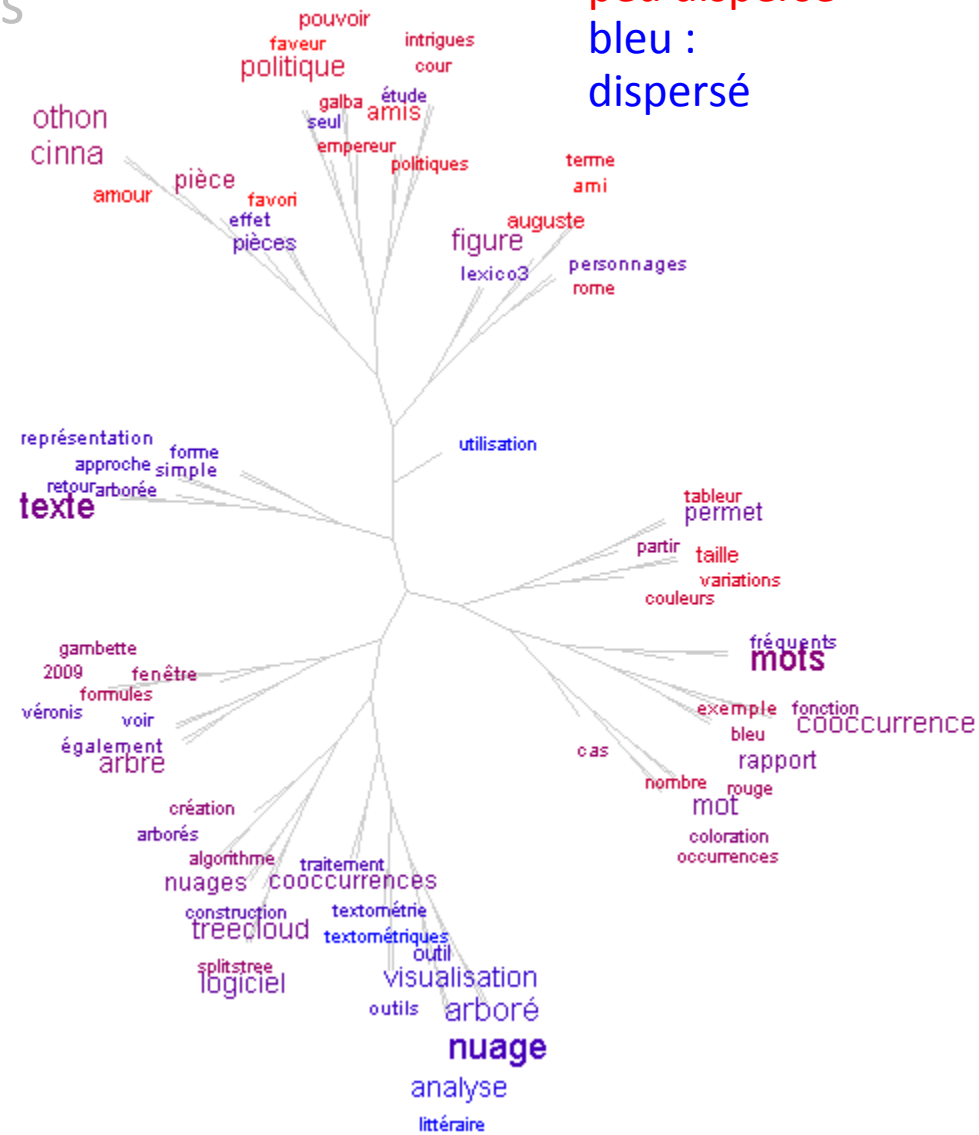
Nuage arboré des mots apparaissant 5 fois ou plus dans l'article d'Amstutz & Gambette, JADT 2010, distance Liddell, fenêtre de 20 mots, coloration chronologique



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 l'article d'Amstutz &
Gambette, JADT 2010,
distance Liddell, fenêtre de
20 mots, coloration
dispersion*

Plan

- Nuages arborés, intérêts et limites
- Construction des nuages arborés
- Options de coloration
- **Utilisations** des nuages arborés
- Prétraitements du texte utilisé
- Perspectives

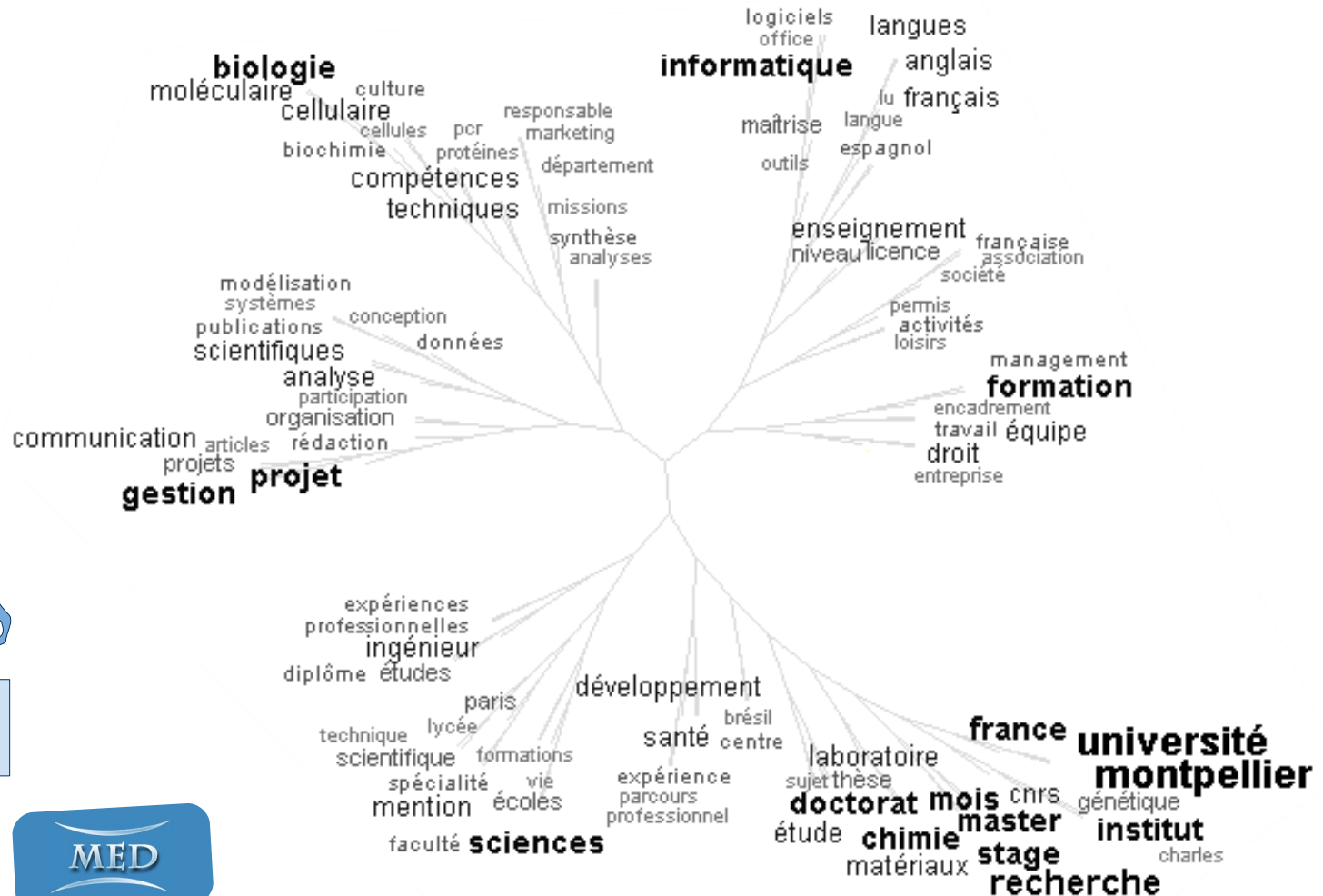
Utilisations des nuages arborés

- **Résumé visuel** des thématiques d'un texte
- Support de **médiation** et d'**argumentation** :
 - **appui visuel** d'une analyse subjective
 - **clarification** de rapports ou discours, lors de la rédaction
- En **analyse textuelle** (réponses aux questions ouvertes, romans, théâtre, corpus médiatique, etc.) :
 - 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**

Support de médiation et d'argumentation

Présenter les compétences des docteurs

Corpus : CV soumis à une rencontre docteurs-entreprises



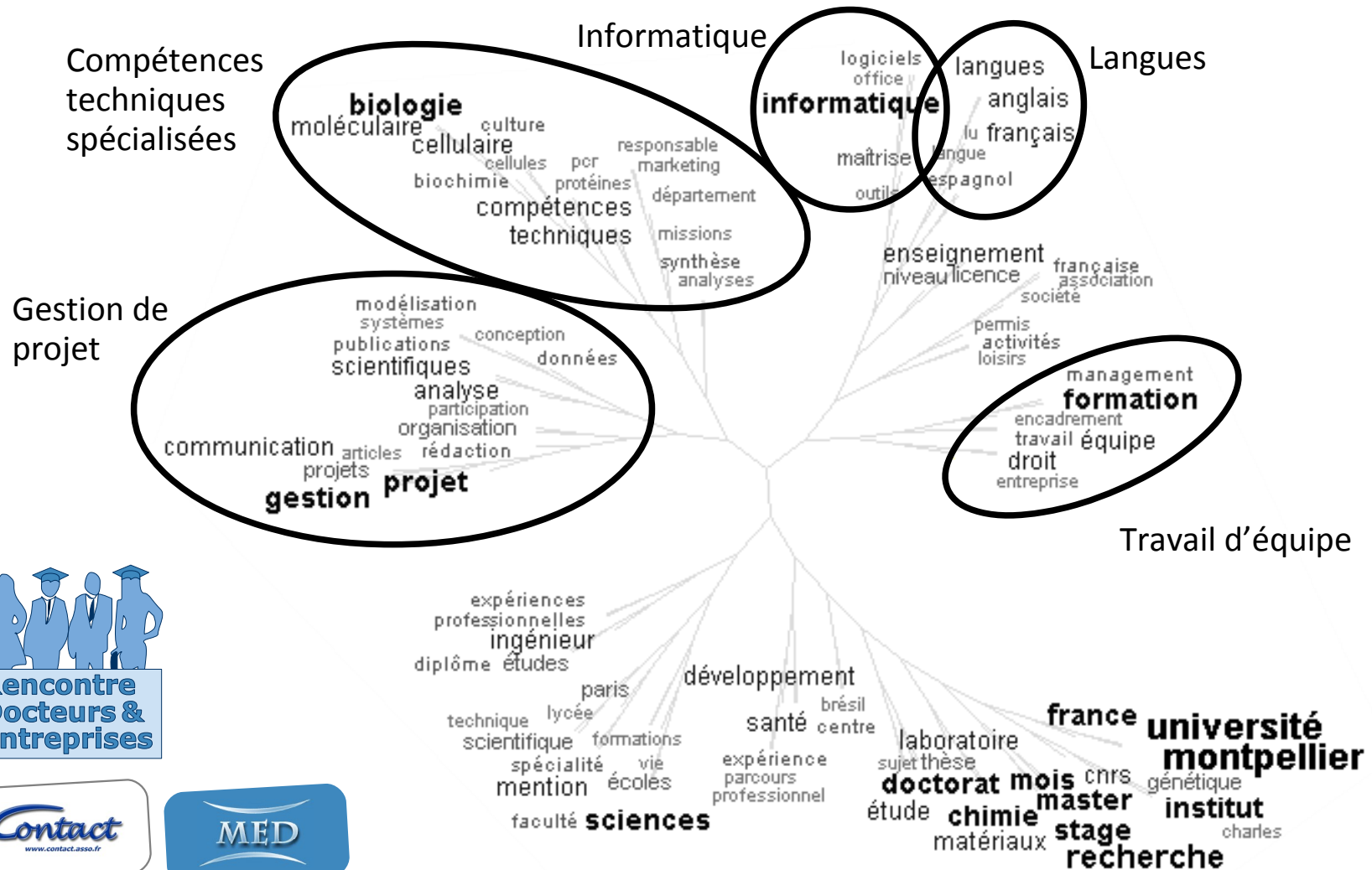
Rencontre
Docteurs &
Entreprises



Support de médiation et d'argumentation

Préserver les compétences des docteurs

Corpus : CV soumis à une rencontre docteurs-entreprises



Rencontre
Docteurs &
Entreprises

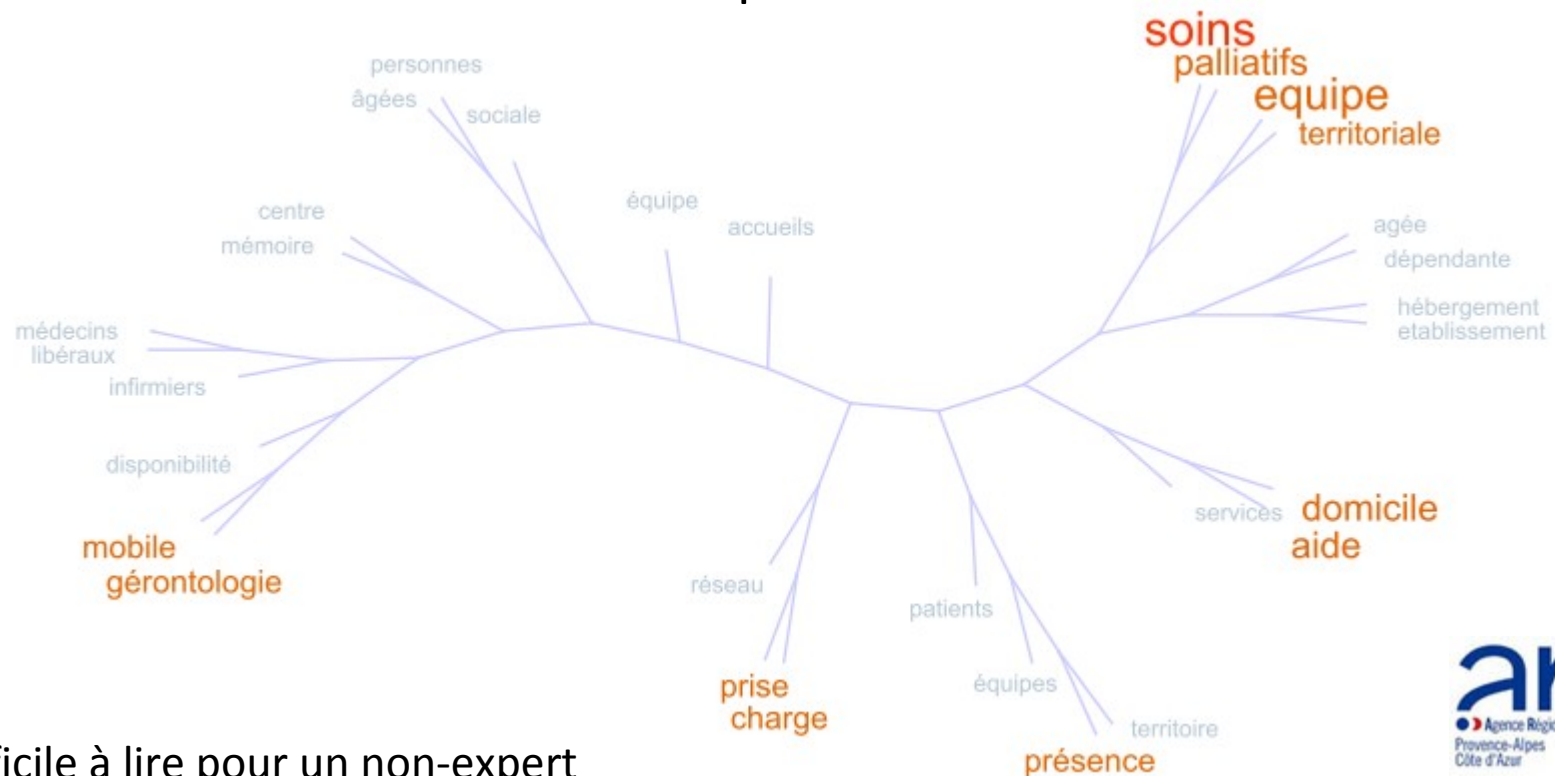


Support de médiation et d'argumentation

Transmettre les résultats d'une consultation

Corpus : réponses à des questions ouvertes à des professionnels de la santé sur le parcours de santé des personnes âgées dans les Alpes de Haute-Provence

Points forts de l'accueil dans le département



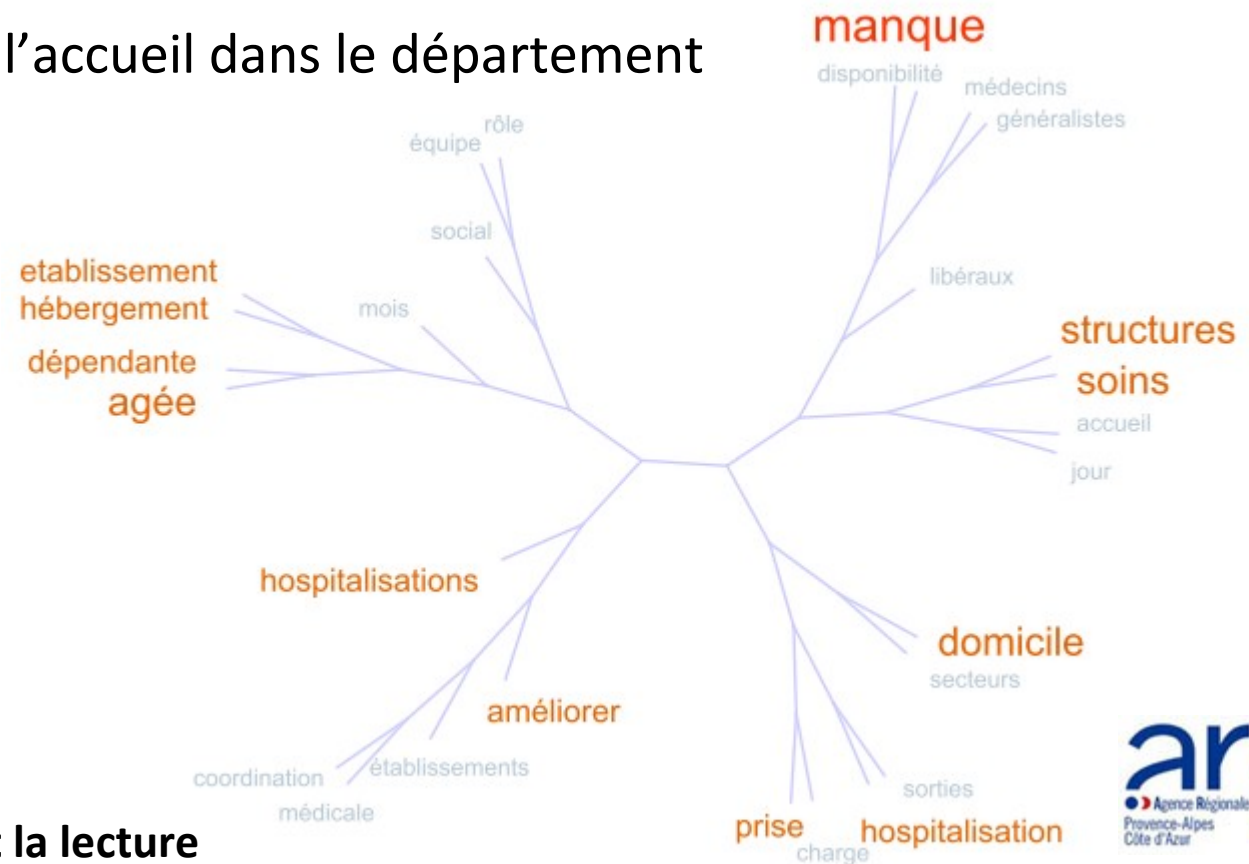
Difficile à lire pour un non-expert
Support cohérent d'accompagnement du discours

Support de médiation et d'argumentation

Transmettre les résultats d'une consultation

Corpus : réponses à des questions ouvertes à des professionnels de la santé sur le parcours de santé des personnes âgées dans les Alpes de Haute-Provence

Points faibles de l'accueil dans le département



Modaux qui guident la lecture

Support cohérent d'accompagnement du discours

Support de médiation et d'argumentation

Transmettre les résultats d'une consultation

Corpus : réponses à des questions ouvertes à des professionnels de la santé sur le parcours de santé des personnes âgées dans les Alpes de Haute-Provence

Points faibles de l'accueil dans le département



Modaux qui guident la lecture

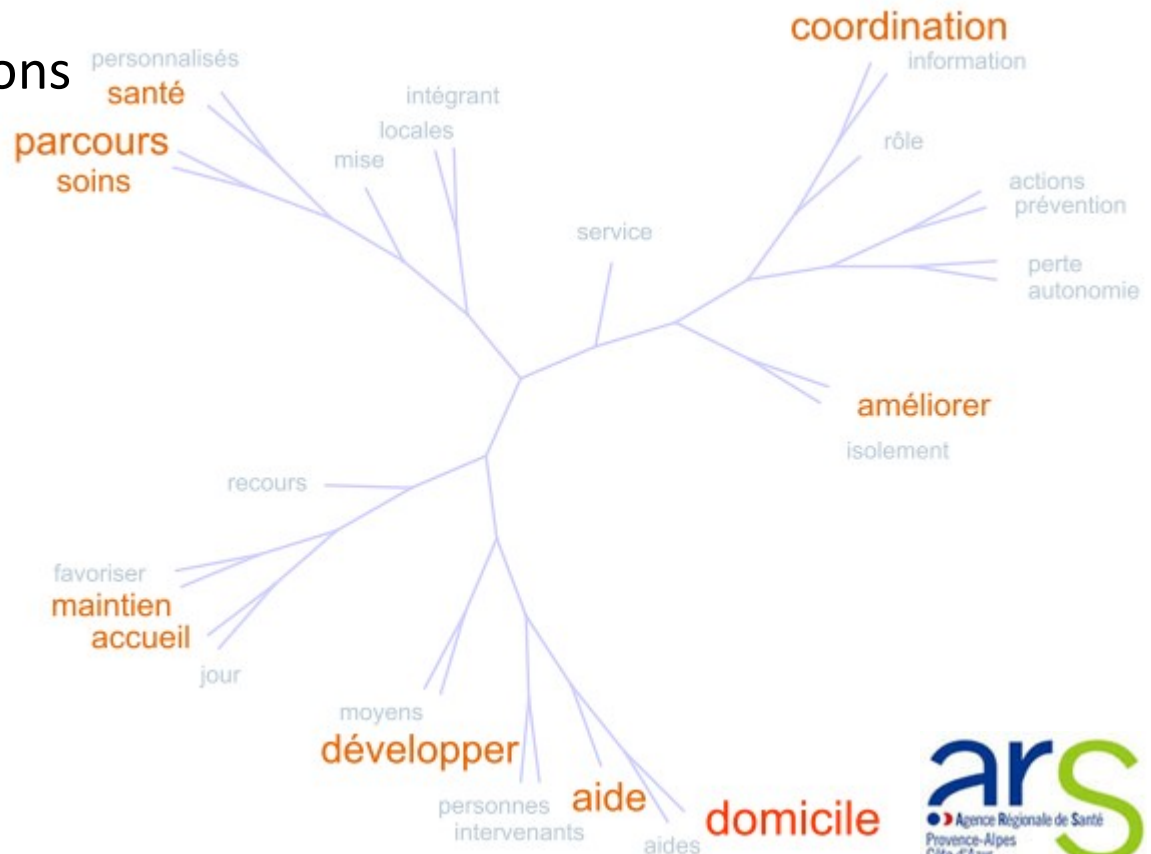
Support cohérent d'accompagnement du discours

Support de médiation et d'argumentation

Transmettre les résultats d'une consultation

Corpus : réponses à des questions ouvertes à des professionnels de la santé sur le parcours de santé des personnes âgées dans les Alpes de Haute-Provence

Suggestions d'améliorations



Verbes qui guident la lecture

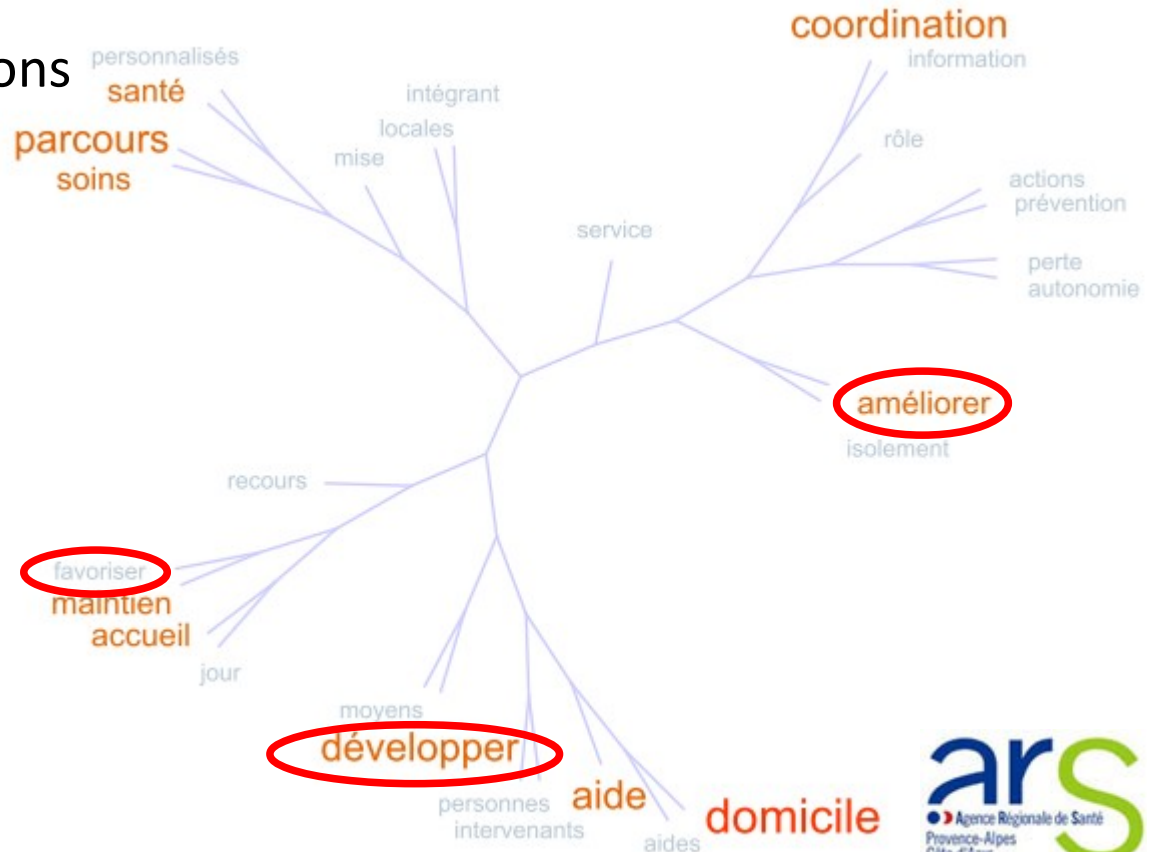
Support cohérent d'accompagnement du discours

Support de médiation et d'argumentation

Transmettre les résultats d'une consultation

Corpus : réponses à des questions ouvertes à des professionnels de la santé sur le parcours de santé des personnes âgées dans les Alpes de Haute-Provence

Suggestions d'améliorations

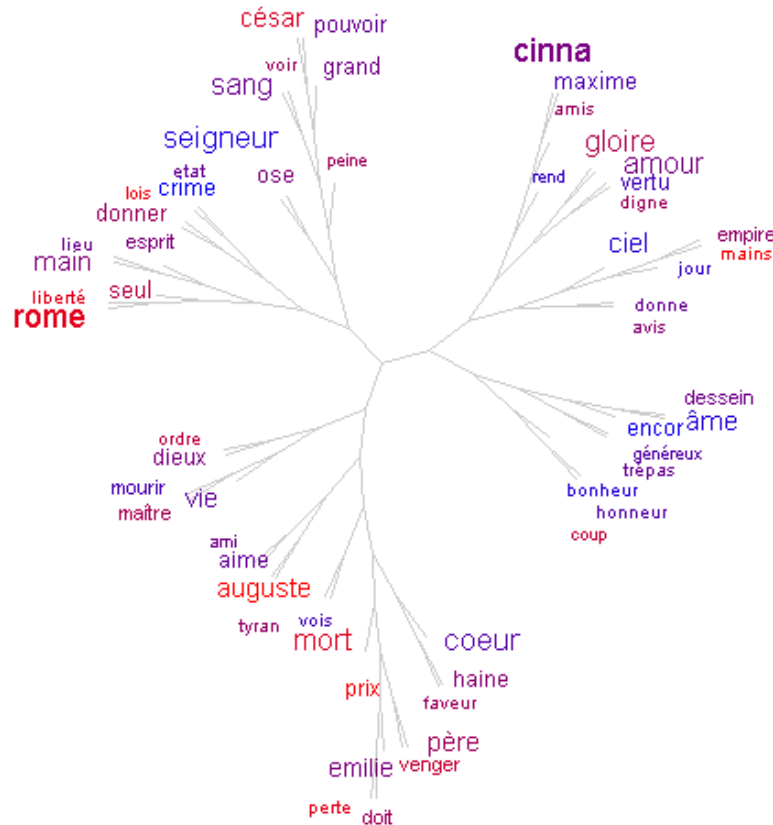


Verbes qui guident la lecture

Support cohérent d'accompagnement du discours

Analyse littéraire : illustration sur *Cinna*

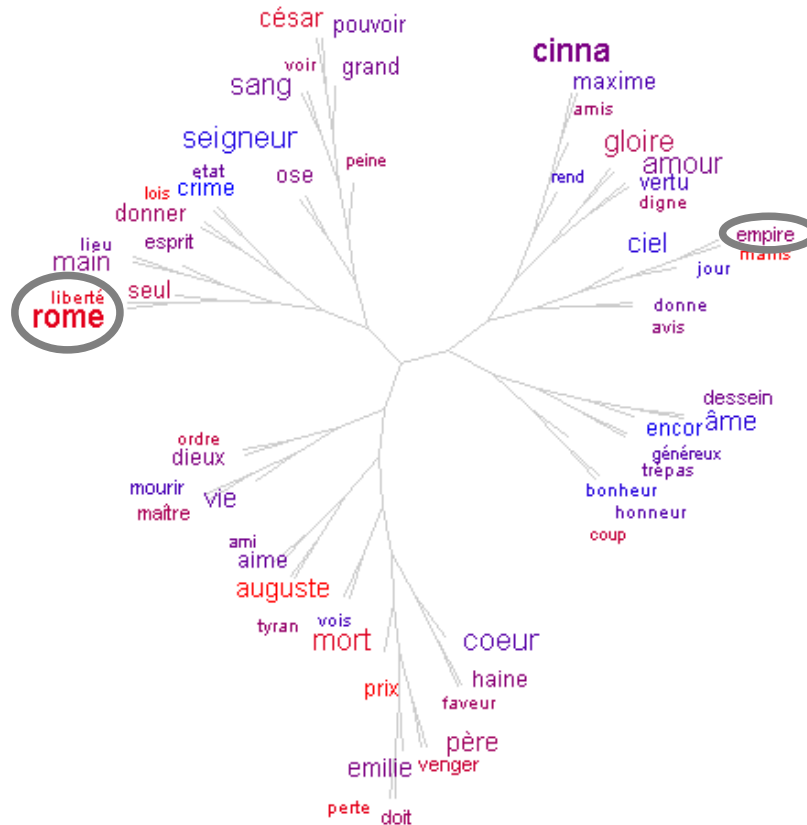
Amstutz & Gambette,
JADT 2010



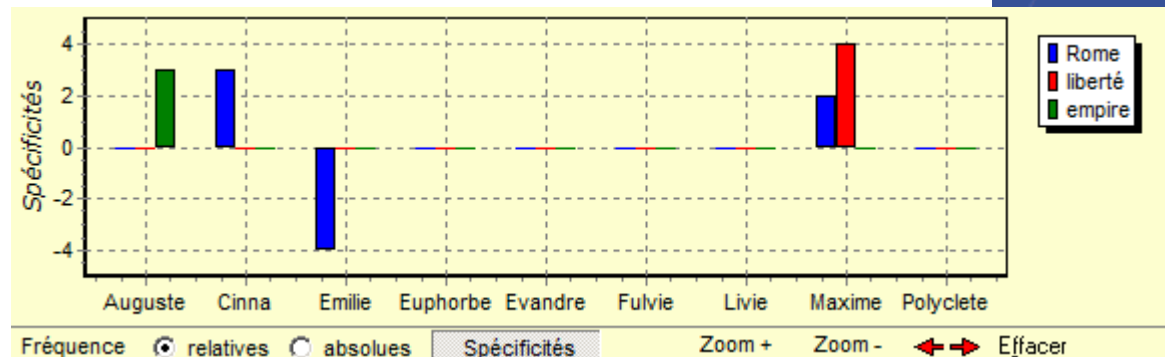
Nuages arborés globaux des 60 mots les plus fréquents dans Cinna de Corneille (distance Liddell, fenêtre de largeur 20), colorés chronologiquement (rouge au début, bleu à la fin)

Analyse littéraire : illustration sur *Cinna*

Amstutz & Gambette,
JADT 2010

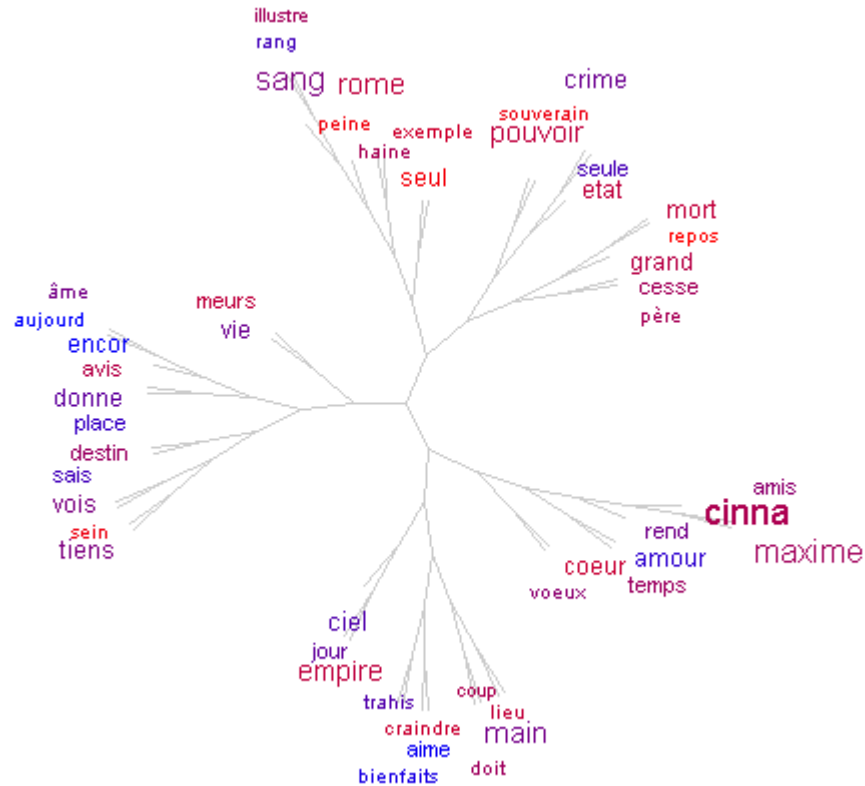


Spécificités d'emploi de « Rome », « liberté » et « empire » chez les différents personnages de Cinna dans Lexico3.



Analyse littéraire : illustration sur *Cinna*

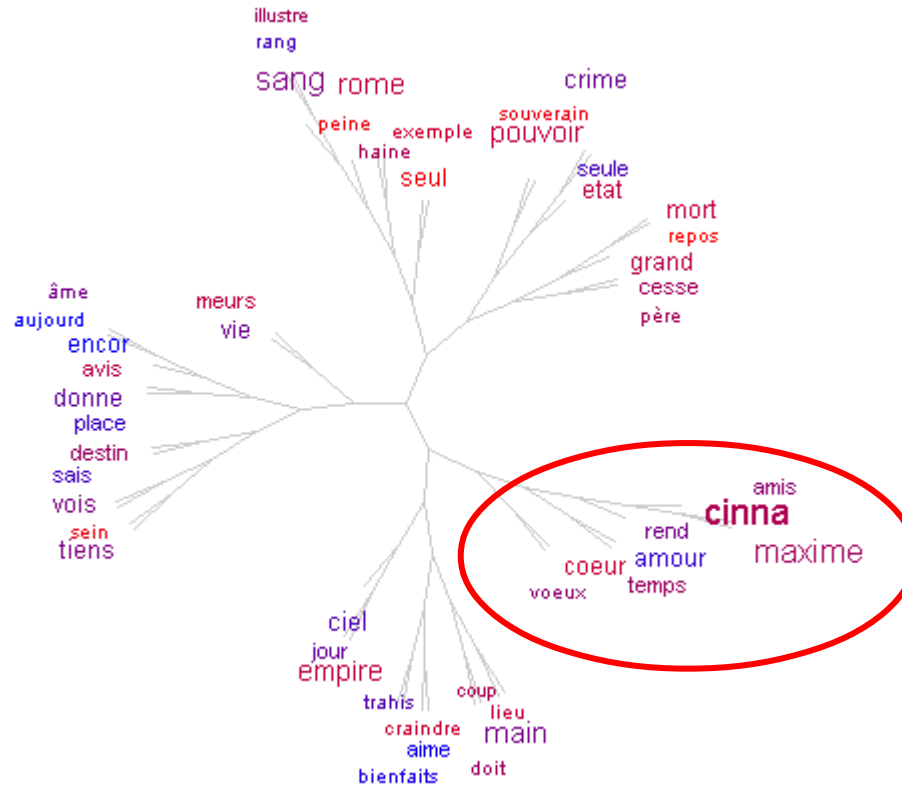
Amstutz & Gambette,
JADT 2010



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

Analyse littéraire : illustration sur *Cinna*

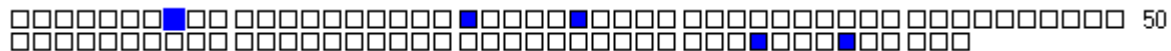
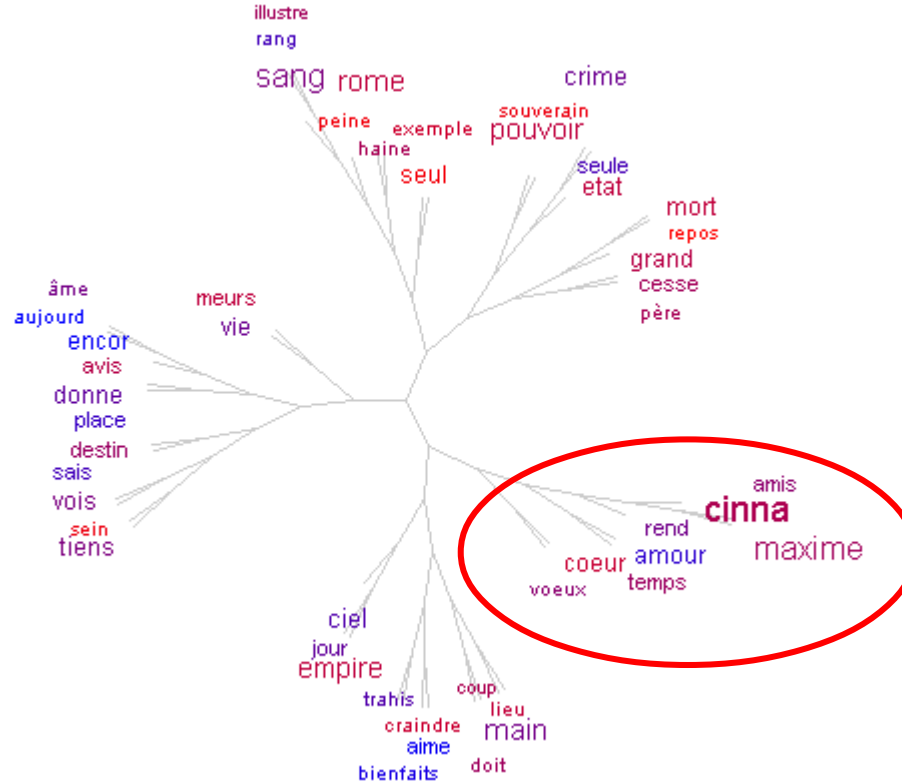
Amstutz & Gambette,
JADT 2010



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

Analyse littéraire : illustration sur *Cinna*

Amstutz & Gambette,
JADT 2010

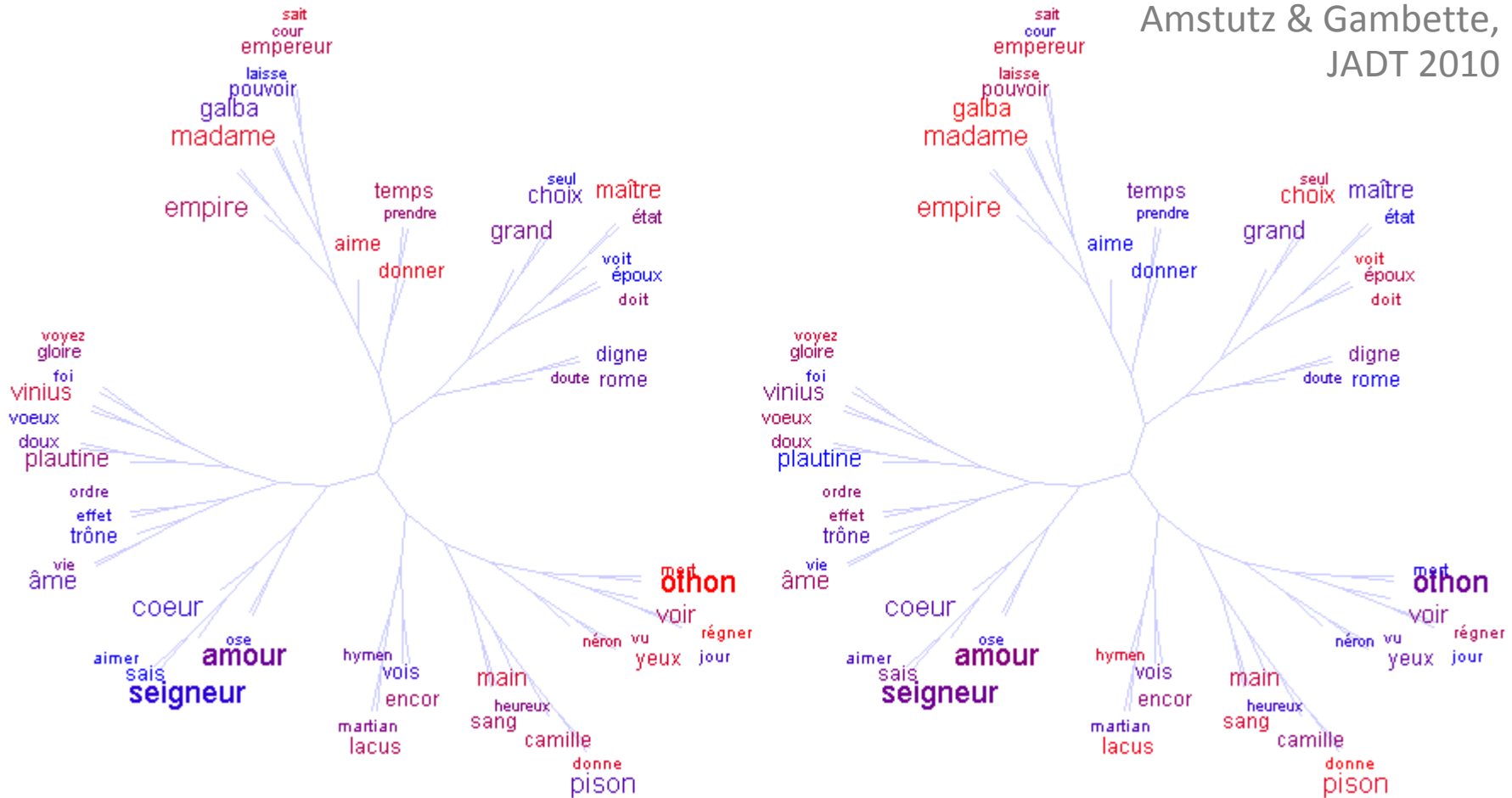


Carte des sections Lexico3 et contextes de « amis » dans les paroles d'Auguste dans *Cinna*.

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**.

Analyse littéraire : illustration sur *Othon*

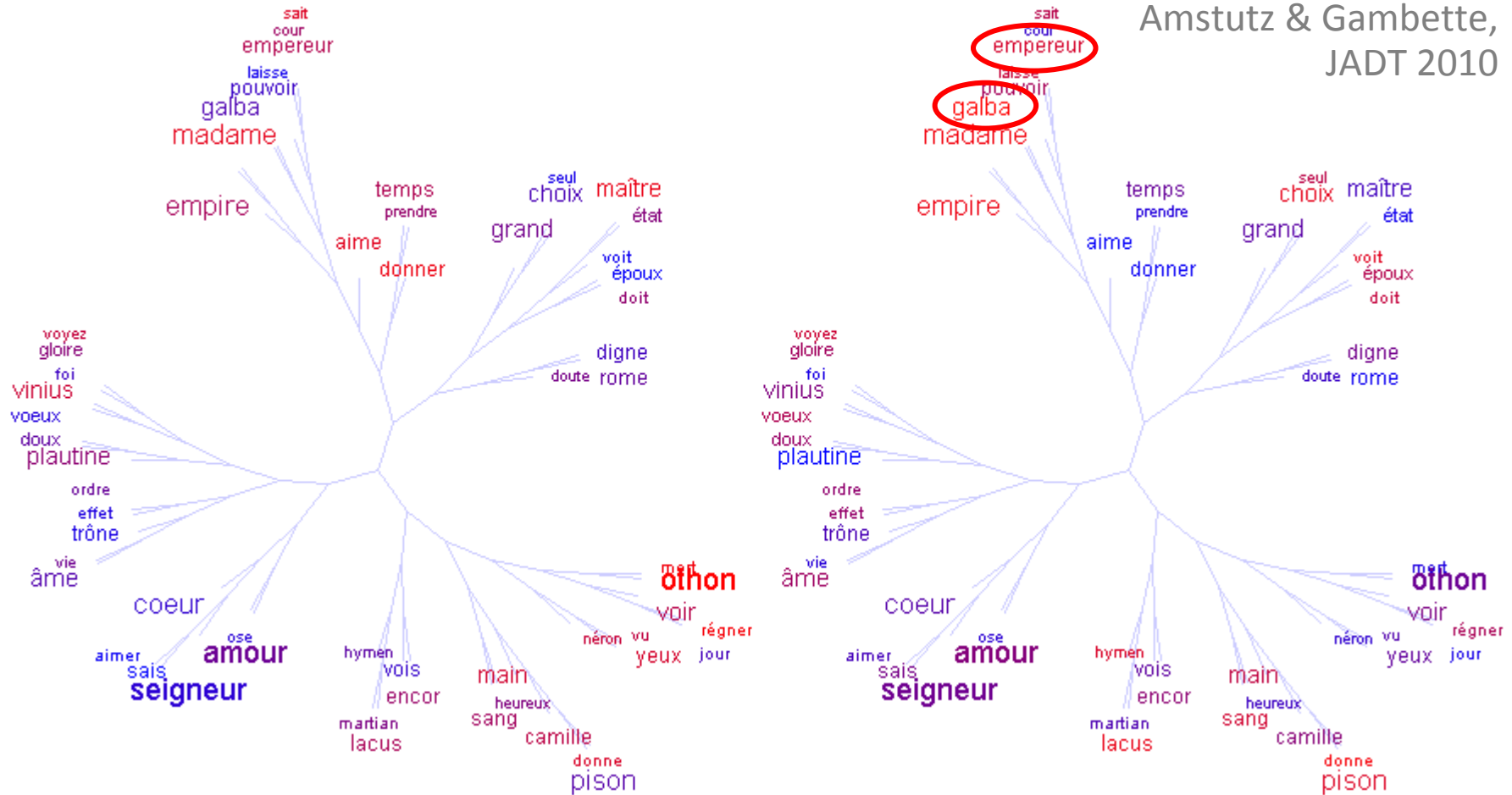
Amstutz & Gambette,
JADT 2010



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 »

Analyse littéraire : illustration sur *Othon*

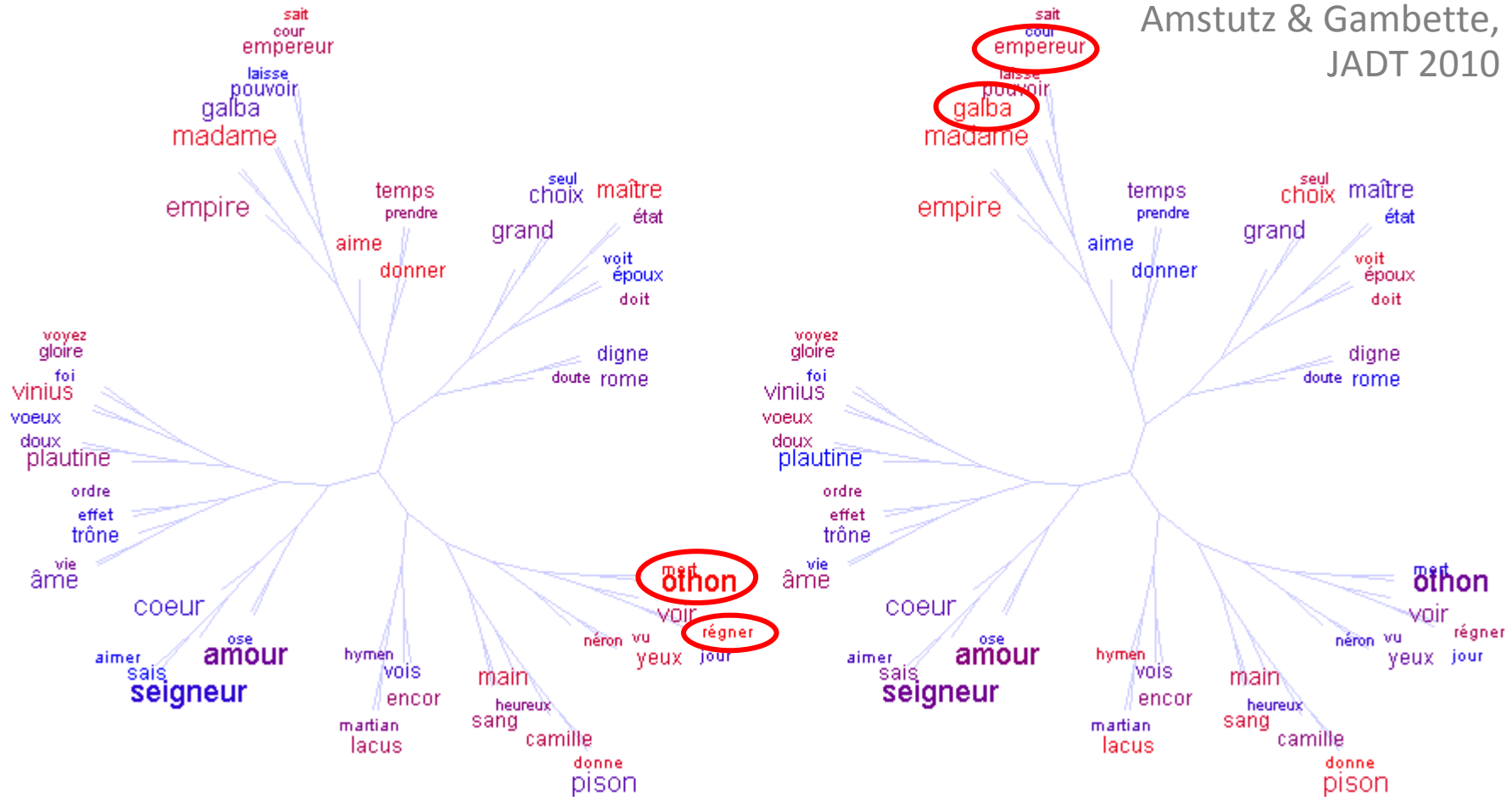
Amstutz & Gambette,
JADT 2010



*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 »*

Analyse littéraire : illustration sur *Othon*

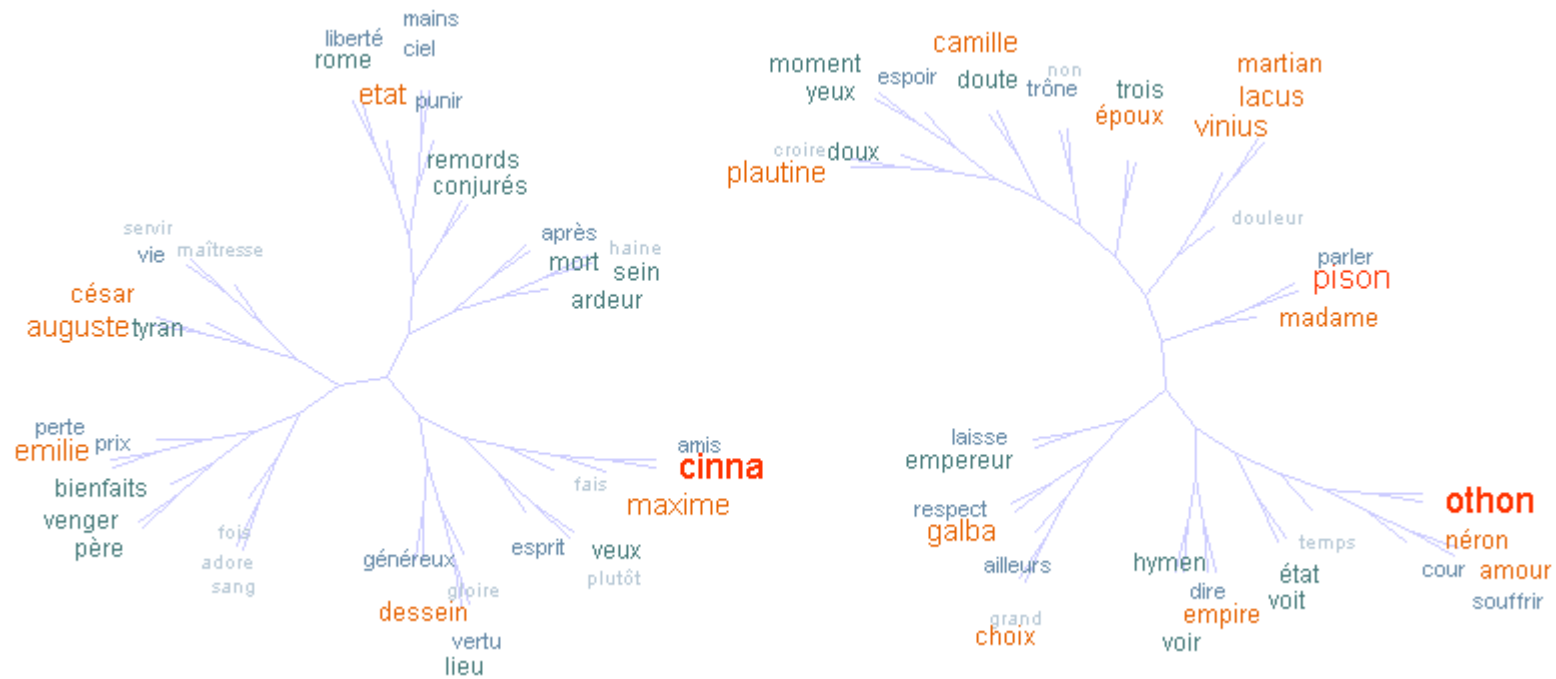
Amstutz & Gambette,
JADT 2010



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 »

Analyse littéraire : comparaison de *Cinna* et *Othon*

Amstutz & Gambette,
JADT 2010



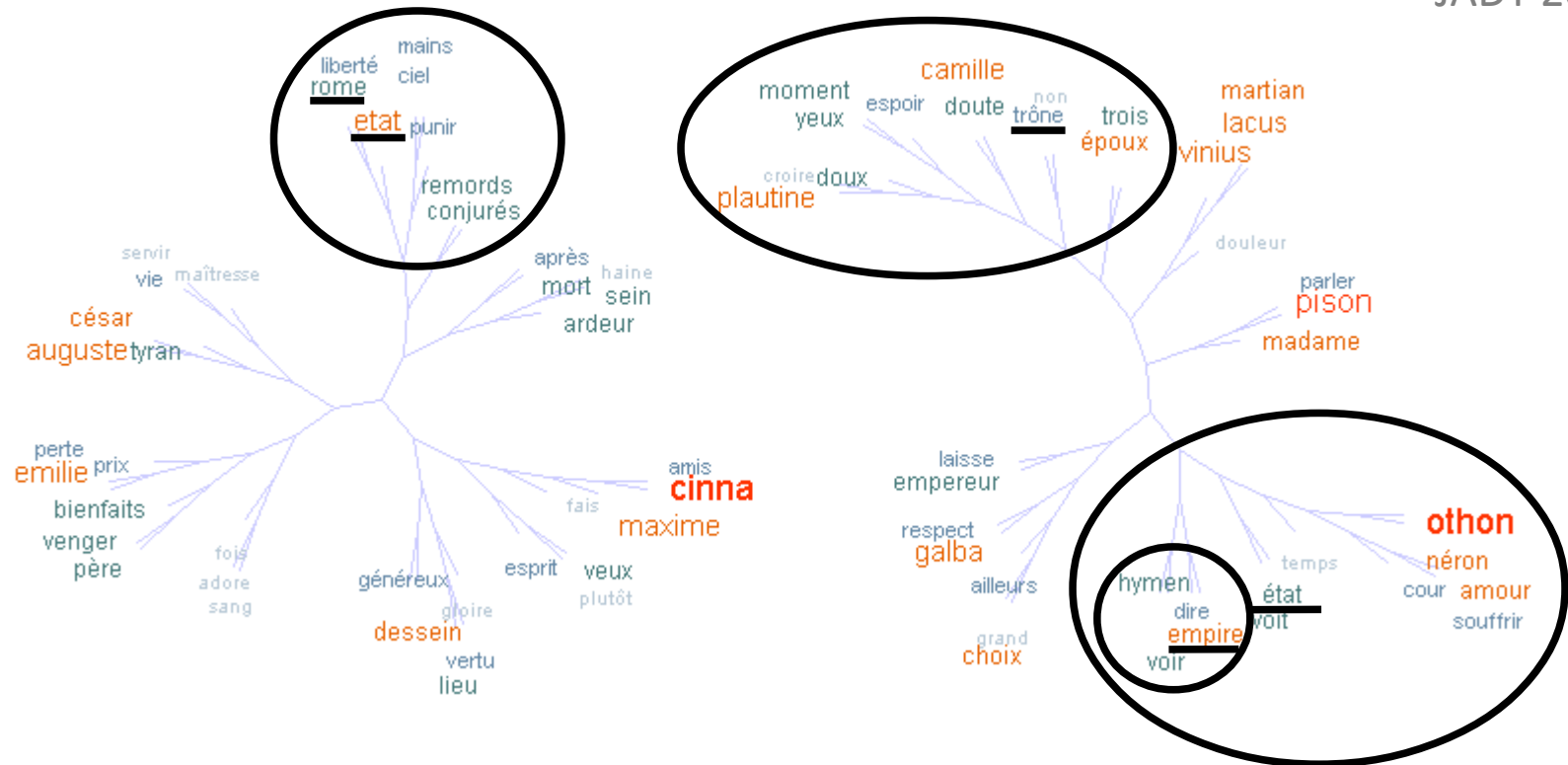
Nuages arborés des mots spécifiques de Cinna et Othon, dimensionnés et colorés d'après leur spécificité calculée dans Lexico3.

Quels moyens au service de la cause politique ?



Analyse littéraire : comparaison de *Cinna* et *Othon*

Amstutz & Gambette,
JADT 2010

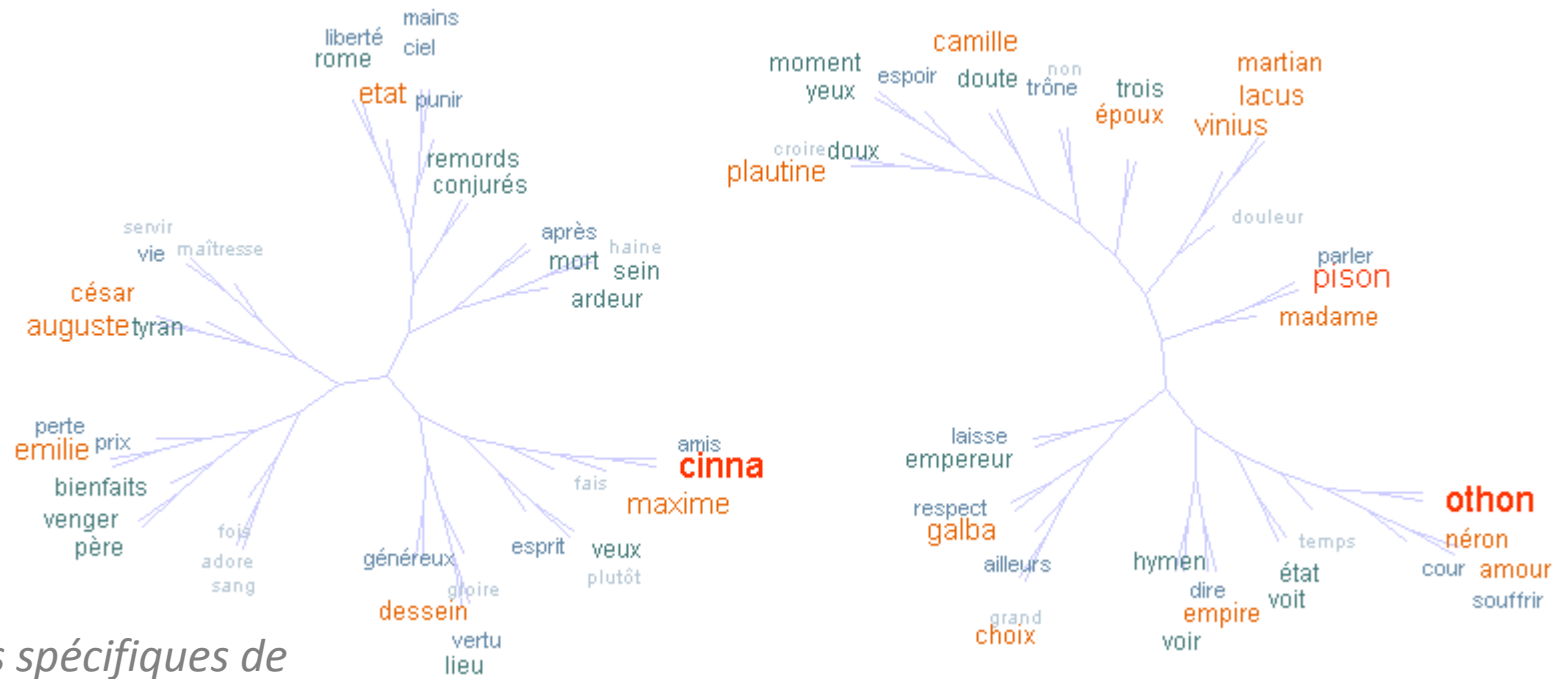


Nuages arborés des mots spécifiques de Cinna et Othon, dimensionnés et colorés d'après leur spécificité calculée dans Lexico3.



Quels moyens au service de la cause politique ?

Analyse littéraire : comparaison de *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 le corpus Mediator

Comparer les articles d'agences et articles de journalistes

Corpus : 595 articles d'agences contre 1496 articles de journalistes de 2011 évoquant l'affaire du Mediator dans la presse française.

Ensemble des articles

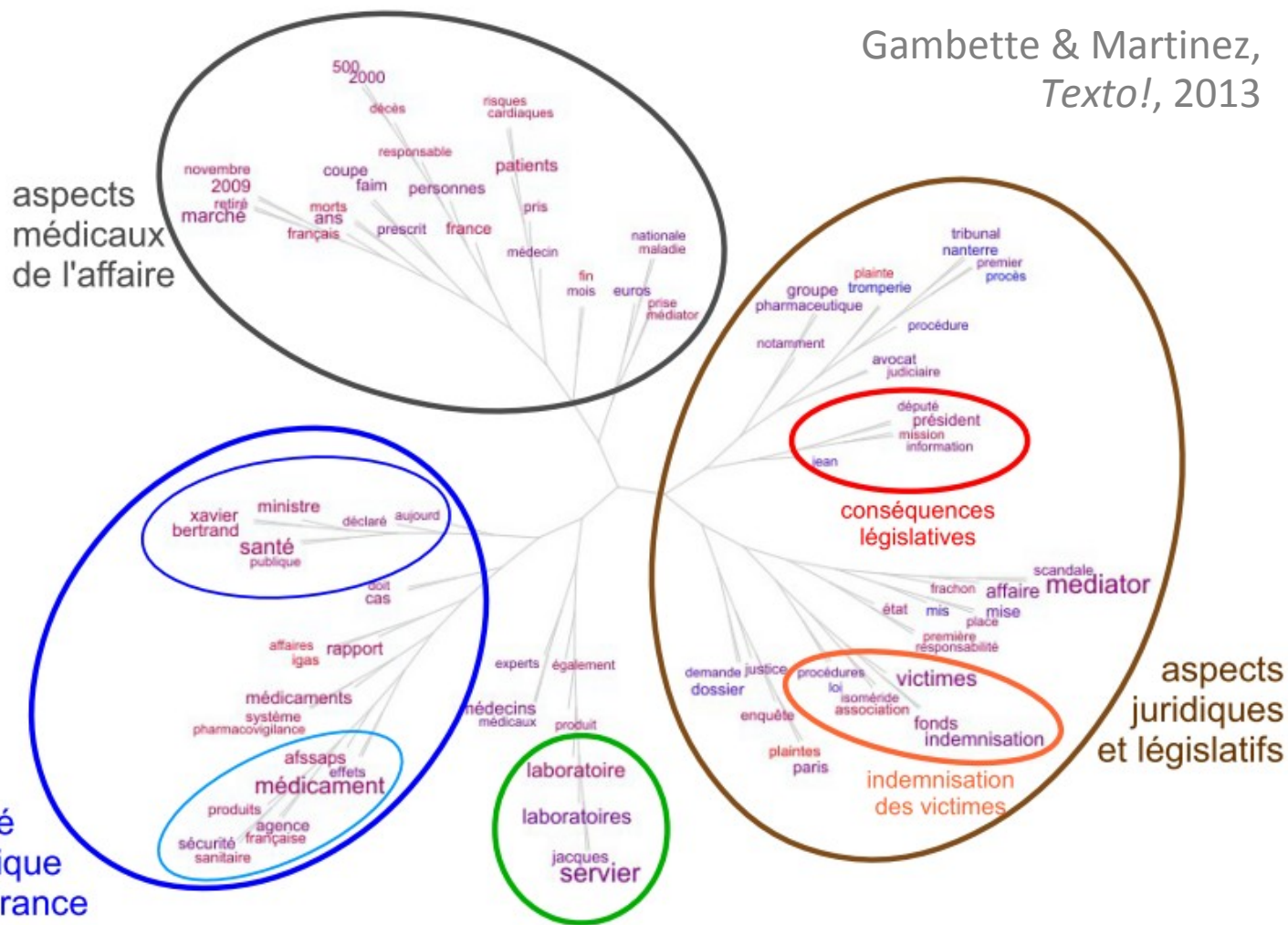
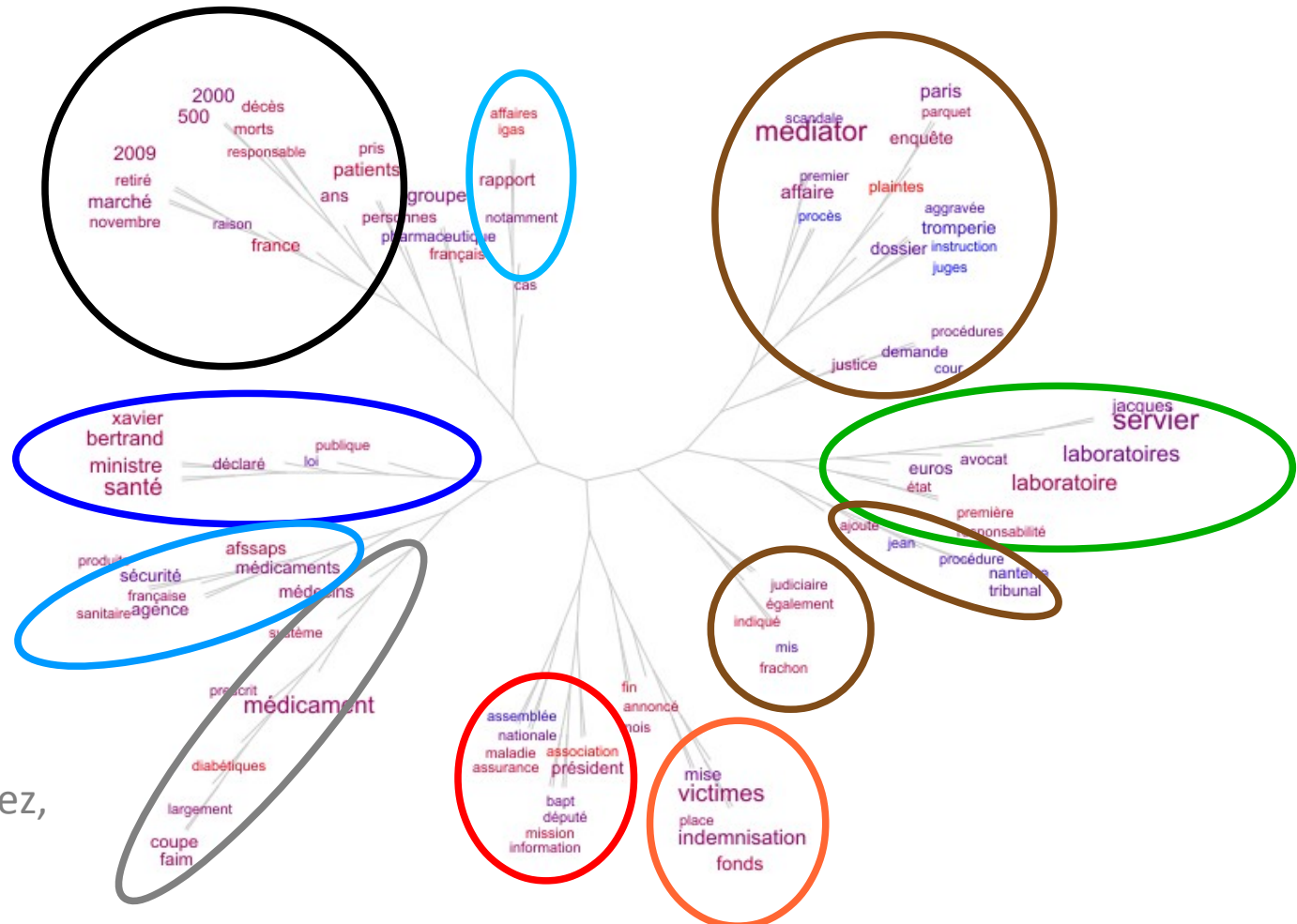


Illustration sur le corpus Mediator

Comparer les articles d'agences et articles de journalistes

Corpus : 595 articles d'agences contre 1496 articles de journalistes de 2011 évoquant l'affaire du Mediator dans la presse française.

Articles
d'agences



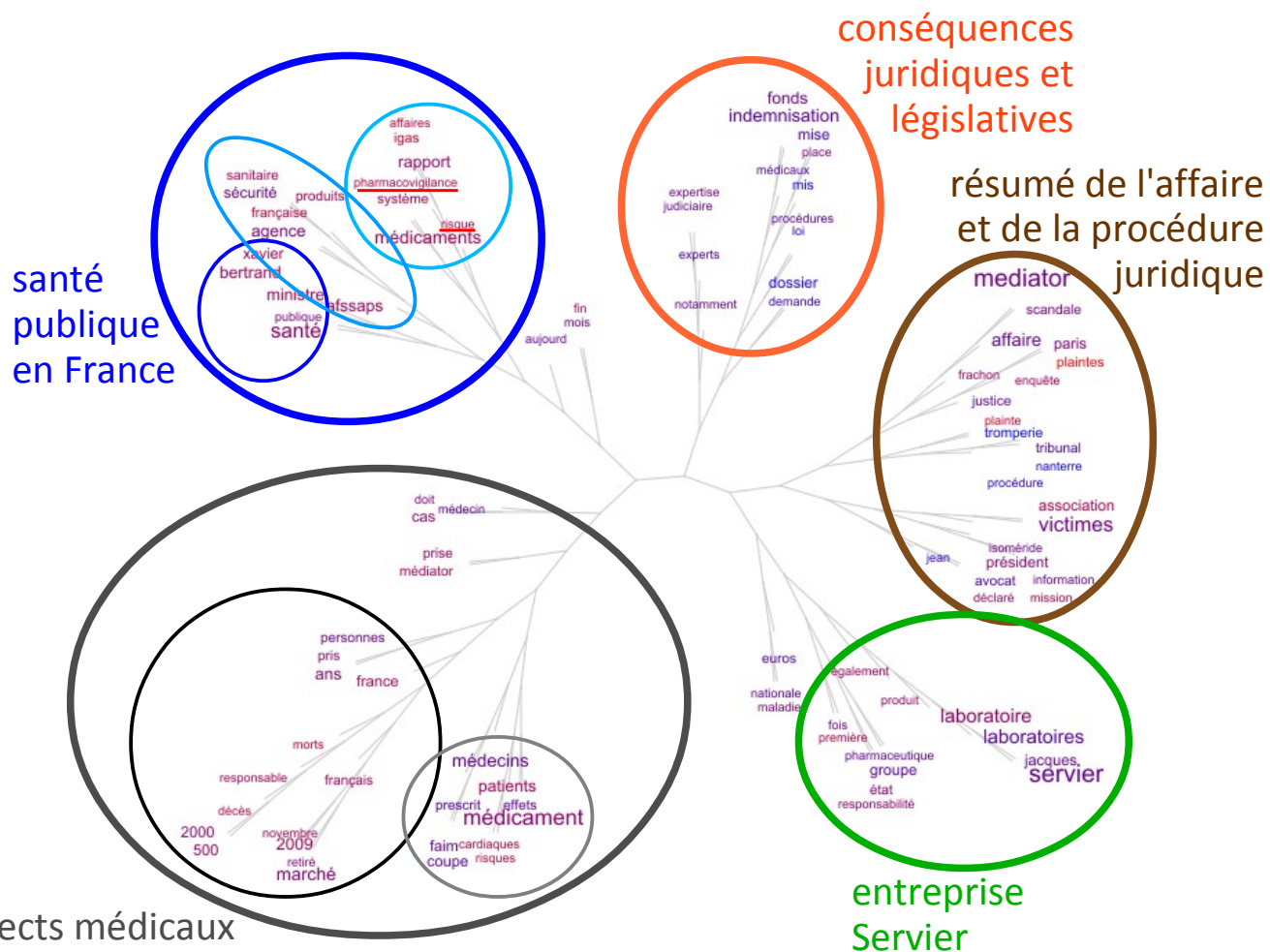
Gambette & Martinez,
Texto!, 2013

Illustration sur le corpus Mediator

Comparer les articles d'agences et articles de journalistes

Corpus : 595 articles d'agences contre 1496 articles de journalistes de 2011 évoquant l'affaire du Mediator dans la presse française.

Articles
de journalistes



Gambette & Martinez,
Texto!, 2013

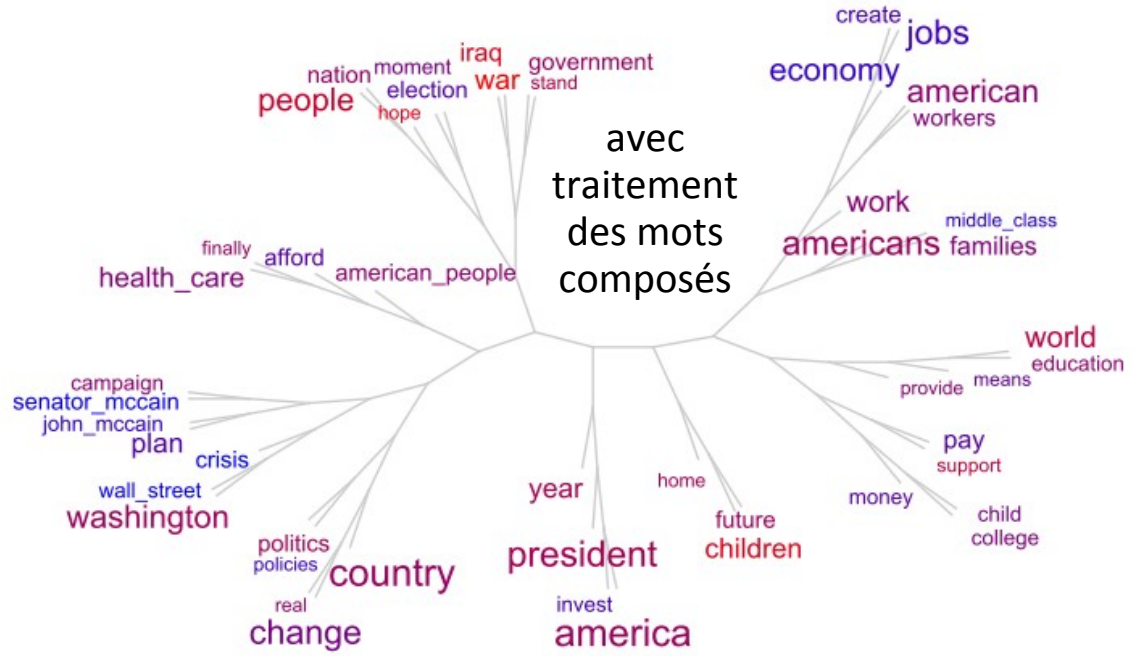
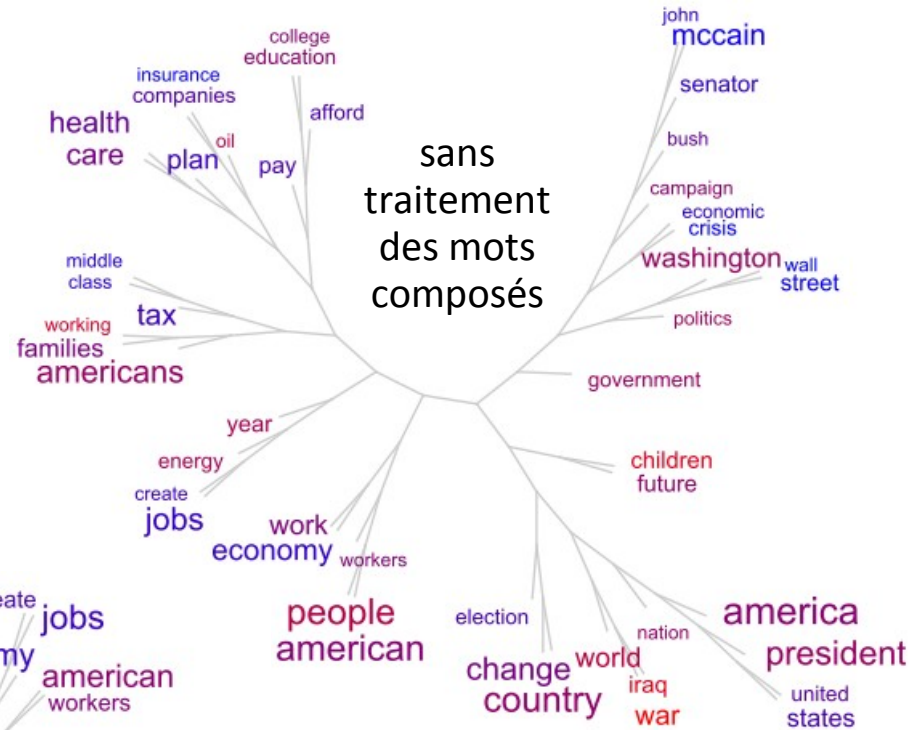
Plan

- Nuages arborés, intérêts et limites
- Construction des nuages arborés
- Options de coloration
- Utilisation des nuages arborés
- **Prétraitements du texte utilisé**
- Perspectives

Prétraitements du texte analysé

Détection des mots composés avec Unitex :

Nuages arborés des 50 mots les plus fréquents dans le discours d'Obama, fenêtres glissantes de 10 mots, distance Liddell, coloration chronologique



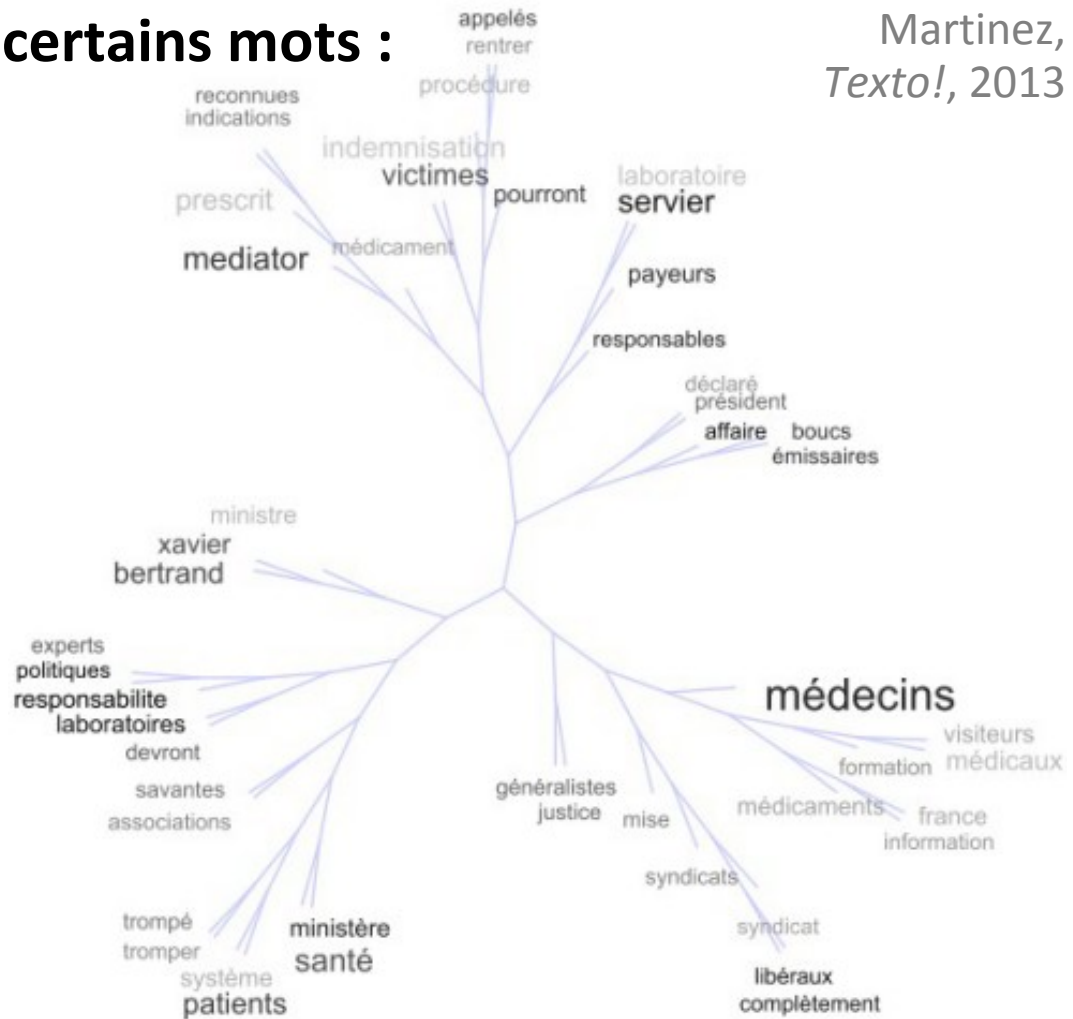
Prétraitements du texte analysé

Gambette &
Martinez,
Texto!, 2013

Focus sur le voisinage de certains mots :

- Contextes de x mots à gauche et y mots à droite du mot cible

Nuage arboré des 50 mots les plus fréquents des contextes (10 mots avant et 10 mots après) du mot médecins dans le sous-corpus des articles avec agence, colorés par le degré de cooccurrence avec le mot responsabilité (en noir pour les mots les plus cooccurents), distance Liddell, fenêtres glissantes de 20 mots



- Motifs trouvés par l'application d'un automate dans Unitex

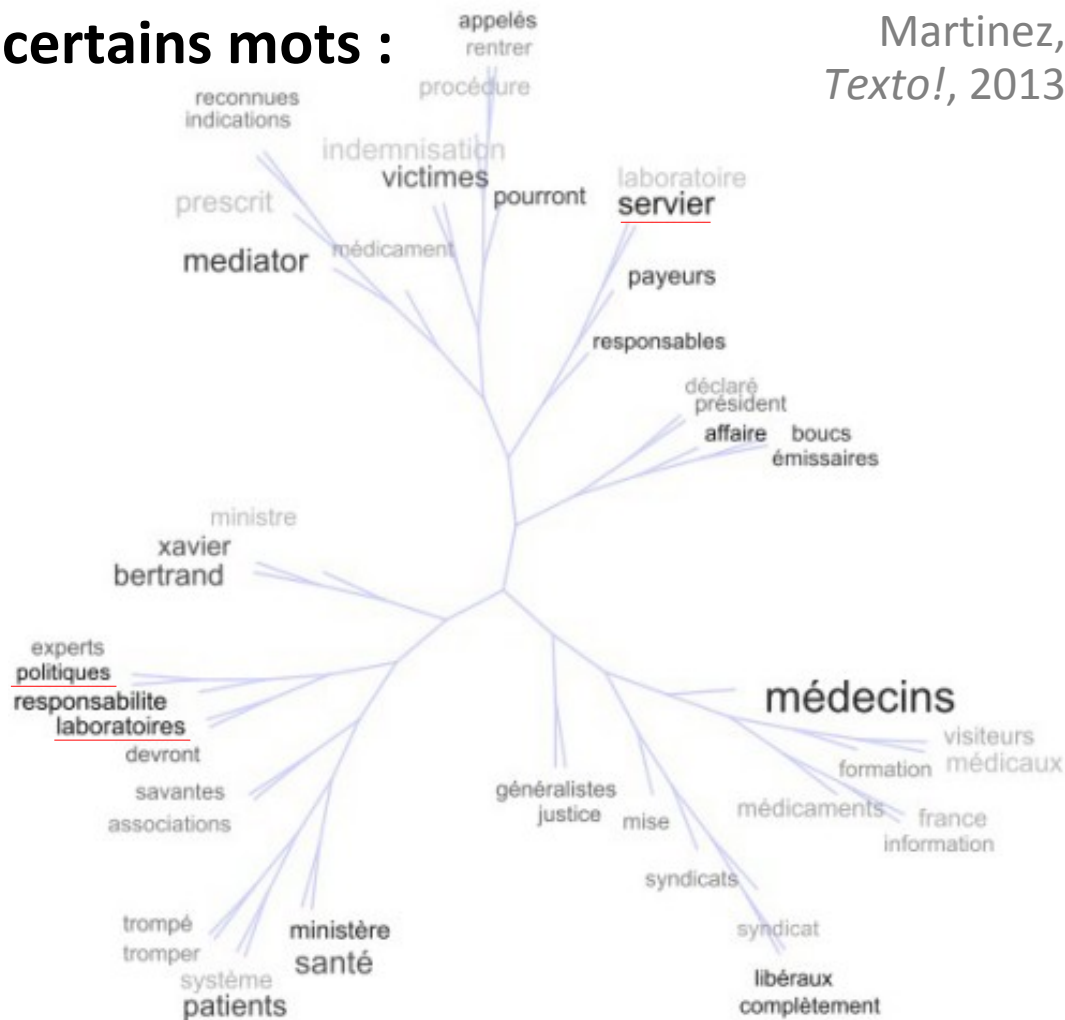
Prétraitements du texte analysé

Gambette &
Martinez,
Texto!, 2013

Focus sur le voisinage de certains mots :

- Contextes de x mots à gauche et y mots à droite du mot cible

*Nuage arboré des 50 mots les plus fréquents des contextes (10 mots avant et 10 mots après) du mot **médecins** dans le sous-corpus des **articles avec agence**, colorés par le degré de cooccurrence avec le mot **responsabilité** (en noir pour les mots les plus cooccurents), distance Liddell, fenêtres glissantes de 20 mots*



- Motifs trouvés par l'application d'un automate dans Unitex

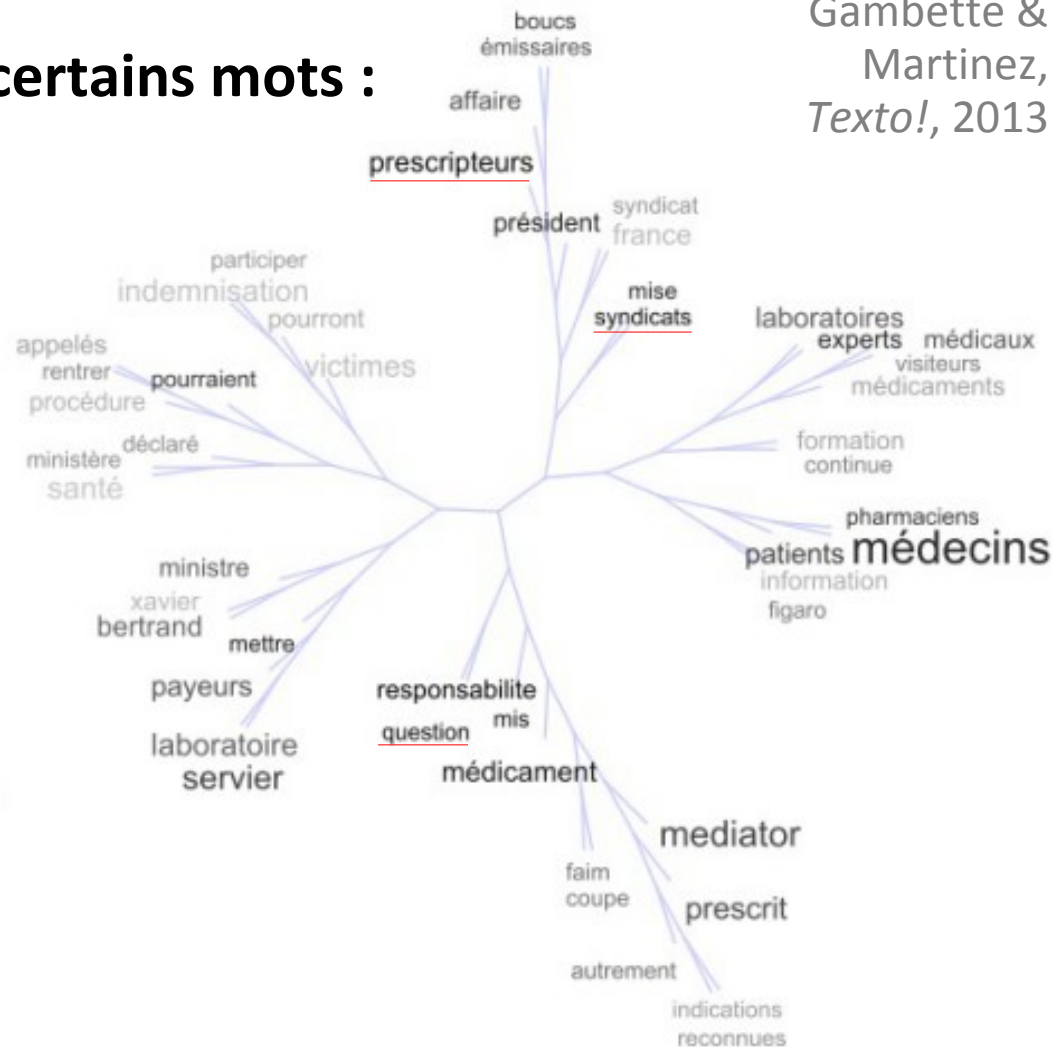
Prétraitements du texte analysé

Gambette & Martinez,
Texto!, 2013

Focus sur le voisinage de certains mots :

- Contextes de x mots à gauche et y mots à droite du mot cible

*Nuage arboré des 50 mots les plus fréquents des contextes (10 mots avant et 10 mots après) du mot **médecins** dans le sous-corpus des **articles de journalistes**, colorés par le degré de cooccurrence avec le mot **responsabilité** (en noir pour les mots les plus cooccurents), distance Liddell, fenêtres glissantes de 20 mots*



- Motifs trouvés par l'application d'un automate dans Unitex

Plan

- Nuages arborés, intérêts et limites
- Construction des nuages arborés
- Options de coloration
- Utilisation des nuages arborés
- Prétraitements du texte utilisé
- **Perspectives**

Perspectives

- intégration de la **visualisation en nuages arborés** avec longueurs de branches post-calculées :
 - avec des outils de prétraitement linguistique :
 - traitement des expressions composées (Unitex)
 - détection des entités nommées
 - étiquetage morphosyntaxique pour coloration (TreeTagger)
 - sélection de groupes syntaxiques dans les concordances (Unitex)
 - dans les outils de textométrie existants
 - par des interfaces d'import/export adaptées
 - pour faciliter le retour au texte
- **comparaison** d'arbres de mots
 - “morphing” d'arbres
 - calculs de dispersion dans les arbres
- développement des **interactions avec Unitex**
 - visualisation des résultats de “Locate” (concordances ou motifs)
 - enrichissement semi-automatique des grammaires

Références

Disponibles sur TreeCloud.org :

P. Gambette & J. Véronis (2009)

Visualising a Text with a Tree Cloud,

IFCS'09, Studies in Classification, Data Analysis, and Knowledge Organization 40,

p. 561-570

<http://www.slideshare.net/PhilippeGambette/visualising-a-text-with-a-tree-cloud>

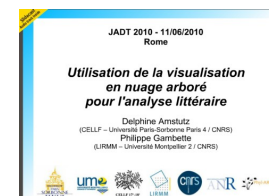


D. Amstutz & P. Gambette (2010)

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

JADT'10 (Proceedings of the 10th International Conference on statistical analysis of textual data), *Statistical Analysis of Textual Data*, p. 227-238

<http://www.slideshare.net/PhilippeGambette/utilisation-de-la-visualisation-en-nuage-arbor-pour-lanalyse-littraire>



P. Gambette, N. Gala & A. Nasr (2012)

Longueur de branches et arbres de mots,

Corpus 11:129-146

<http://www.slideshare.net/PhilippeGambette/longueur-de-branches-et-arbres-de-mots>



W. Martinez & P. Gambette (2013)

L'affaire du Médiateur au prisme de la textométrie,

Texto! XVIII(4)

<http://www.revue-texto.net/index.php?id=3318>