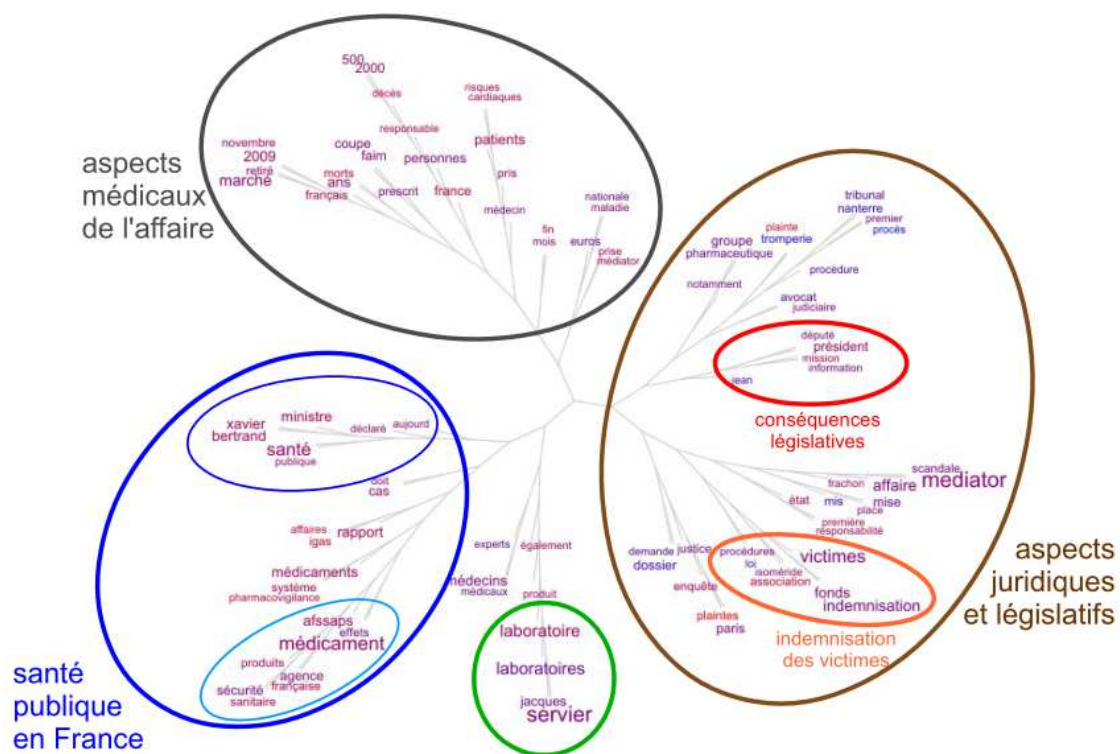


# Annexes de l'article "L'affaire du Mediator au prisme de la textométrie"

Gambette Philippe - Univ. Paris-Est Marne la Vallée.  
Martinez William - ILTEC, Lisbonne.

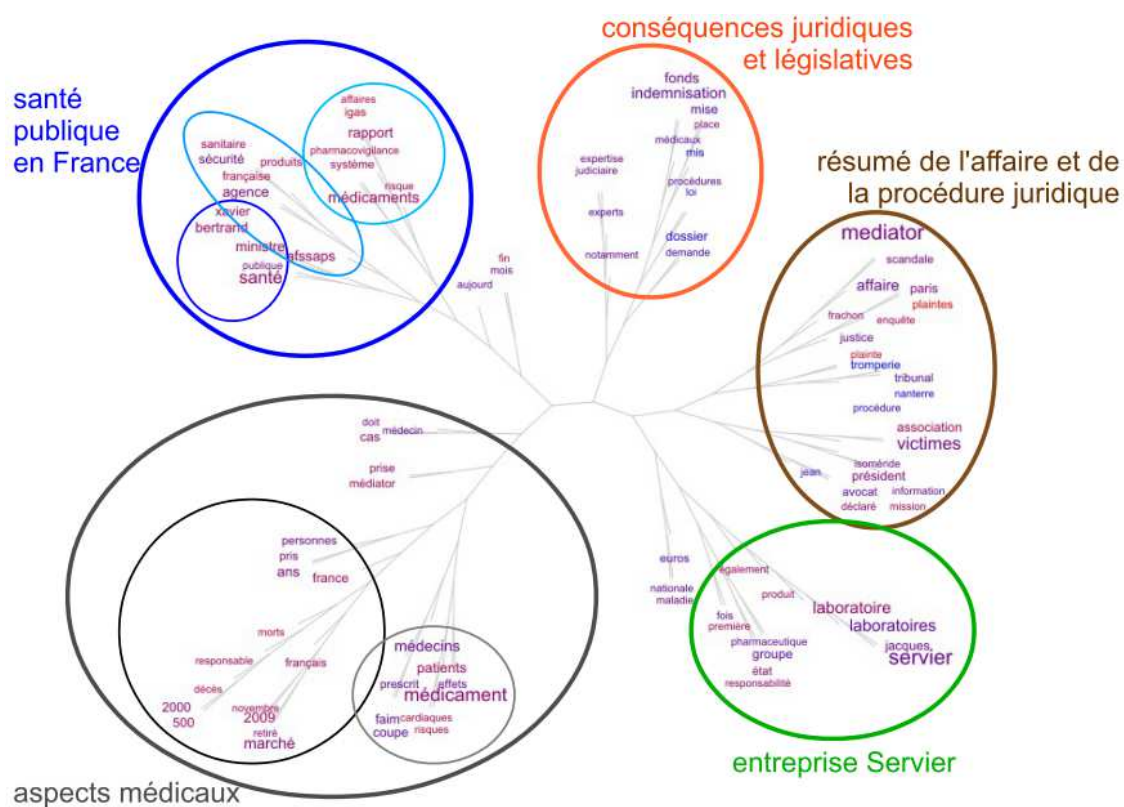
*En raison de contraintes de format, nous nous sommes limités dans notre article à une sélection de figures en niveaux de gris et de taille réduite. Cette annexe fournit l'ensemble de ces figures en assurant une meilleure lisibilité.*

## Corpus total

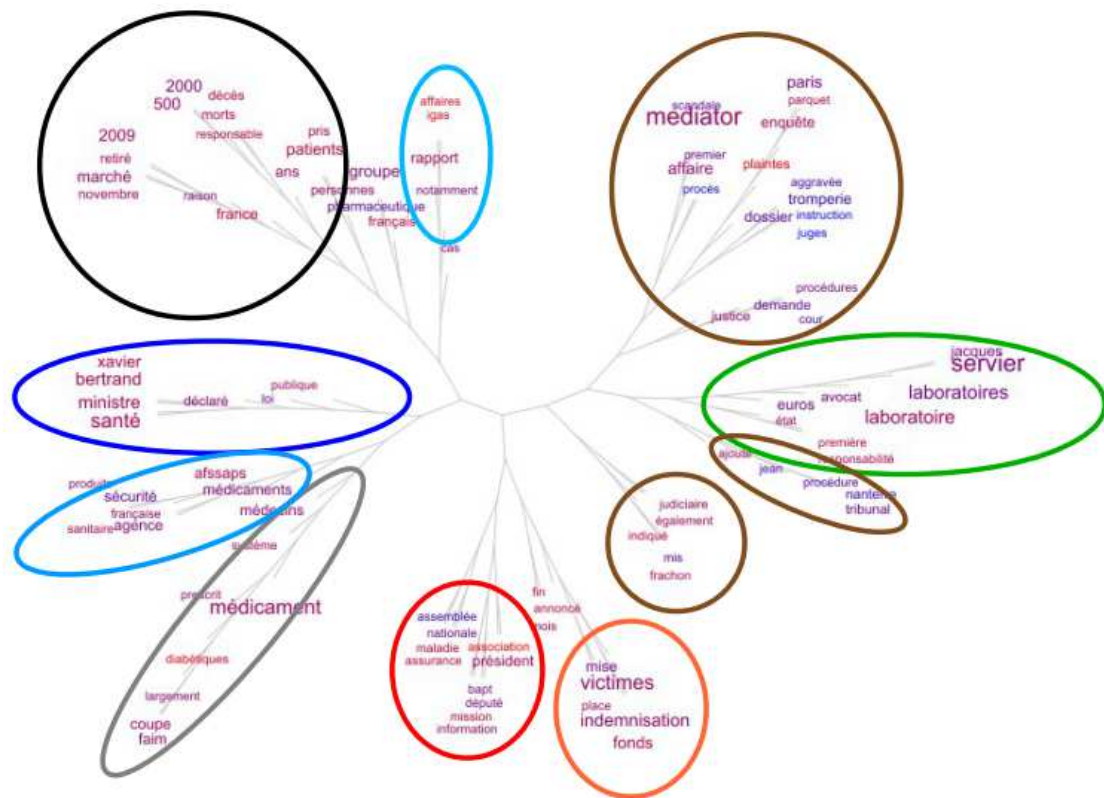


**Figure 1.** Nuage arboré des 100 mots les plus fréquents du corpus total, colorés chronologiquement (en bleu pour les mots de la fin du corpus et en rouge pour les mots du début du corpus), construit par TreeCloud avec la formule Liddell, et des fenêtres glissantes de 20 mots

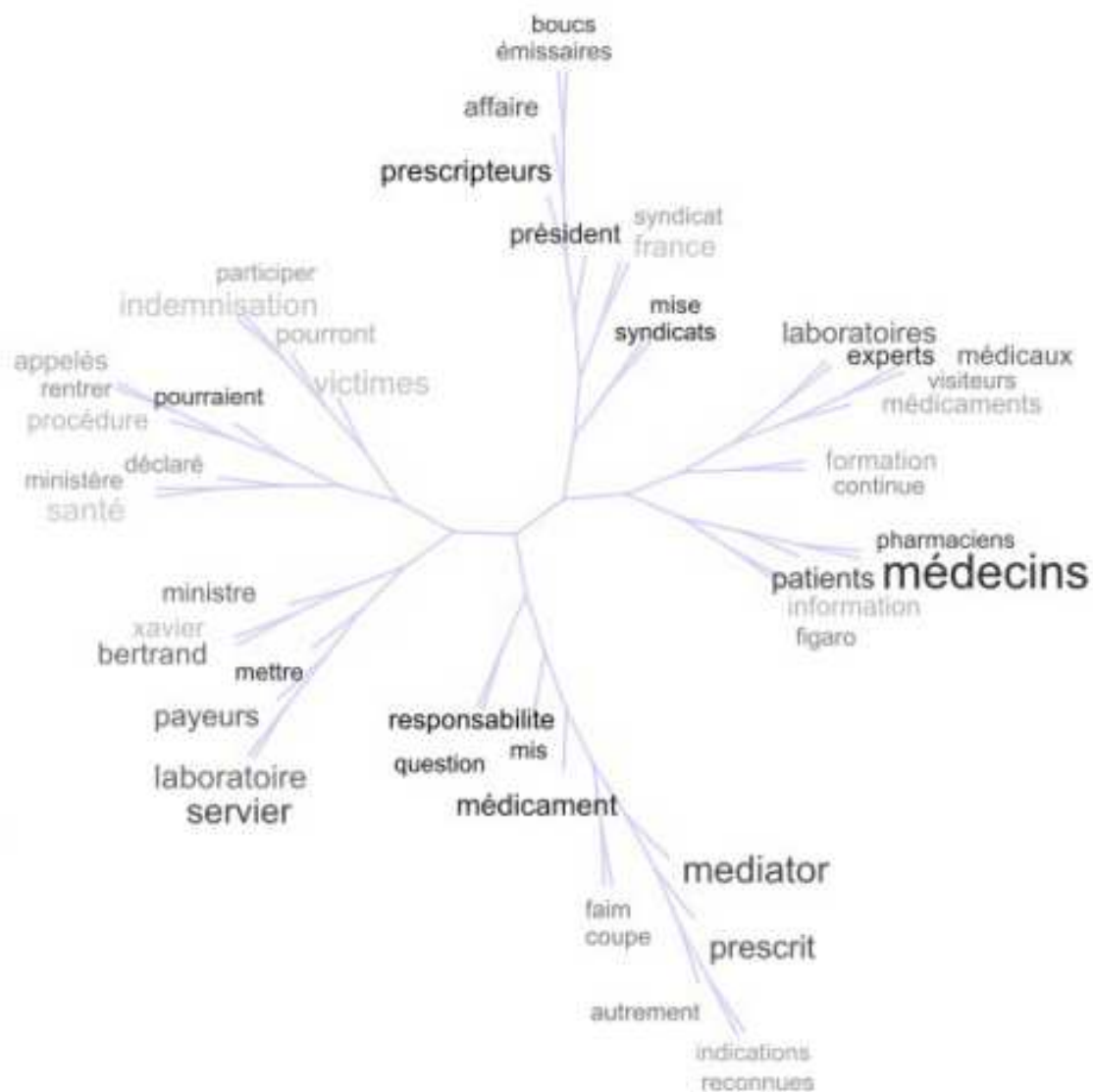
## Corpus avec agence / sans agence



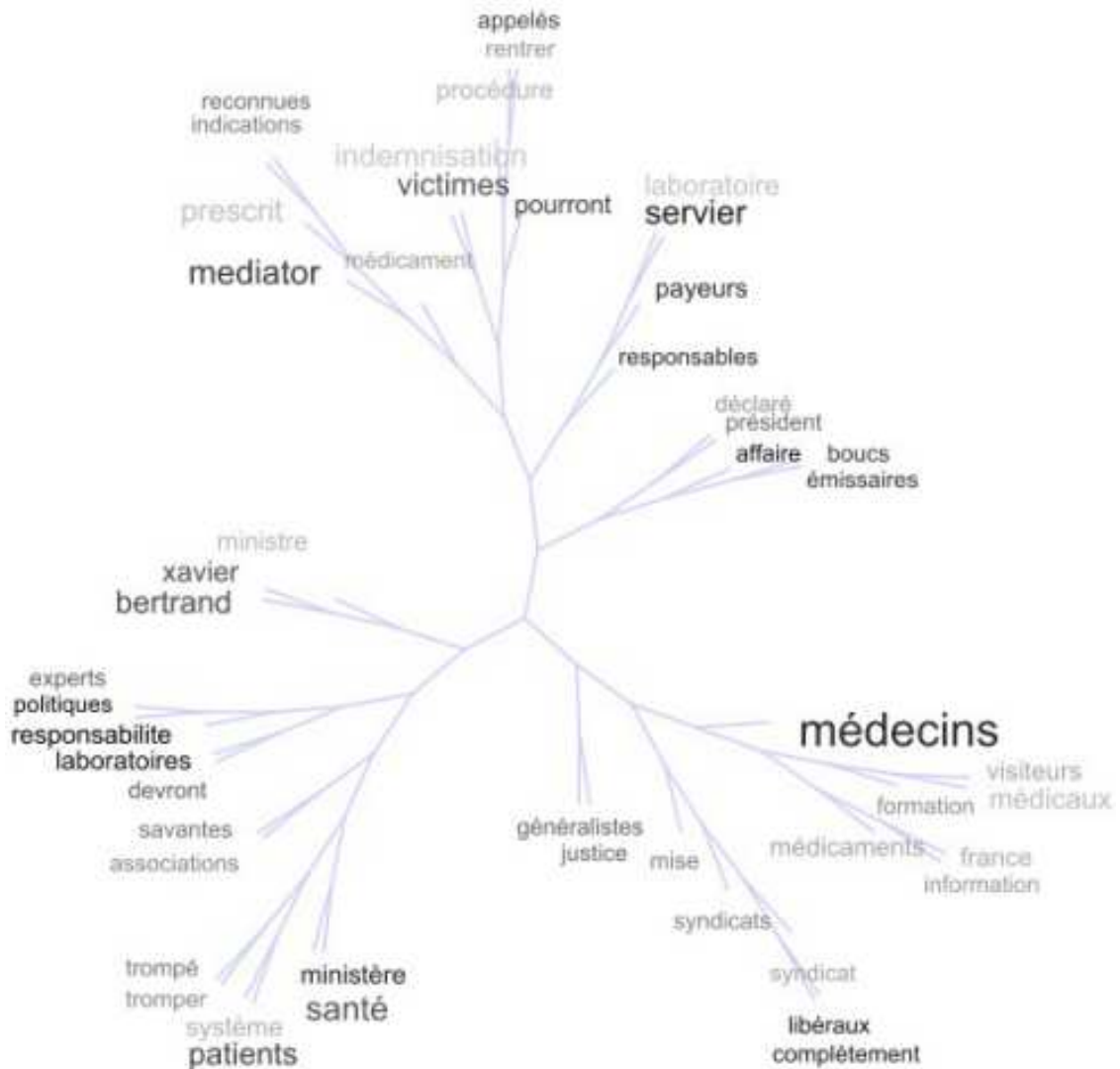
**Figure 2.** Nuage arboré des 100 mots les plus fréquents du sous-corpus des articles sans agence, colorés chronologiquement (en bleu pour les mots de la fin du corpus et en rouge pour les mots du début du corpus), construit par TreeCloud avec la formule Liddell, et des fenêtres glissantes de 20 mots



**Figure 3.** Nuage arboré des 100 mots les plus fréquents du sous-corpus des articles avec agence, colorés chronologiquement (en bleu pour les mots de la fin du corpus et en rouge pour les mots du début du corpus), construit par TreeCloud avec la formule Liddell, et des fenêtres glissantes de 20 mots



**Figure 4.** 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 *responsabilités* (en noir pour les mots les plus cooccurents), construit par TreeCloud avec la formule Liddell, et des fenêtres glissantes de 20 mots



**Figure 5.** 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 *responsabilités* (en noir pour les mots les plus cooccurents), construit par TreeCloud avec la formule Liddell, et des fenêtres glissantes de 20 mots

## Acteurs

Les nuages arborés des acteurs permettent de résumer leur rôle dans l'affaire du Médiateur, en montrant éventuellement les évolutions grâce à la coloration chronologique.

La liste des acteurs a été établie par l'analyse suivante, semi-automatique, en excluant le recours à des méthodes entièrement automatiques d'extraction d'entités nommées :

- identification visuelle de termes faisant référence à des acteurs dans le nuage arboré du corpus total : *afssaps, bertrand, frachon, igas, jacques, jean, servier, xavier* ;
- construction du nuage des contextes (3 mots à droite) de ces termes (voire figure 6) ;
- extraction des verbes conjugués de ce nuage : *accusé, affirmé, annoncé, assuré, auditionné, cherché, commercialisé, déclaré, demande, demandé, dénonce, dénoncé, devra, doit, entendu, estime, estimé, indiqué, mis, précisé, présenté, proposé, refuse, remis, rendu, souhaite, souligné* ;
- construction du nuage des contextes (3 mots à gauche) de ces verbes, en excluant ces verbes eux-mêmes (voir figure 7) ;
- identification d'acteurs supplémentaires dans ce nuage : *avim, bap, biopharma, claire, cnam, gérard, irène, michel, nicolas, oniam, sarkozy*.

Ceci conduit donc à la liste suivante d'acteurs identifiés sans faire référence à des prénoms et noms : *afssaps, avim, biopharma, cnam, igas, oniam*.

Nous avons complété cette liste par l'identification d'acteurs désignés par leur prénom suivi de leur nom, par la méthode suivante :

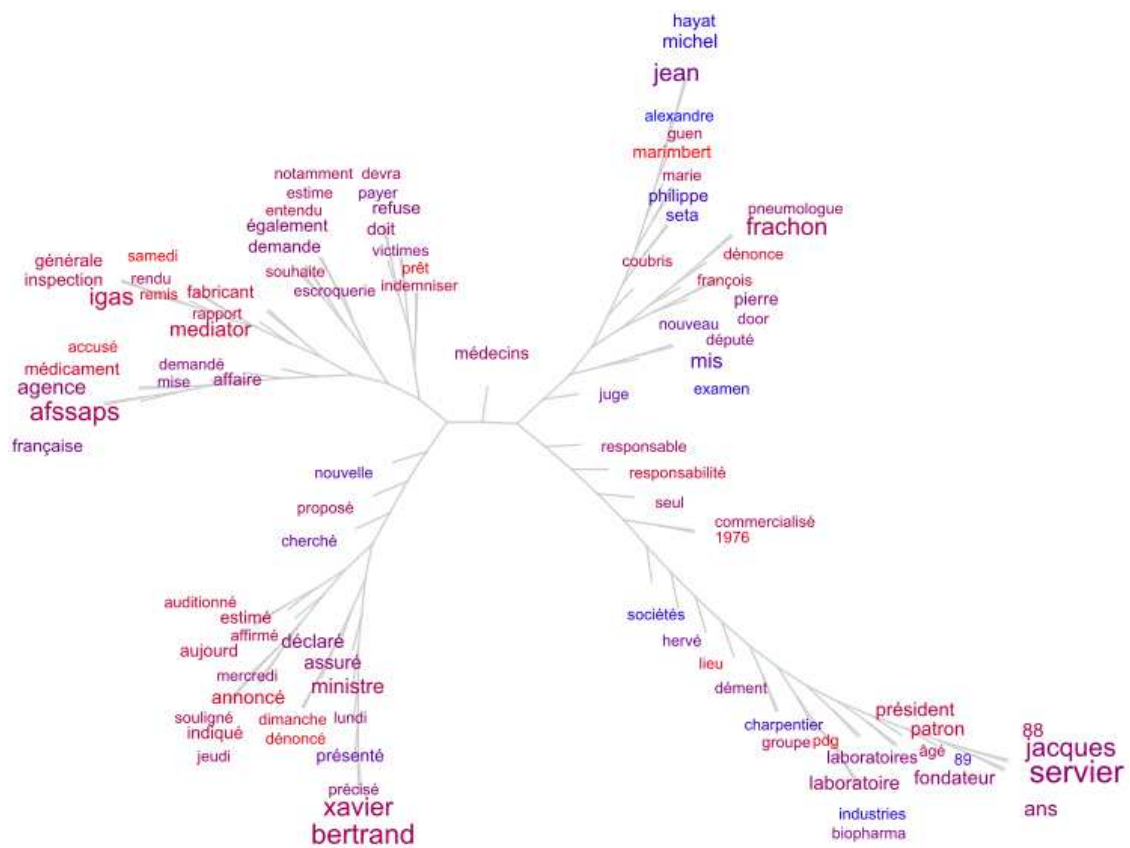
- sélection des prénoms les plus fréquents du corpus (dans la liste des termes classés par fréquence décroissante) :

*bertrand, xavier, jacques, jean, irène, gérard, michel, joseph, dominique, bernard, charles, nicolas, philippe, vincent, marie, françois, isabelle, claude, lucy, pierre, claire, alexandre ;*

- extraction des contextes (2 mots à droite) de ces prénoms ;
- sélection, à l'aide d'un nuage arboré (voir figure 8), des noms les plus fréquents dans les contextes et association avec les prénoms correspondant.

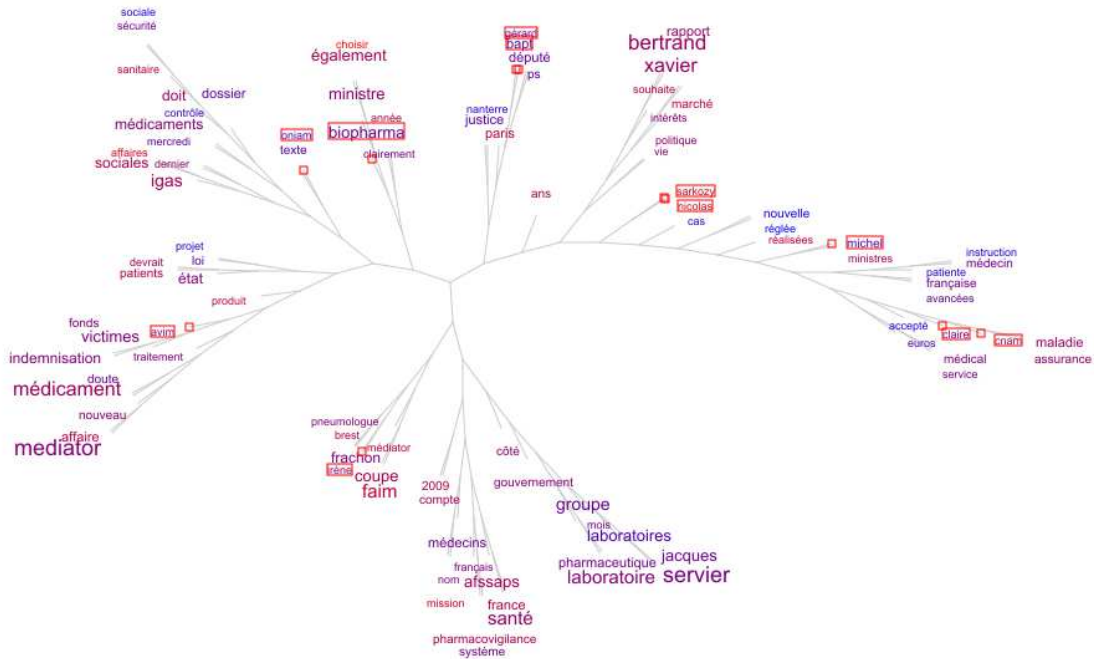
La liste obtenue est la suivante (prénom puis nom puis nombre d'occurrences identifiées) :

- Jacques Servier 1013
- Xavier Bertrand 1208
- Irène Frachon 454
- Gérard Bapt 335
- Charles-Joseph Oudin 202
- Nicolas Sarkozy 192
- Isabelle Prévost-Desprez 157
- Lucy Vincent 131
- Dominique-Michel Courtois 119
- Bernard Debré 117

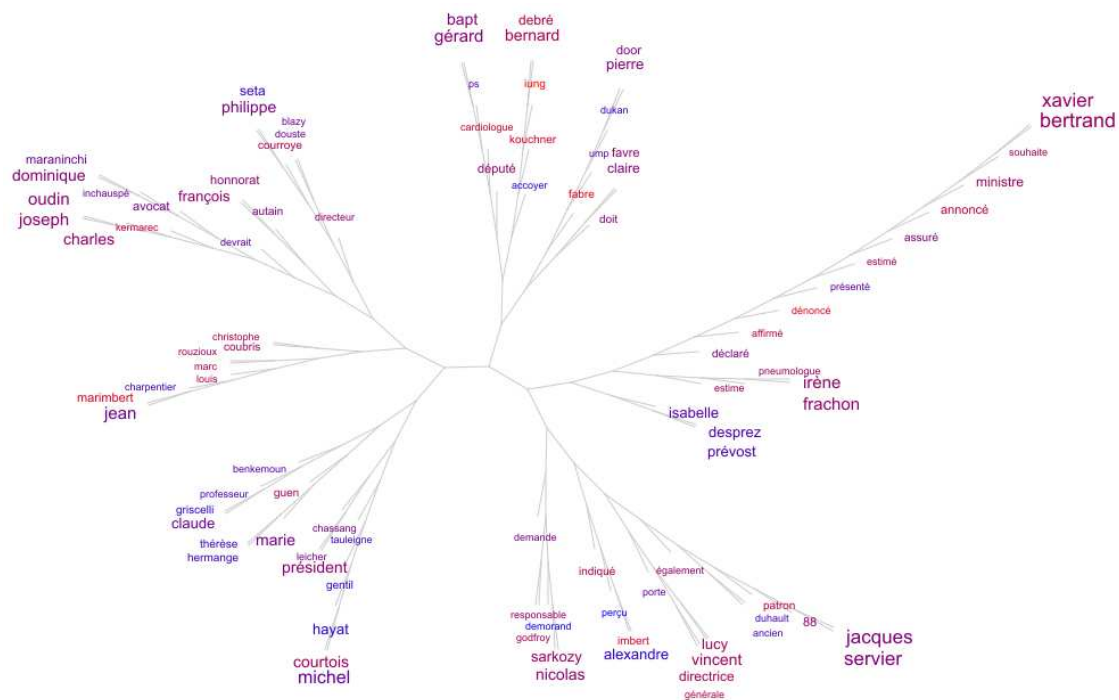


**Figure 6.** Nuage arboré des 100 mots les plus fréquents des contextes (3 mots à droite) des mots *afssaps*, *bertrand*, *frachon*, *igas*, *jacques*, *jean*, *servier* et *xavier* dans l'ensemble du corpus, colorés chronologiquement (en bleu pour les mots de la fin du corpus et en rouge pour les mots du début du corpus), construit par TreeCloud avec la formule Liddell, et des fenêtres glissantes de 20 mots

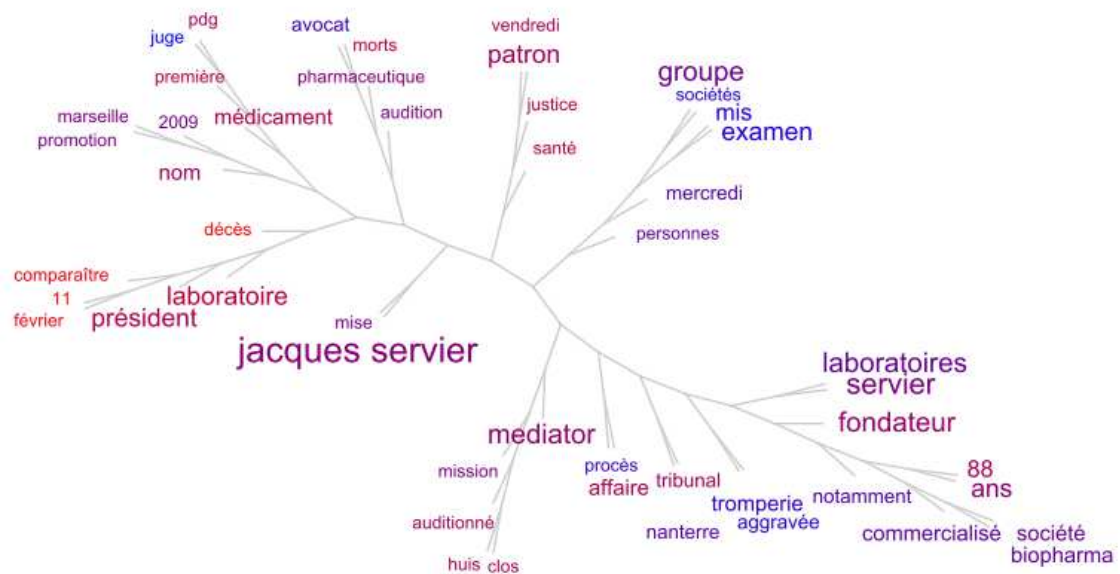




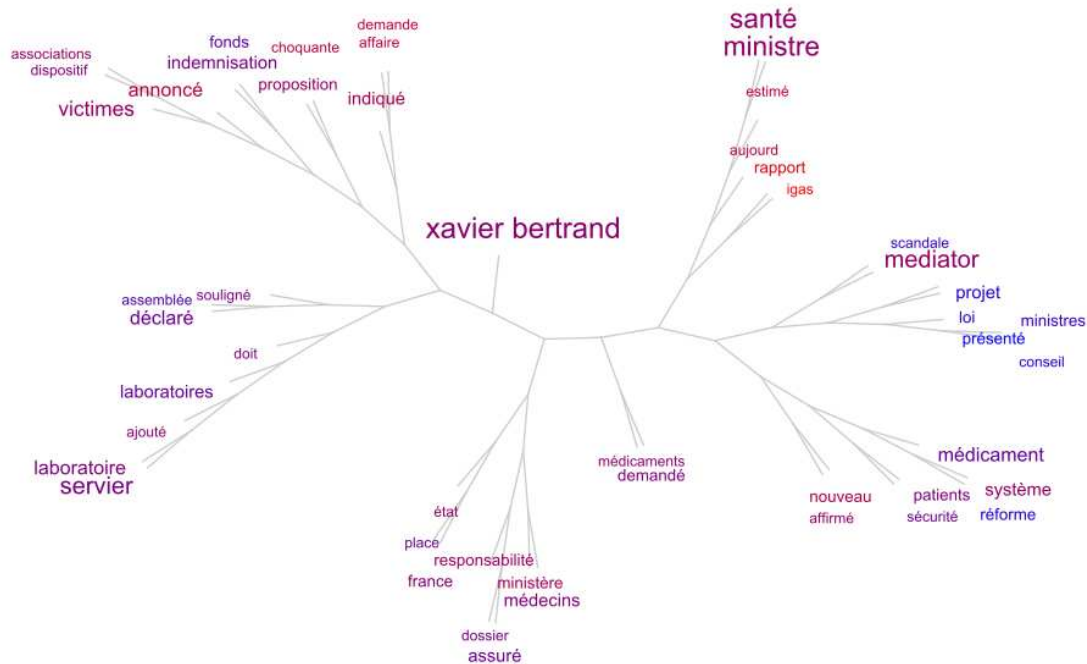
**Figure 7.** Nuage arboré des 100 mots les plus fréquents des contextes (3 mots à gauche) des mots *accusé, affirmé, annoncé, assuré, auditionné, cherché, commercialisé, déclaré, demande, demandé, dénonce, dénoncé, devra, doit, entendu, estime, estimé, indiqué, mis, précisé, présenté, proposé, refuse, remis, rendu, souhaite, souligné* dans l'ensemble du corpus, colorés chronologiquement (en bleu pour les mots de la fin du corpus et en rouge pour les mots du début du corpus), construit par TreeCloud avec la formule Liddell, et des fenêtres glissantes de 20 mots



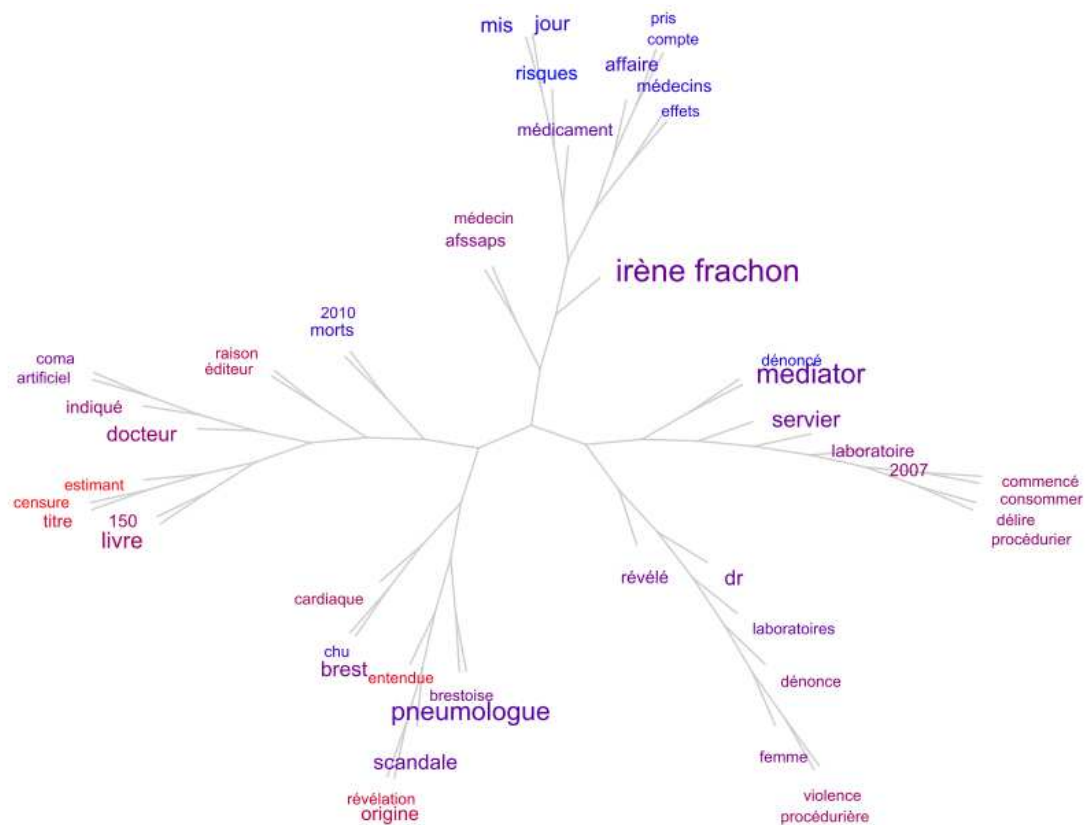
**Figure 8.** Nuage arboré des 100 mots les plus fréquents des contextes (2 mots à droite) des mots *bertrand*, *xavier*, *jacques*, *jean*, *irène*, *gérard*, *michel*, *joseph*, *dominique*, *bernard*, *charles*, *nicolas*, *philippe*, *vincent*, *marie*, *françois*, *isabelle*, *claire*, *lucy*, *pierre*, *claire*, *alexandre* dans l'ensemble du corpus, colorés chronologiquement (en bleu pour les mots de la fin du corpus et en rouge pour les mots du début du corpus), construit par TreeCloud avec la formule Liddell, et des fenêtres glissantes de 20 mots



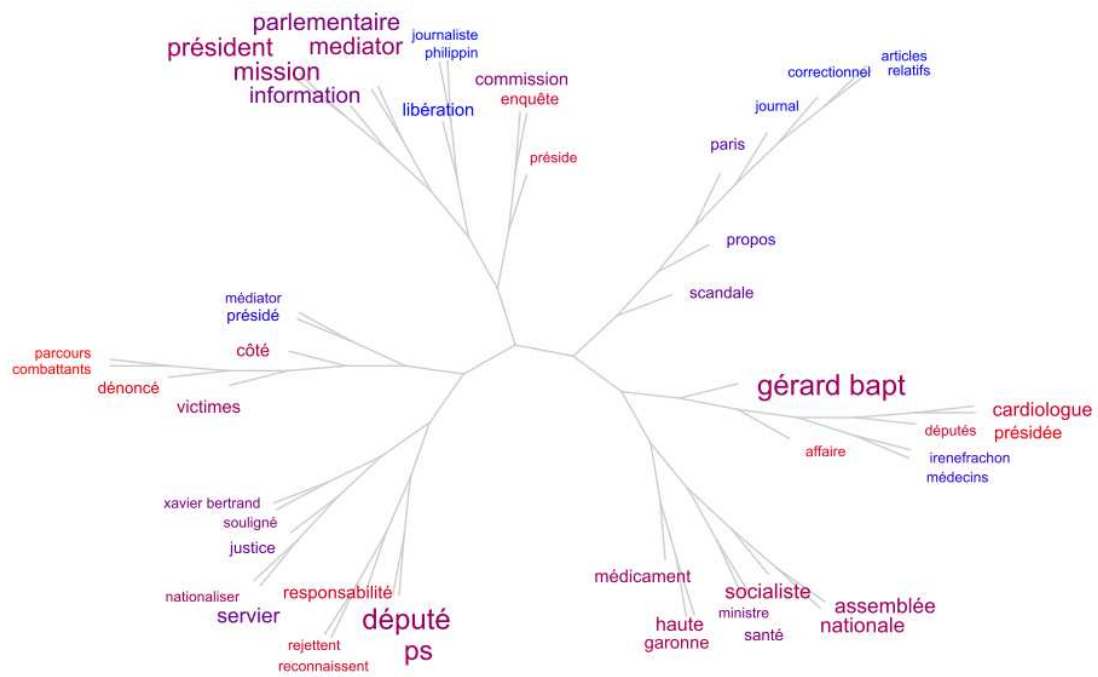
**Figure 9.** Nuage arboré des 50 mots les plus fréquents des contextes (10 mots à gauche et 10 mots à droite) de l'expression *jacques servier* dans l'ensemble du corpus, colorés chronologiquement (en bleu pour les mots de la fin du corpus et en rouge pour les mots du début du corpus), construit par TreeCloud avec la formule Liddell, et des fenêtres glissantes de 20 mots



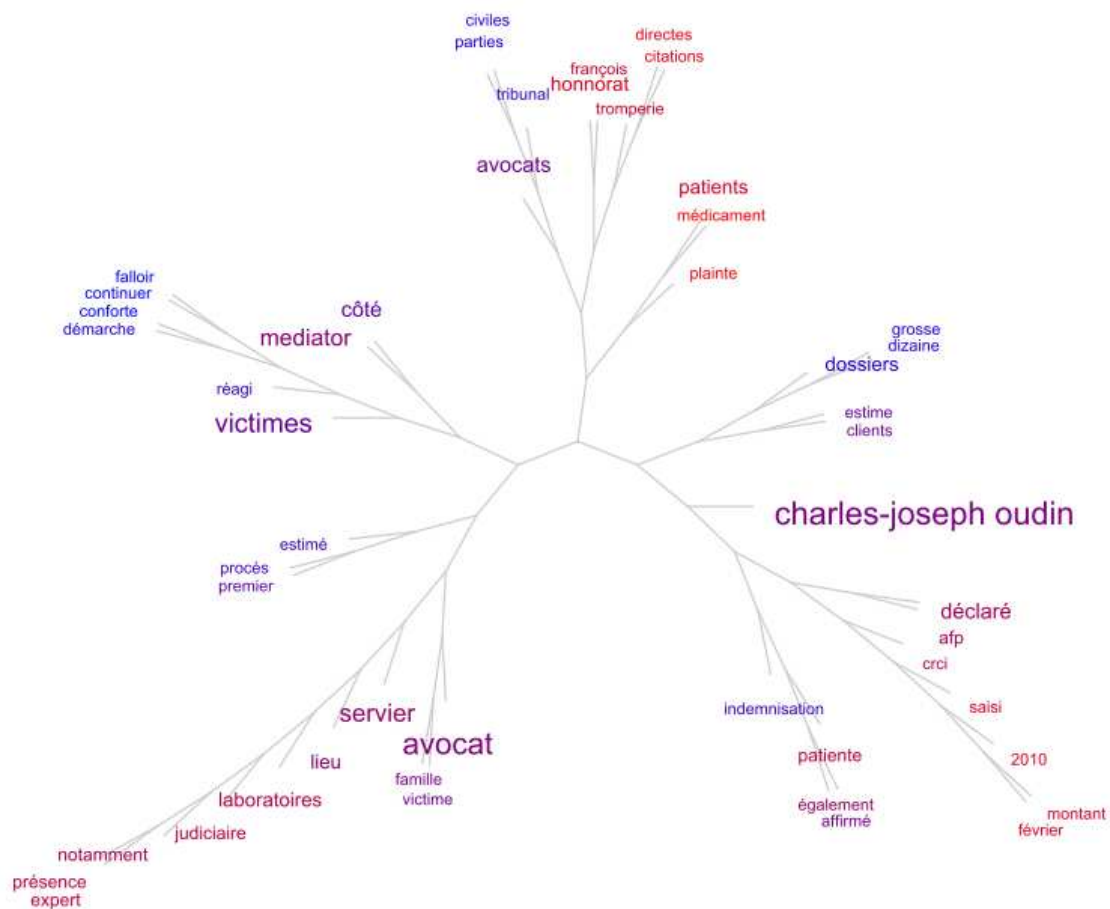
**Figure 10.** Nuage arboré des 50 mots les plus fréquents des contextes (10 mots à gauche et 10 mots à droite) de l'expression *xavier bertrand* dans l'ensemble du corpus, colorés chronologiquement (en bleu pour les mots de la fin du corpus et en rouge pour les mots du début du corpus), construit par TreeCloud avec la formule Liddell, et des fenêtres glissantes de 20 mots



**Figure 11.** Nuage arboré des 50 mots les plus fréquents des contextes (10 mots à gauche et 10 mots à droite) de l'expression *irène frachon* dans l'ensemble du corpus, colorés chronologiquement (en bleu pour les mots de la fin du corpus et en rouge pour les mots du début du corpus), construit par TreeCloud avec la formule Liddell, et des fenêtres glissantes de 20 mots



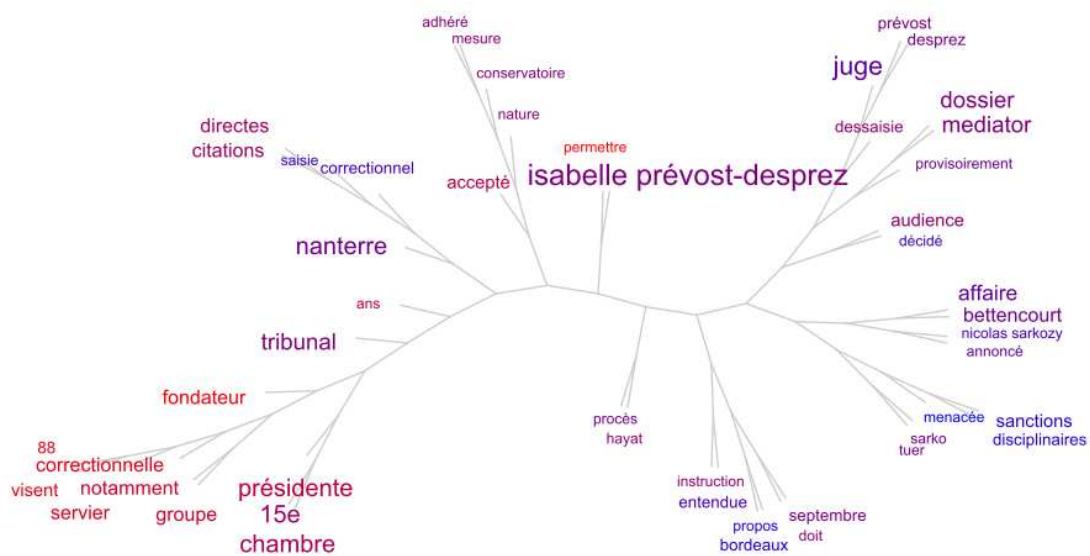
**Figure 12.** Nuage arboré des 50 mots les plus fréquents des contextes (10 mots à gauche et 10 mots à droite) de l'expression *gérard bapt* dans l'ensemble du corpus, colorés chronologiquement (en bleu pour les mots de la fin du corpus et en rouge pour les mots du début du corpus), construit par TreeCloud avec la formule Liddell, et des fenêtres glissantes de 20 mots



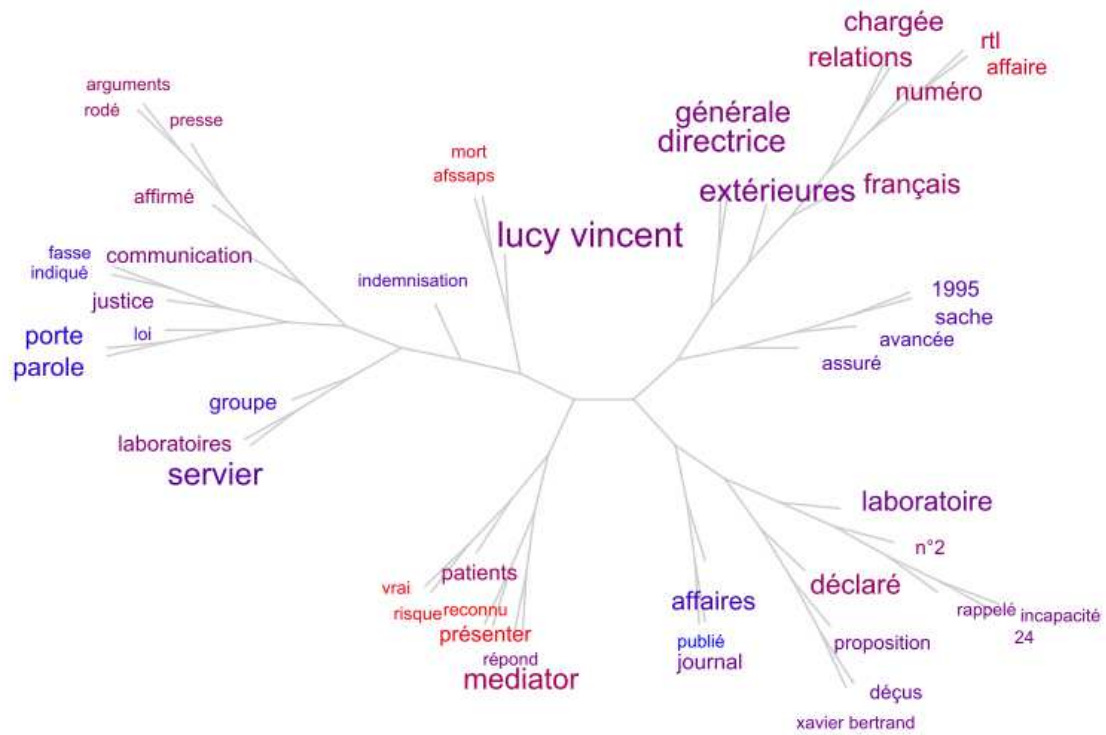
**Figure 13.** Nuage arboré des 50 mots les plus fréquents des contextes (10 mots à gauche et 10 mots à droite) de l'expression *charles joseph oudin* dans l'ensemble du corpus, colorés chronologiquement (en bleu pour les mots de la fin du corpus et en rouge pour les mots du début du corpus), construit par TreeCloud avec la formule Liddell, et des fenêtres glissantes de 20 mots



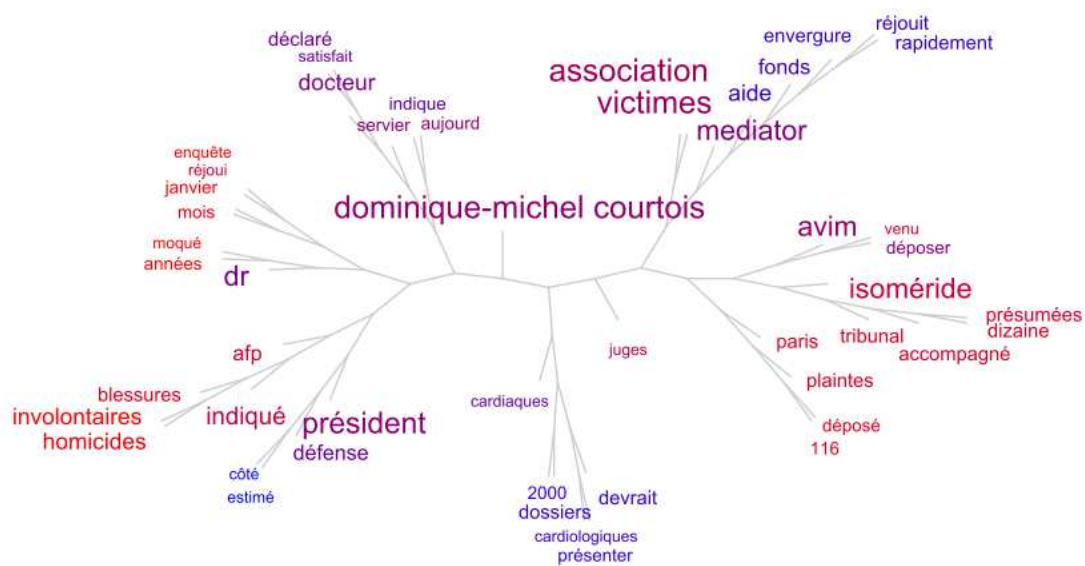




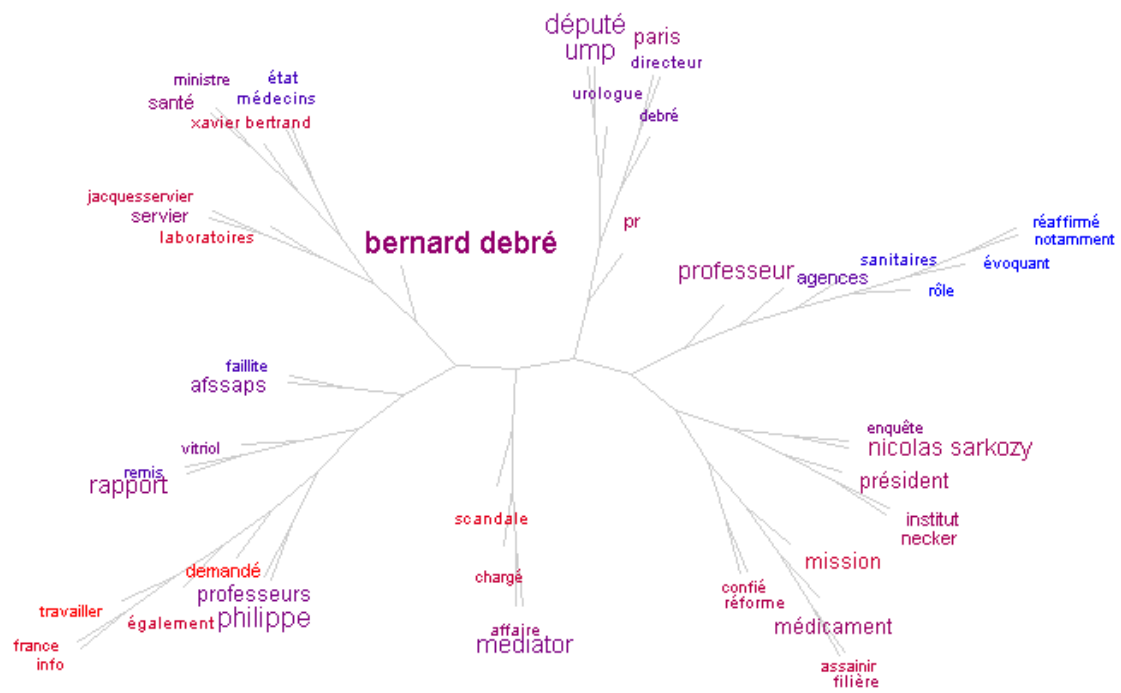
**Figure 15.** Nuage arboré des 50 mots les plus fréquents des contextes (10 mots à gauche et 10 mots à droite) de l'expression *isabelle prévost desprez* dans l'ensemble du corpus, colorés chronologiquement (en bleu pour les mots de la fin du corpus et en rouge pour les mots du début du corpus), construit par TreeCloud avec la formule Liddell, et des fenêtres glissantes de 20 mots



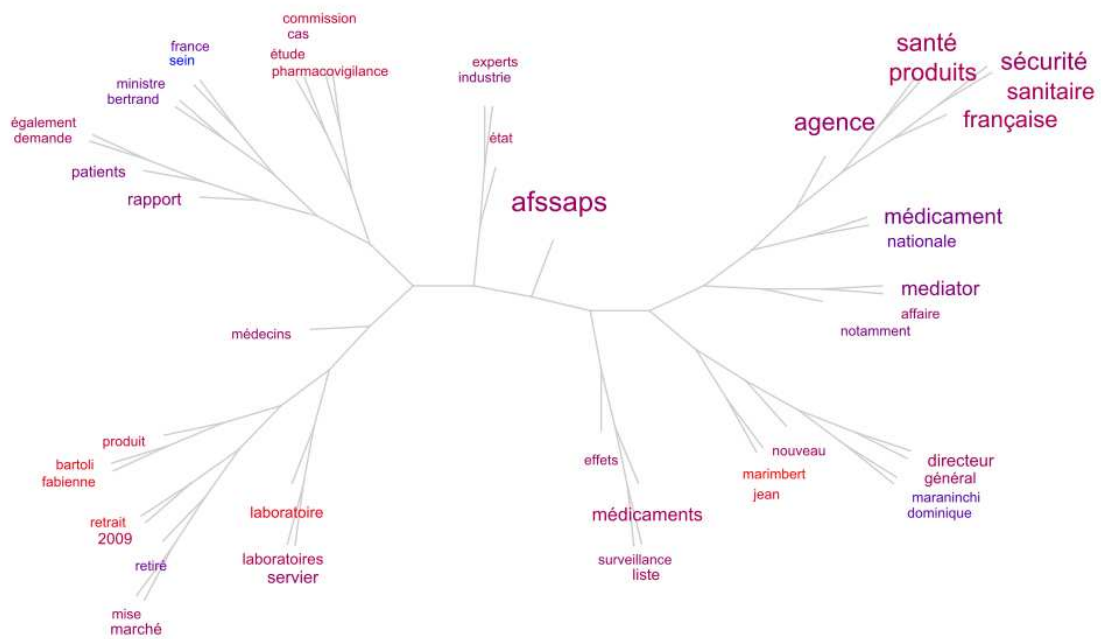
**Figure 16.** Nuage arboré des 50 mots les plus fréquents des contextes (10 mots à gauche et 10 mots à droite) de l'expression *lucy vincent* dans l'ensemble du corpus, colorés chronologiquement (en bleu pour les mots de la fin du corpus et en rouge pour les mots du début du corpus), construit par TreeCloud avec la formule Liddell, et des fenêtres glissantes de 20 mots



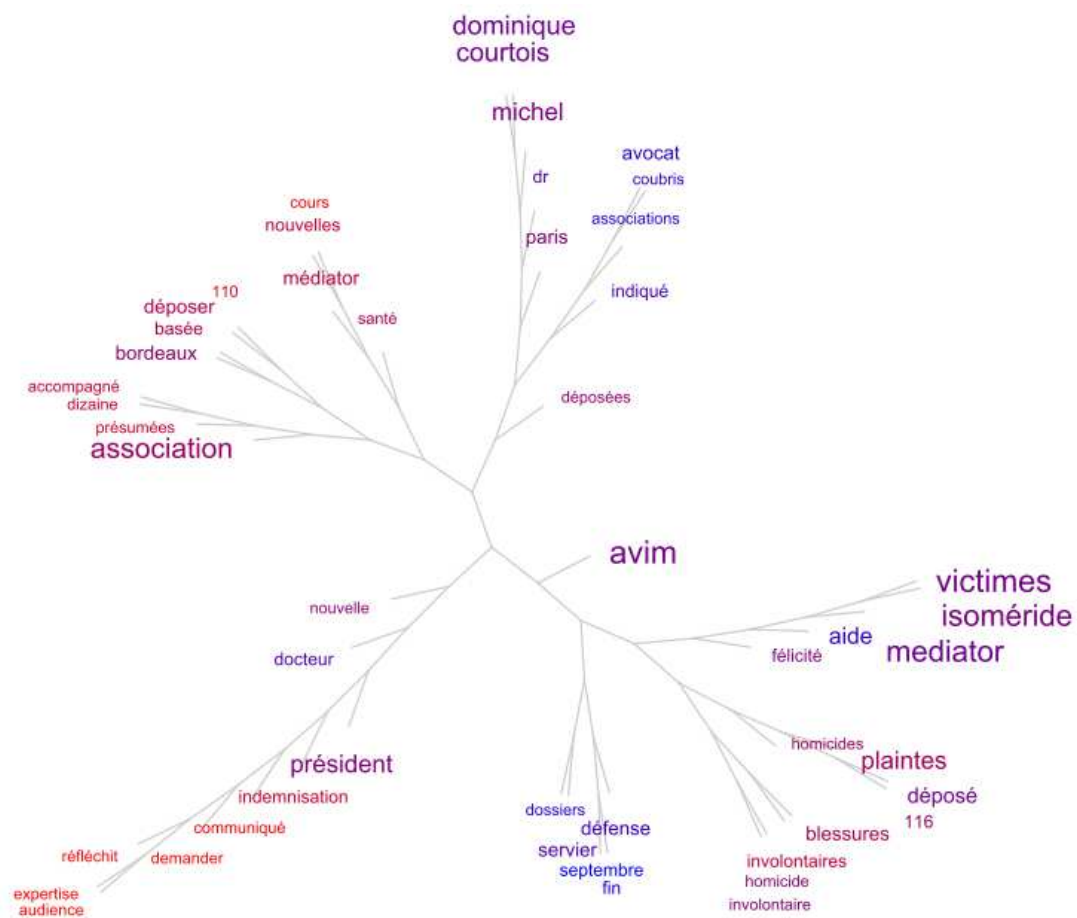
**Figure 17.** Nuage arboré des 50 mots les plus fréquents des contextes (10 mots à gauche et 10 mots à droite) de l'expression *dominique michel courtois* dans l'ensemble du corpus, colorés chronologiquement (en bleu pour les mots de la fin du corpus et en rouge pour les mots du début du corpus), construit par TreeCloud avec la formule Liddell, et des fenêtres glissantes de 20 mots



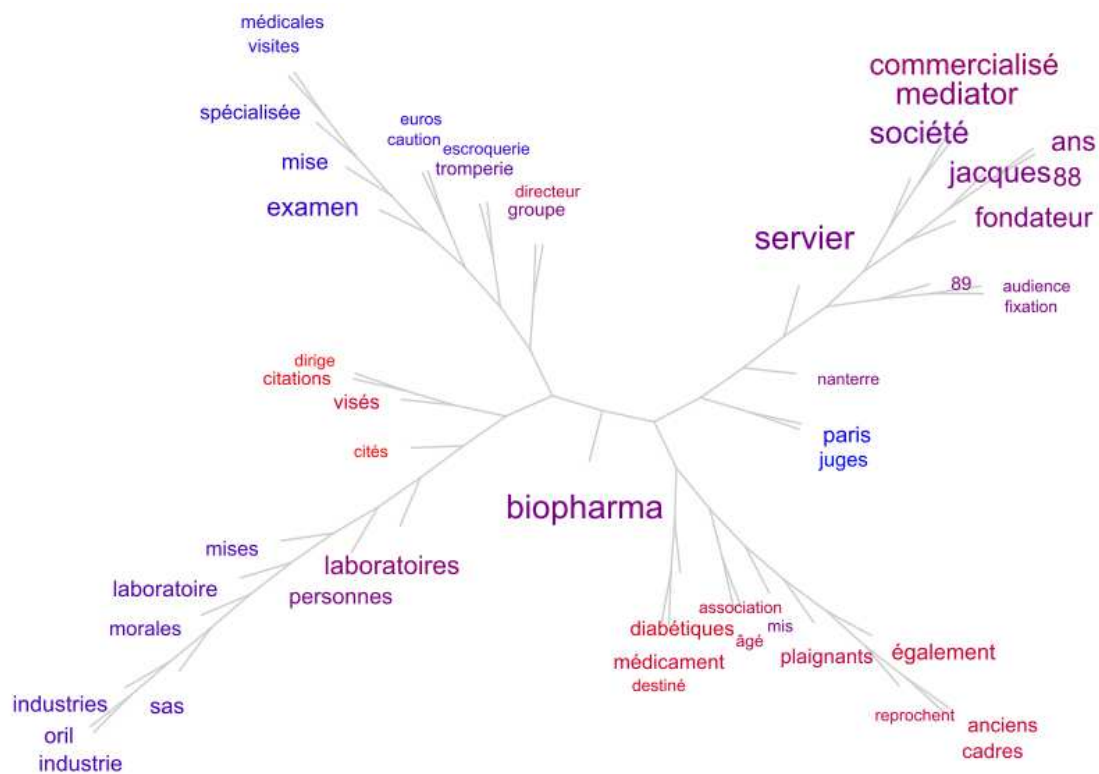
**Figure 18.** Nuage arboré des 50 mots les plus fréquents des contextes (10 mots à gauche et 10 mots à droite) de l'expression *bernard debré* dans l'ensemble du corpus, colorés chronologiquement (en bleu pour les mots de la fin du corpus et en rouge pour les mots du début du corpus), construit par TreeCloud avec la formule Liddell, et des fenêtres glissantes de 20 mots



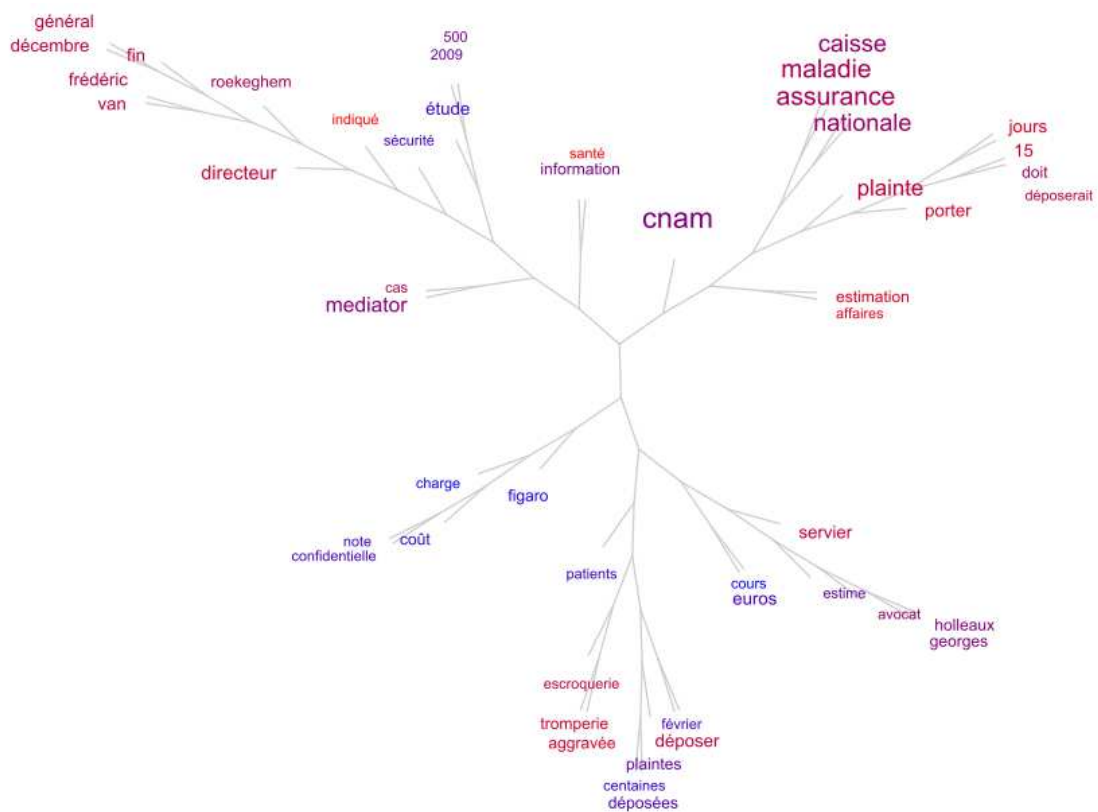
**Figure 19.** Nuage arboré des 50 mots les plus fréquents des contextes (10 mots à gauche et 10 mots à droite) du mot *afssaps* dans l'ensemble du corpus, colorés chronologiquement (en bleu pour les mots de la fin du corpus et en rouge pour les mots du début du corpus), construit par TreeCloud avec la formule Liddell, et des fenêtres glissantes de 20 mots



**Figure 20.** Nuage arboré des 50 mots les plus fréquents des contextes (10 mots à gauche et 10 mots à droite) du mot *avim* dans l'ensemble du corpus, colorés chronologiquement (en bleu pour les mots de la fin du corpus et en rouge pour les mots du début du corpus), construit par TreeCloud avec la formule Liddell, et des fenêtres glissantes de 20 mots

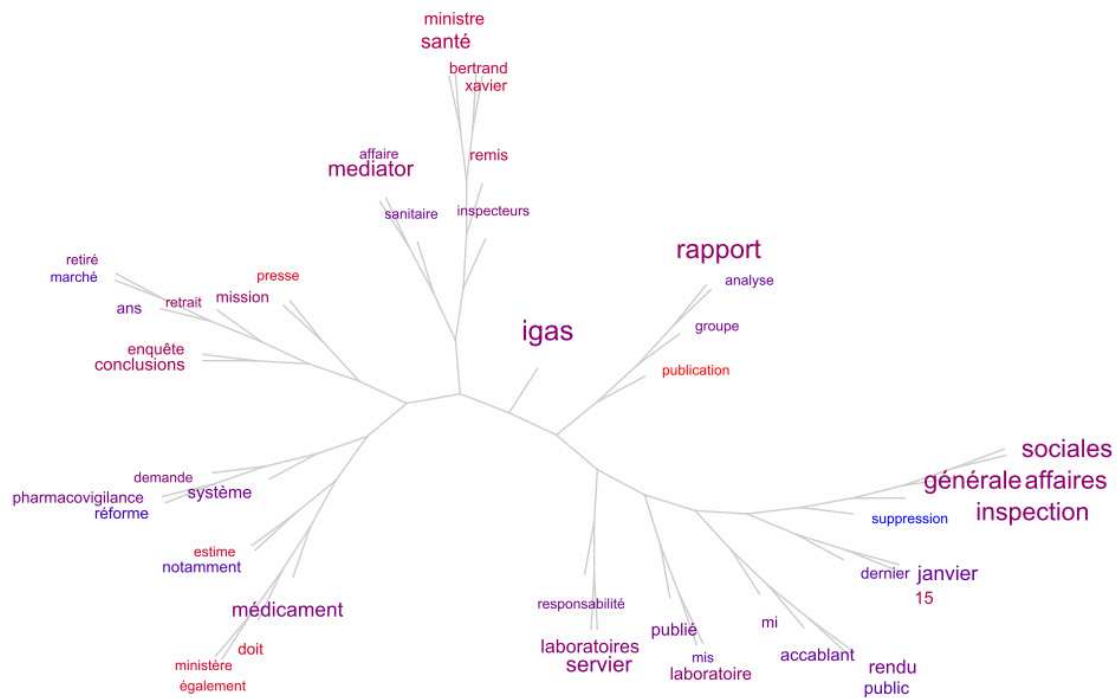


**Figure 21.** Nuage arboré des 50 mots les plus fréquents des contextes (10 mots à gauche et 10 mots à droite) du mot *biopharma* dans l'ensemble du corpus, colorés chronologiquement (en bleu pour les mots de la fin du corpus et en rouge pour les mots du début du corpus), construit par TreeCloud avec la formule Liddell, et des fenêtres glissantes de 20 mots

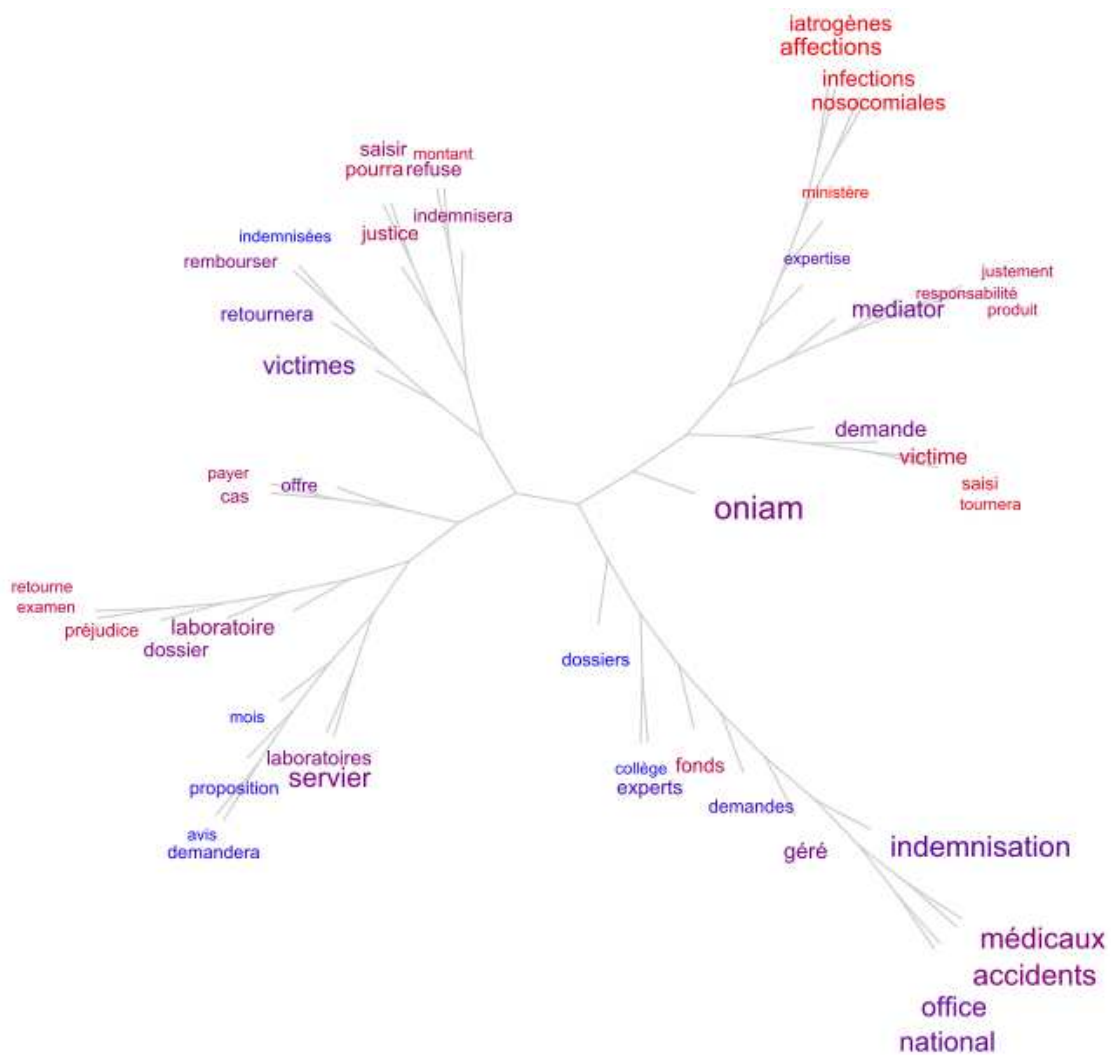


**Figure 22.** Nuage arboré des 50 mots les plus fréquents des contextes (10 mots à gauche et 10 mots à droite) du mot *cnam* dans l'ensemble du corpus, colorés chronologiquement (en bleu pour les mots de la fin du corpus et en rouge pour les mots du début du corpus), construit par TreeCloud avec la formule Liddell, et des fenêtres glissantes de 20 mots





**Figure 23.** Nuage arboré des 50 mots les plus fréquents des contextes (10 mots à gauche et 10 mots à droite) du mot *igas* dans l'ensemble du corpus, colorés chronologiquement (en bleu pour les mots de la fin du corpus et en rouge pour les mots du début du corpus), construit par TreeCloud avec la formule Liddell, et des fenêtres glissantes de 20 mots



**Figure 24.** Nuage arboré des 50 mots les plus fréquents des contextes (10 mots à gauche et 10 mots à droite) du mot *oniam* dans l'ensemble du corpus, colorés chronologiquement (en bleu pour les mots de la fin du corpus et en rouge pour les mots du début du corpus), construit par TreeCloud avec la formule Liddell, et des fenêtres glissantes de 20 mots