

## Analyser des pétitions en ligne : un corpus complexe

- Christine Barats, Céditec, Université Paris Est Créteil, en délégation à l'ISCC, CNRS
- Philippe Gambette, Université Paris-Est, LIGM, UMR8049

# Une analyse pluridisciplinaire

---

## Regards pluriels, objet commun

- **Approches** : sociologique et politique, juridique et socio-discursive
- **Dispositif étudié** : lapetition.be, +12.000 pétitions, 3,25 millions de signatures entre le 31 octobre 2006 et le 12 février 2015
- **Problématiques**: quid du rapport au politique, quid des formes de participation ? D'engagement ? Quid écritures web ? Quid des pratiques transmédiatiques (plateforme pétition, sites, réseaux sociaux, autres formes de relais)

# Articuler différents types de données et de méthodes

---

Diversité des données, approches qualitatives et quantitatives

Entretien concepteur de la base, questionnaires et entretiens auprès des signataires, données issues de la base (plusieurs types de corpus)

## **Zoom données issues de la base :**

- Données alphanumériques (titre de l'appel, texte de l'appel, commentaires, destinataire(s), porteur(s) de l'appel)
- Nombre de signataires, horodatage, début/fin pétition

Exigence d'anonymisation des données (Latzko-Thot, Proulx, 2016 ; Greffet, 2016)

# Quelles approches ?

---

**-Approche longitudinale** : corpus de titres, textes d'appel et commentaires, intérêt (dimension heuristique et diversité des hypothèses) et limite (non stabilité du dispositif dans le temps, formulaire et design)

D'où importance des entretiens avec concepteur et signataires (contextualiser les données)

**-Approche thématique** (11 pétitions survol Bruxelles aéroport ; 147 pétitions scission/unité Belgique, dont 1 pétition=120 000 signatures)

# Données et corpus : apports de l'AD et analyses exploratoires

---

**Corpus global des pétitions**, toutes catégories confondues : **12 522 pétitions** (chacune signée au moins une fois, certaines pétitions n'ont pas de signataire, même pas celle de leur porteur, au moins un mot dans le texte d'appel)

- **9 catégories** proposées sur le site lapetition.be :
  - Art et culture = 652
  - Droits de l'Homme = 1738
  - Environnement, nature et écologie = 1034
    - Humour/Insolite = 397
    - Loisirs = 795
    - **Politique** = 2276
  - Protection animale = 1378
    - Social = 1806
    - **Autres** = 2446

# Rubriques et longueur des textes d'appel (Visualisation : PG)

---

# Premières observations sur les champs sélectionnés par le porteur de la pétition

---

- Variété des interprétations possibles des catégories proposées (rubrique, « à l'attention de ») : usages distancés ou non de ces catégories, possible détournement, de même qu'éventuels mésusages (ne pas surestimer les « compétences »)
- Quid des présupposés et choix des concepteurs (entretiens)
- Catégorie "**Autres**" = le **plus grand nombre de pétitions**=indice difficulté d'appropriation des catégories ? Ou « stratégie » de détournement ?
- « Autres » et volume Texte d'appel : nombre de mots varie de 1 à 8322 mots  
= **Caractère très hétérogène catégorie « Autres »**
- Catégorie "**Politique**" : deuxième rubrique en nombre de pétitions, 60% des textes d'appel > 100 mots

# Premières observations...

## variété des pratiques

---

- **Illustrations** : catégorie “Droits de l’homme” = variété de textes d’appel = difficulté d’appréhension de cette catégorie, ou bien son détournement par des porteurs de projet
- Exemple : titre “allo”, catégorie “Droits de l’Homme”, texte “cava ? 13/05/08”
- Approche « thématique » Survol Bruxelles (2014) : 11 pétitions (4 « politique », 6 « environnement », 1 « art et culture ») dont une pétition=20 000 signataires  
rubrique « politique » (NIMBY : not in my backYard vs montée en généralité
  - Lisibilité des données : test du dispositif par les porteurs, pétitions = un mot = tests (une seule signature ou pas!)  
Exemple : “+”, “ERREUR”, “test”, “f”, “rien”, “RAS”), ou URL= “  
<http://www.blog4ever.com/fichier/94004070306065618.pdf>”  
Ou suppressions de pétition (“x” et “...” sont des textes et des titres de pétitions ayant recueilli 35 signatures)



# Premières observations... hétérogénéité des données

---

- Usages pluriels du dispositif socio-technique :
  - Choix de la rubrique par le porteur de la pétition (usages, mésusages, détournement, test)
  - Présence ou non de commentaires (prise de position, adhésion ou non, écriture web)

## Situer les données quantitatives

**Relative stabilité du dispositif** : évolution de la base et évolution du formulaire de signature et du design

- Matérialité discursive (longueur, aspects morphosyntaxiques) et approches (lexicale, sémantique, énonciative, argumentative) :
  - Titre de l'appel
  - Texte de l'appel
  - Commentaires
  - Signatures

# Premières observations... hétérogénéité des données

---

**Diversité, hétérogénéité des pétitions et des données** : quid des modalités de constitution des corpus, quid de la lisibilité des corpus (Alceste, messages de condoléances, G.Truc, 2011) ?

Choix des outils... et corpus

## Des corpus illisibles ?

- Données alphanumériques (titre de l'appel, texte de l'appel, commentaires, porteur(s) de l'appel)
- Multilinguisme
- Maitrise/registre de la langue
- Étirements, majuscules, graphies, ponctuation, émoticônes sites web URL
- Nombre de signataires, horodatage, début/fin pétition
- **Préparation des corpus et nécessaire adaptation des outils** (cf exposé : JML, développement TextObserver)

# Modalités de constitution des corpus: apports de l'AD

---

Des corpus exploratoires : titres et textes d'appel, texte d'appel et commentaires (survol aéroport)

rubrique=catégorie à interroger

**Questions** : y a-t-il un « genre » pétition avec spécificités ? Quelles sont les caractéristiques des commentaires ? Quid du lexique ? Les formes d'adresse ? Présence de phraséologie ?

Contextualiser les données + apport des entretiens pour approfondir pistes, vérifier des observations (rubrique « animaux »/liens facebook, réseau associatif)

Apport pluridisciplinarité (AD/sc. Po) : les commentaires qui infirment acte d'adhésion (marginal, mais présent)

•Exemple: une pétition en faveur du rattachement de la Wallonie à la France (« rattachement a la France », catégorie « Politique »)

« Vous faite tous chier maintenant!!!!Francophone ou neerlandophone , Si tu n'aime pas ton pays part ! Et toi ,SI TU AIME LA FRANCE , VA Y HABITER !  
MAIS MOI , TU NE M'Y EMÈNERA PAS !!!  
Belge et fier de l'être ! ».

# Diapo à supprimer ?

## Corpus constitués ou en cours de constitution

---

- **Corpus de pétitions en cours d'analyses**
- Pétition scission
- Textes d'appel des pétitions (toutes rubriques si > 10 signatures)
- Titres des textes d'appel des pétitions (toutes rubriques si > 10 signatures)
- Textes d'appel rubrique politique (si > 10 signatures)
- Aéroport, survol Bruxelles - commentaires (pétition 1478)
- 
- **Corpus en cours de constitution**
- Pétition scission
- Pétition Tokio Hotel
- Pétition Catch sur AB4

# Premiers commentaires sur le(s) destinataire(s) explicite(s) de l'appel

---

mention "A l'attention de / Ter attentie van"

## Catégorie « Politique »

« Gouvernement Bruxellois », « ministre de l'immigration », « Ministre Président de la Région Bruxelloise », « Le Ministre de l'Enseignement Supérieur , Jean-Claude Marcourt, ainsi que les différentes présidences de partis politiques. »

## Différents cas de figure, dont destinataire est en fait le signataire potentiel

### Catégorie « Humour/Insolite »

« des étudiants, non-étudiants, trucs et autres plantes... », « tout le monde »,

### Catégorie « Droits de l'Homme »

« victimes décedées »

### Catégorie « Politique »

« Tous », « Schaerbeekois(es) », « tout public »

# Premiers commentaires sur le(s) destinataire(s) explicite(s) de l'appel

---

## « Ton » dans l'appellation du destinataire

« Nos (trop) chers élus » (C:H/I; P: « Pour remplacer Charles Michel par Tatayet »)

« Nos chers politiciens » (C:P ; P: « Les impôts et taxes pour tous »)

« potager » (C:DH ; P: « laissé tranquille les pommes de terre » (1)

« des connards qui ne vive pas avec » (C:P; P: « Immigration + refugies = danger mortel de perte d'identité et de guerre civile ! » (17))

« des enfoires qui ne vive pas avec » (C:P; P: « Stop a l'immigration » (77))

« des menteursssss » (C:P; P: « l'enfant noyé échoué sur une plage » (7))

« Nos députés, qui ne nous écoutent jamais. » (C:P; P: « NON aux immigrés en Belgique, il y en a assez ! » (176))

# Prendre en compte les spécificités du corpus

---

Pétitions bilingues (conception même du site : mention “A l’attention de / Ter attentie van”)

-

Commentaires entièrement dans une autre langue que le français ou bilingues

Commentaires bilingues

\*\*\*\* \*id\_14178 \*idsign\_4381033 TROP DE BRUIT IS TE VEEL !!!

\*\*\*\* \*id\_14178 \*idsign\_4340275 Trop is te veel !

\*\*\*\* \*id\_14178 \*idsign\_4480071 Het was fijn geweest deze petitie ook in het Nederlands te vinden, mais bon...

Étirements et majuscules \*\*\*\* \*id\_14178 \*idsign\_4544902 STOOOOOOOP!!!!

# Quid de la normalisation...

---

\*\*\*\* \*id\_14178 \*idsign\_4448709 (correction des accents)

Nous avons **evite** les communes (moins **cheres**) proches de **l'aeroport** pour **eviter** le bruit.

\*\*\*\* \*id\_14178 \*idsign\_4428472

Bonjour **alor** je trouve **sa** affreux de vouloir **mettre** les **avion** juste **au dessus** de nous oh les **bruit** vont **meme** pas nous laisser dormir et en plus qui nous **garantie** que **aucune** avion va tomber sur nous moi je suis contre **sa** ....

Se pose alors la question de la normalisation orthographique des données : rendre les données utilisables par des logiciels habitués à des graphies standards

Une solution : développer annotation des corpus...



# Premiers résultats d'analyse des textes de pétitions

---

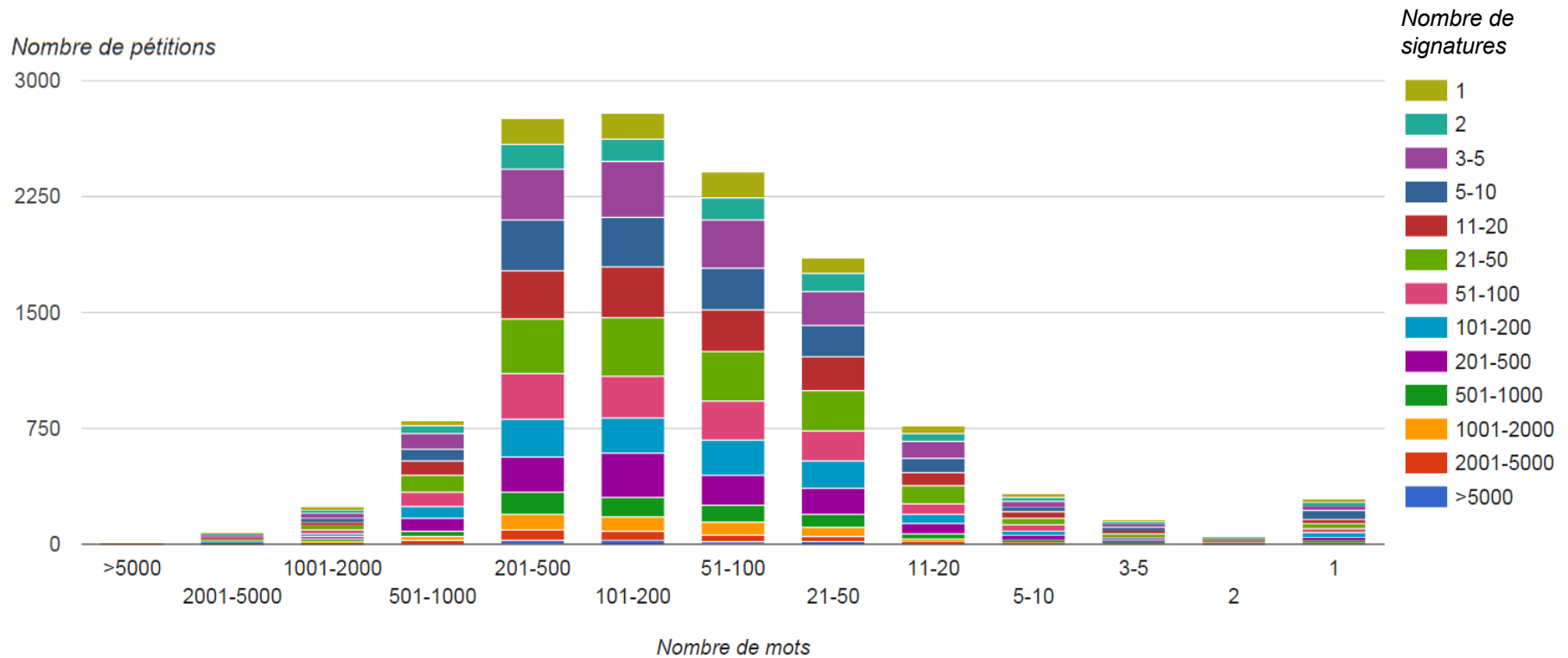
- Lien entre nombre de signatures et nombre de mots
- Lexique des débuts et fins de texte
- Lien entre nombre de signatures et lexique

# Premiers résultats d'analyse des textes de pétitions

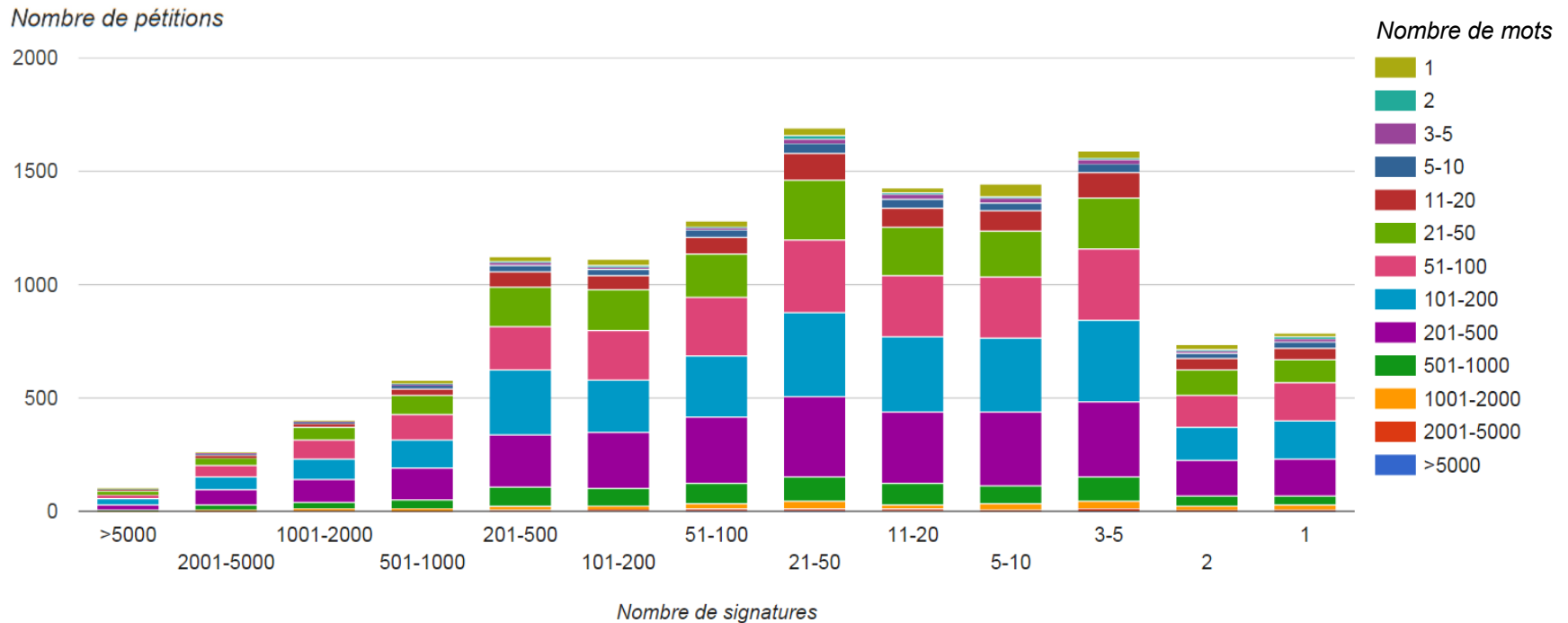
---

- Lien entre nombre de signatures et nombre de mots
- Lexique des débuts et fins de texte
- Lien entre nombre de signatures et lexique

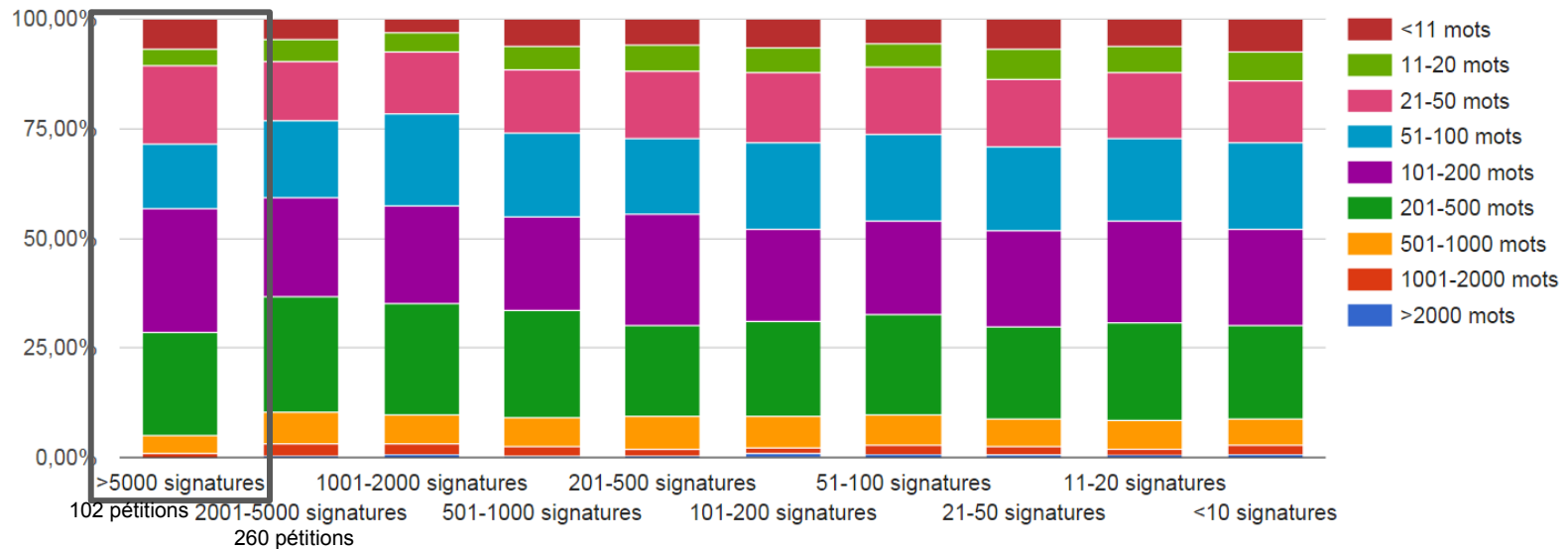
# Nombre de pétitions dans chaque intervalle de nombres de signatures en fonction du nombre de mots de la pétition



# Nombre de pétitions dans chaque intervalle de nombres de mots en fonction du nombre de signatures de la pétition



# Pourcentage de pétitions dans chaque intervalle de nombres de mots en fonction du nombre de signatures de la pétition



# Lexique des débuts et fins de textes

10 mots les plus fréquents des débuts (10 premiers mots) et fins (10 derniers mots) de textes politiques

Visualisation CFP :  
[www.treecloud.org/cfp/](http://www.treecloud.org/cfp/)  
(P. Gambette & N. Lechevrel)

## 10 premiers mots

(144) pétition  
(112) monsieur  
(96) belgique  
(76) ministre  
(50) citoyens  
(49) bonjour  
(47) président  
(46) belge  
(45) gouvernement  
(44) pays  
(41) messieurs  
(40) petition  
(37) stop  
(37) premier  
(36) appel  
(35) belges  
(35) madame  
(34) chers  
(34) france  
(33) peuple

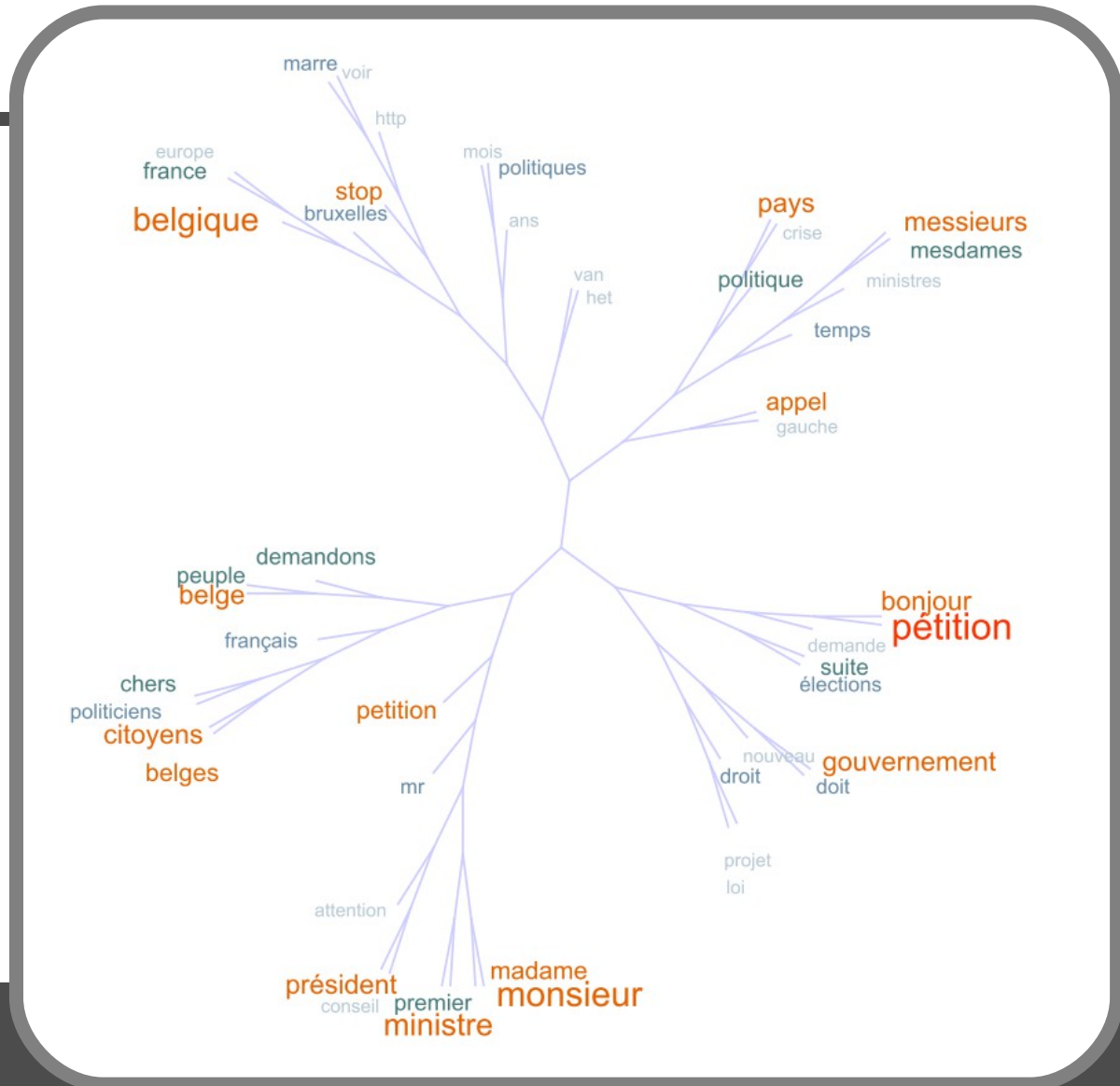
## 10 derniers mots

pétition (121)  
merci (114)  
com (69)  
www (66)  
belgique (64)  
pays (59)  
http (59)  
signer (41)  
be (40)  
france (39)  
signez (38)  
peuple (37)  
droit (33)  
citoyens (31)  
ministre (30)  
fr (29)  
europe (28)  
monde (28)  
gouvernement (28)  
van (27)

# Débuts de textes politiques

50 mots les plus fréquents

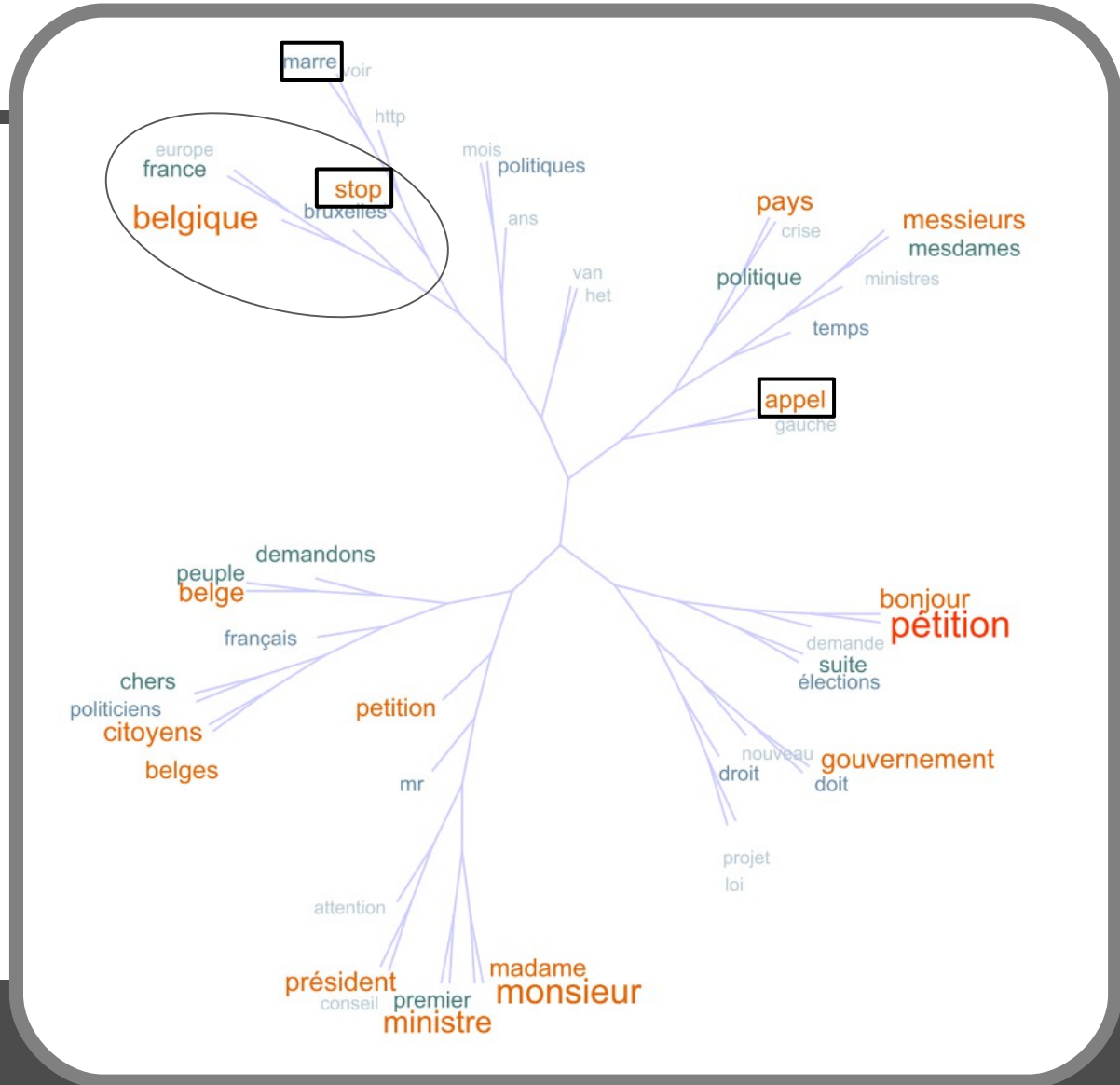
Coloration, taille :  
fréquence parmi les  
début (10 mots) de  
pétitions politiques



# Débuts de textes politiques

50 mots les plus fréquents

Coloration, taille :  
fréquence parmi les  
début (10 mots) de  
pétitions politiques



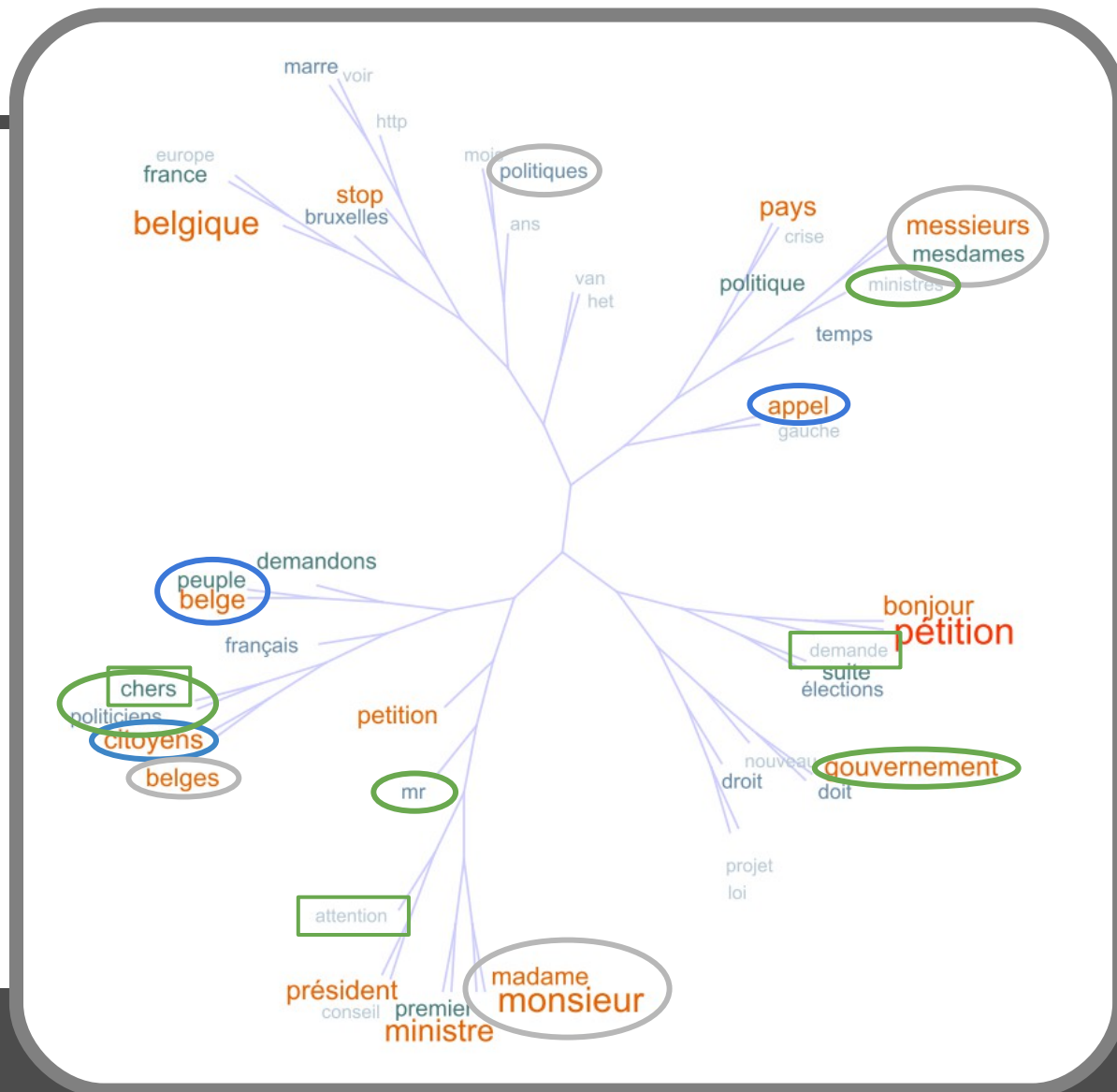


# Débuts de textes politiques

50 mots les plus fréquents

Coloration, taille :  
fréquence parmi les  
débuts (10 mots) de  
pétitions politiques

décideurs  
signataires



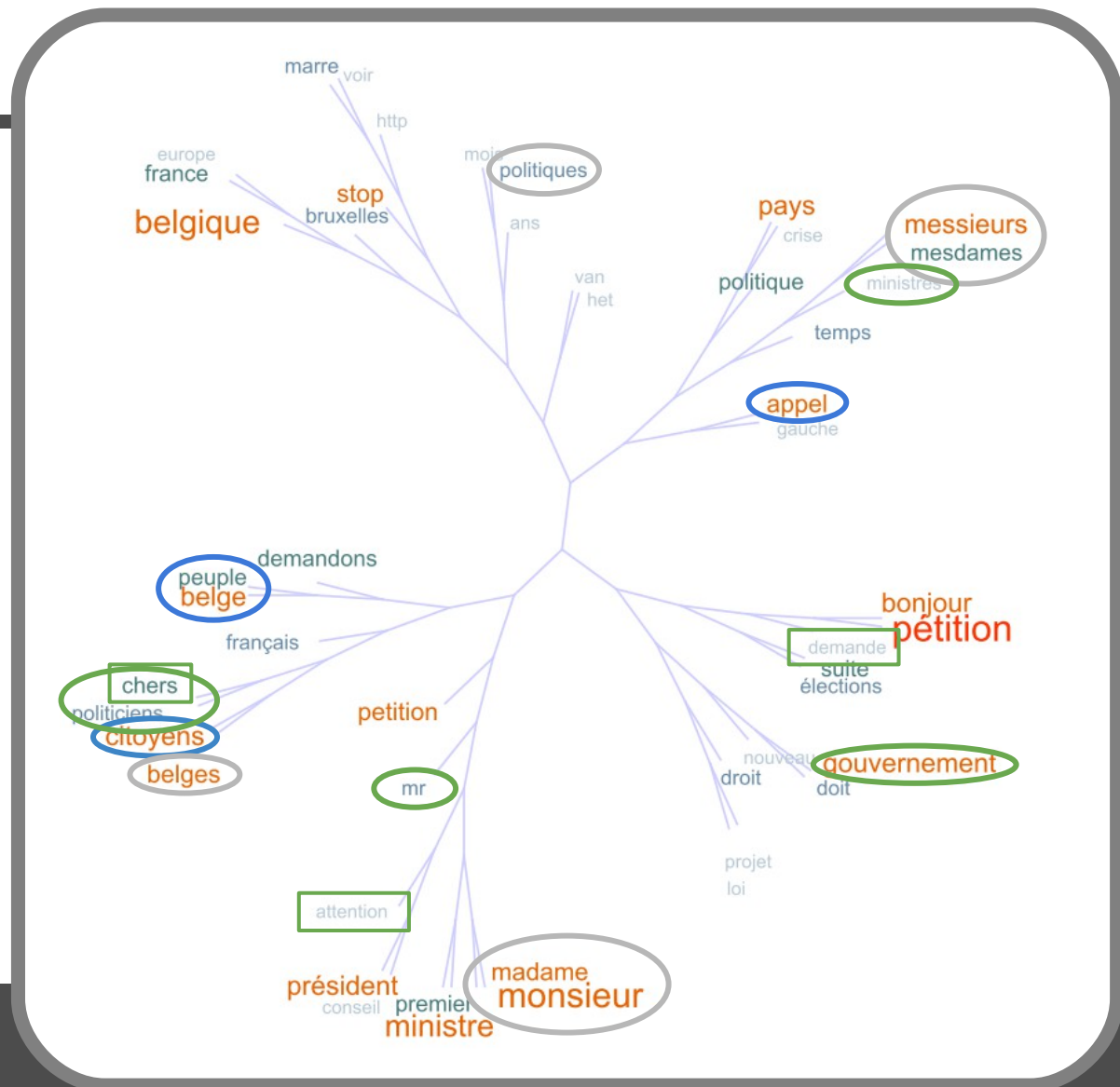
# Débuts de textes politiques

## Relevé manuel :

6,5% d'adresse aux destinataires des pétitions

3,8% d'adresse aux signataires des pétitions

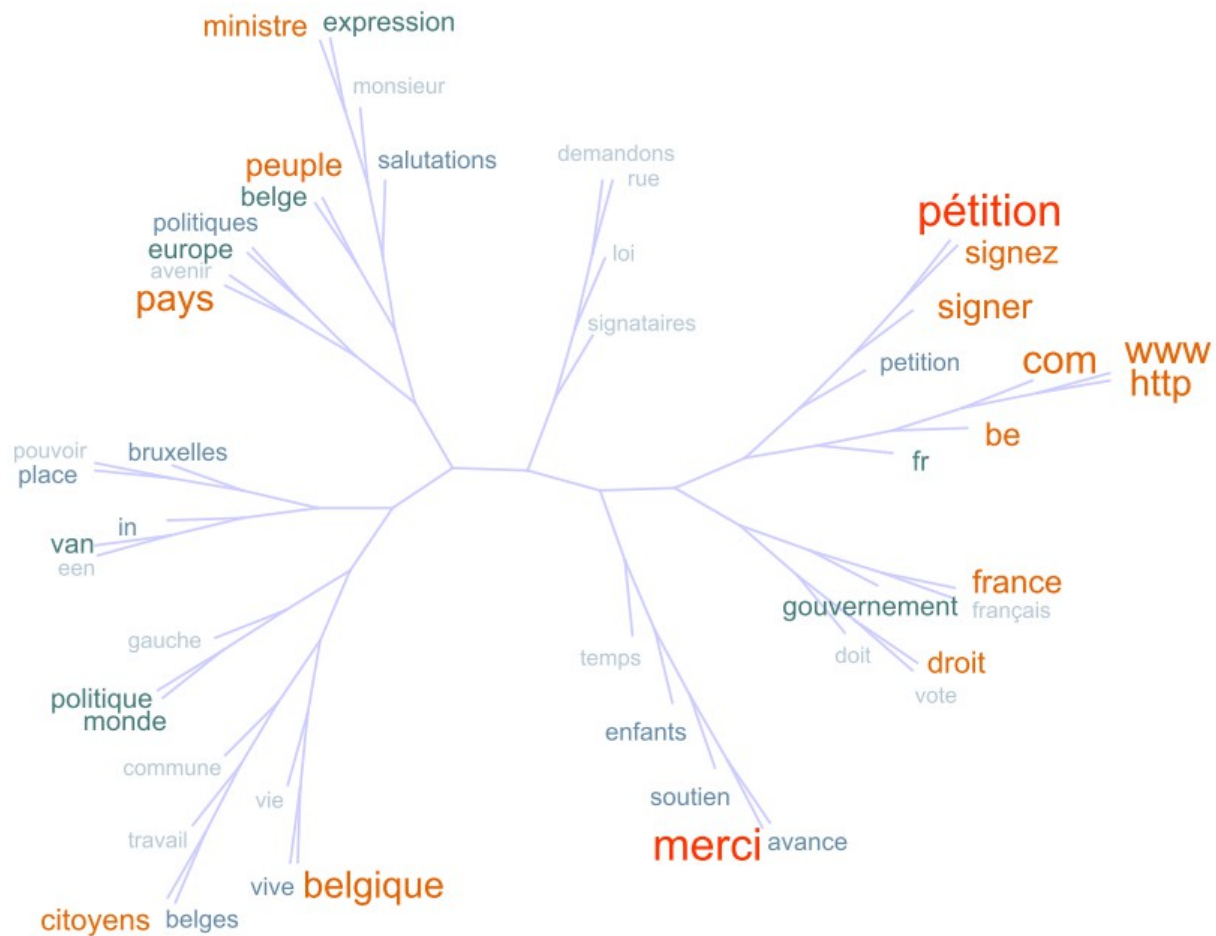
décideurs  
signataires



# Fins de textes politiques

50 mots les plus fréquents

Coloration, taille : fréquence parmi les fins (10 mots) de pétitions politiques



# Fins de textes politiques

50 mots les plus fréquents

Coloration, taille : fréquence parmi les fins (10 mots) de pétitions politiques



# Fins de textes politiques

50 mots les plus fréquents

Coloration, taille : fréquence parmi les fins (10 mots) de pétitions politiques



## Relevé de http ou www :

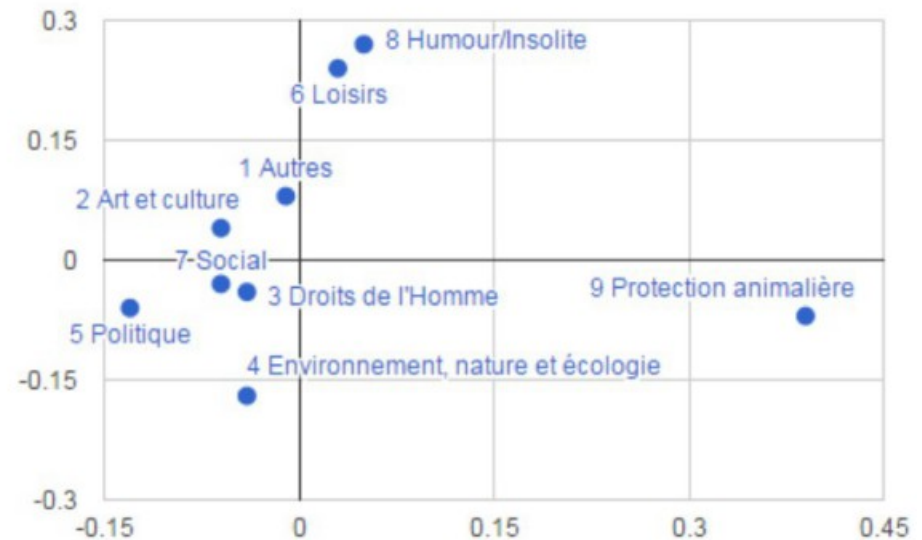
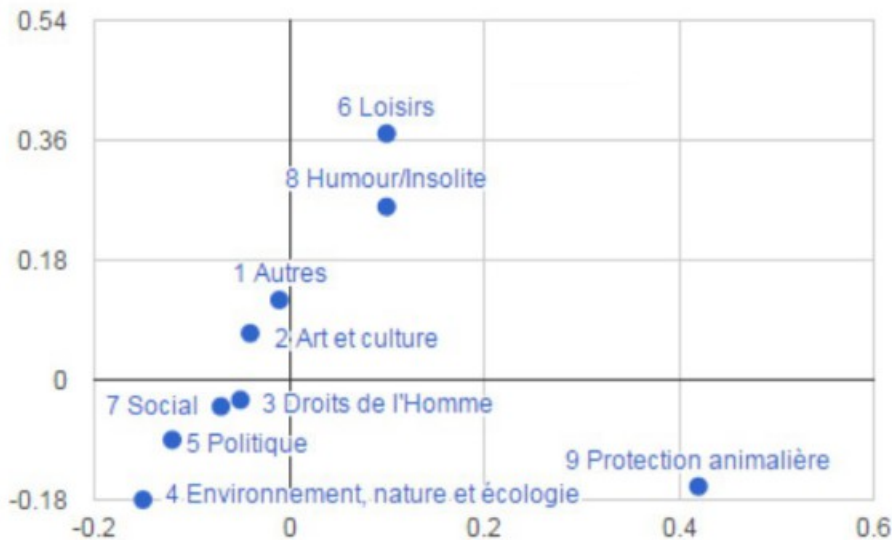
dans 4.1% des fins de textes politiques

dans 1.1% des débuts de textes politiques

(2.9% des fins "loisirs", 5.9% des fins "animaux" ;

0.6% des débuts "loisirs", 2.1% des débuts "animaux")

# Lien entre rubrique et débuts et fins de texte

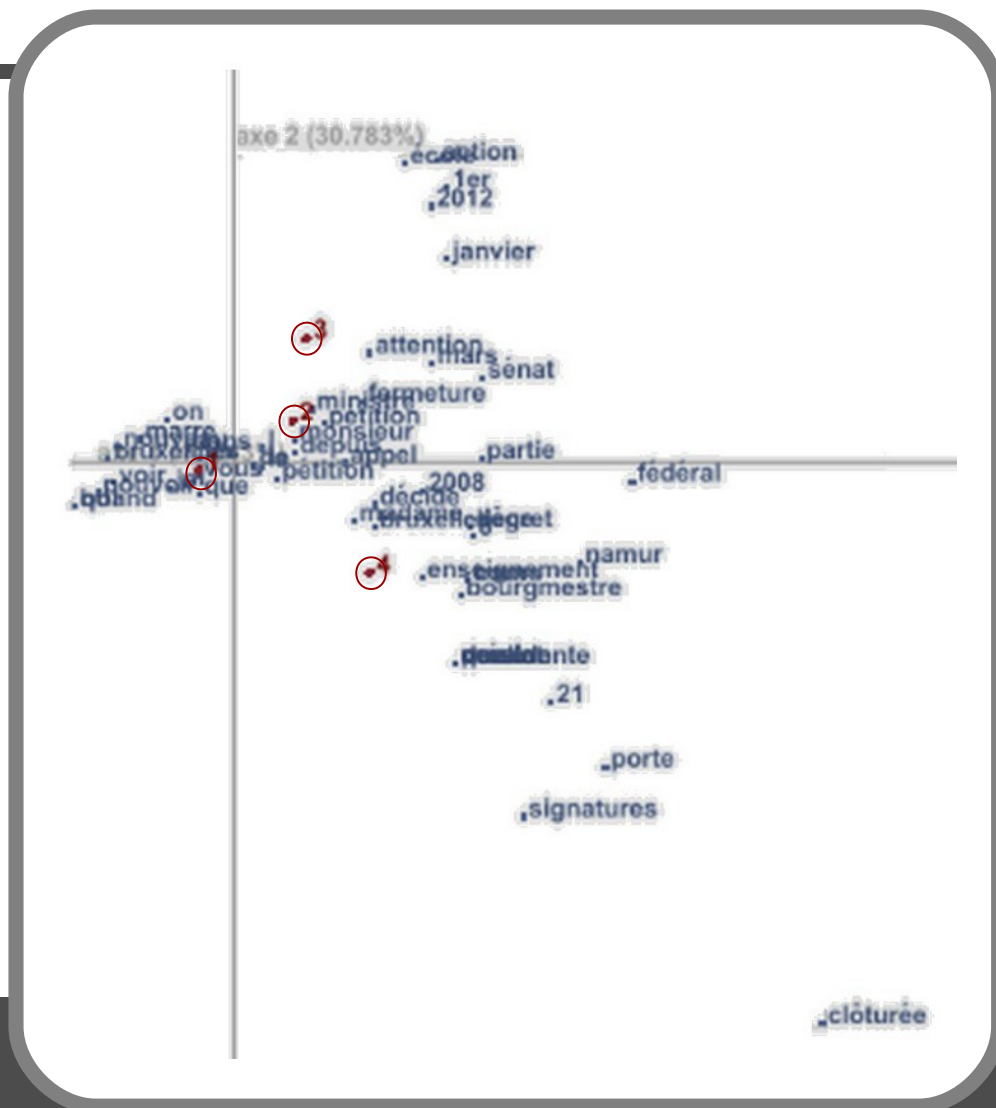


Analyses factorielles des correspondances sur les 10 premiers mots (à gauche) et les 10 derniers mots (à droite) des textes de pétitions, réunis par rubrique

# Début des pétitions politiques

Analyse factorielle des correspondances des **débuts** de textes politiques (10 premiers mots)

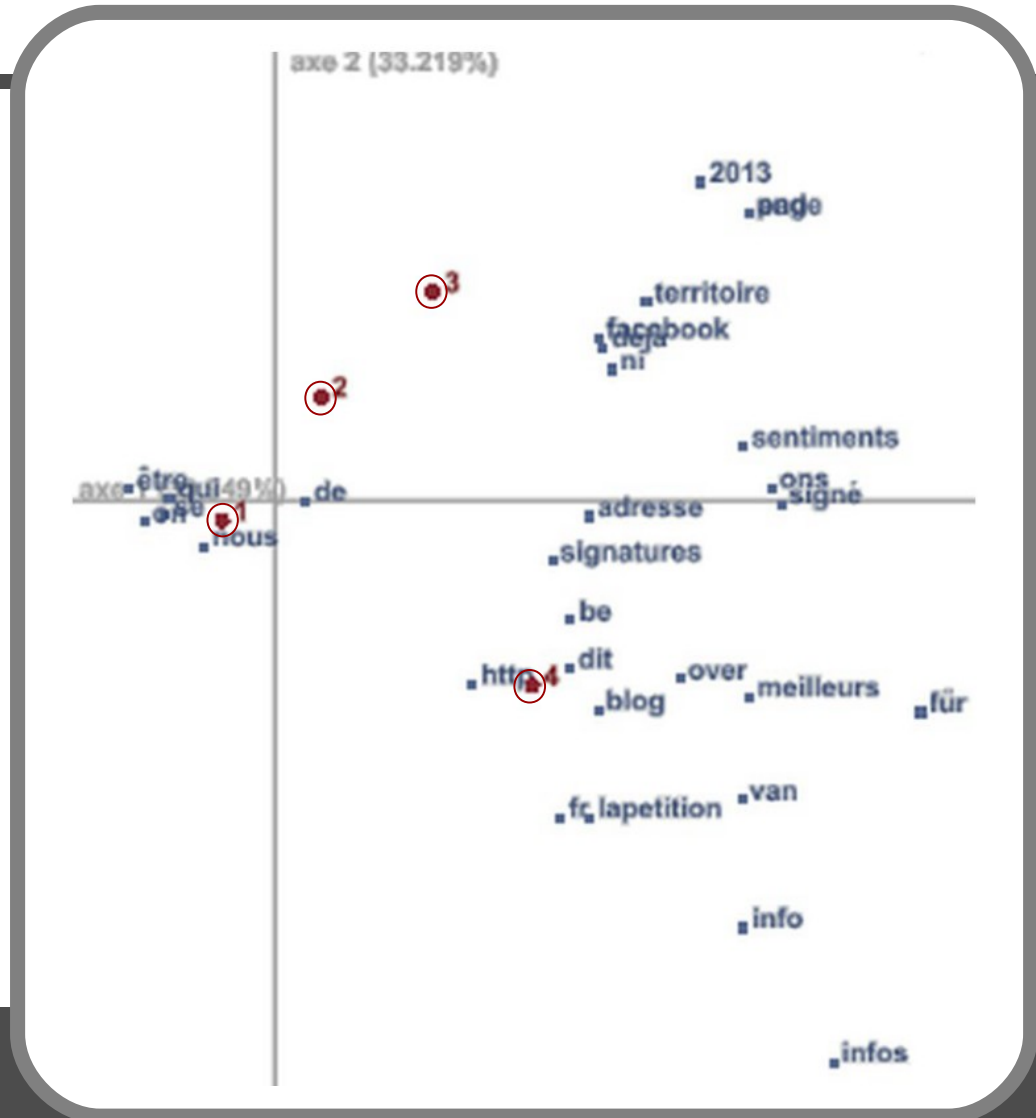
- ① : 1 à 49 signatures (1581)
- ② : 50 à 199 signatures (341)
- ③ : 200 à 499 signatures (164)
- ④ : 500 signatures et plus (190)



# Fin des pétitions politiques

Analyse factorielle des correspondances des fins de textes politiques (10 premiers mots)

- ① : 1 à 49 signatures (1581)
- ② : 50 à 199 signatures (341)
- ③ : 200 à 499 signatures (164)
- ④ : 500 signatures et plus (190)



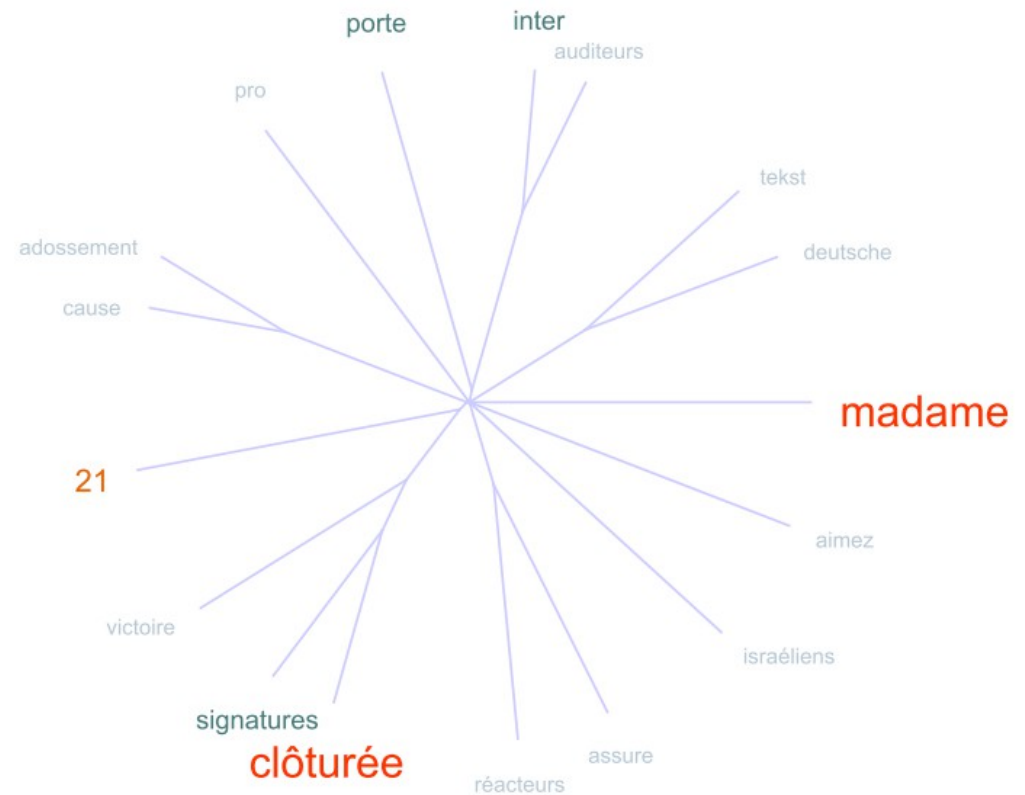


# Début des pétitions politiques les + signées

Vocabulaire **spécifique** des **débuts** (10 mots) de pétitions **politiques** signées **500 fois ou plus** par rapport aux autres débuts de pétitions politiques

Score de spécificité au moins égal à 2

Coloration, taille : fréquence parmi les débuts de pétitions politiques signées 500 fois ou plus

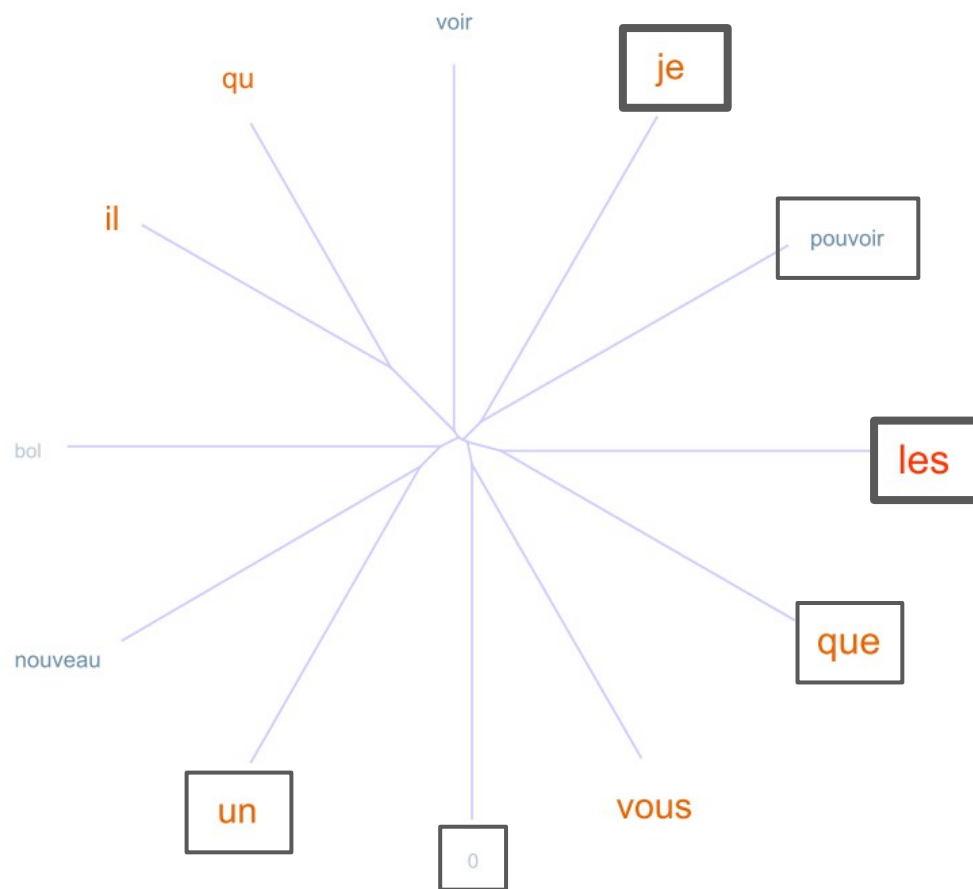


# Début des pétitions politiques les + signées

Vocabulaire **spécifique** des **débuts** (10 mots) de pétitions **politiques** signées **moins de 50 fois** par rapport aux autres débuts de pétitions politiques

Score de spécificité au moins égal à 1.5

Coloration, taille : fréquence parmi les débuts de pétitions politiques signées moins de 50 fois



# Début des pétitions politiques les + signées

Vocabulaire **spécifique** des **débuts** (10 mots) de pétitions **politiques** signées **moins de 50 fois** par rapport aux autres débuts de pétitions politiques

Score de spécificité au moins égal à 1.5

Coloration, taille : fréquence parmi les débuts de pétitions politiques signées moins de 50 fois

