

Analyse sémantique computationnelle des débats publics : objectifs et modèles

Sabine Ploux, CAMS EHESS-CNRS
sabine.ploux@ehess.fr

Master Humanités numériques et computationnelles de PSL





Plan du cours

1. Une courte introduction sur les débats publics
2. Les analyses s'appuient sur
 - ▶ Analyse du discours
 - ▶ Traitement automatique du langage
 - ▶ Les humanités numériques : croiser l'analyse des textes à des données historiques et géographiques, démographiques, économiques...
3. Quelques modèles et méthodes
4. Le Grand débat national, exemples de résultats
5. Perspectives

QUELQUES REPÈRES SUR LES GRANDES CONSULTATIONS D'OPINION



- ▶ Les Cahiers de doléances, 1789
- ▶ Revendications locales au moment de la création des départements (1790-1799) (cf. M-V. Ozouf-Marignier)
- ▶ Le Grand débat national (2018-2019)

L'EXEMPLE DU GRAND DÉBAT NATIONAL LANCÉ EN RÉPONSE AU MOUVEMENT DES GILETS JAUNES (2018-)



- ▶ Les ronds-points et le mouvement des Gilets jaunes
- ▶ L'opération "Mairies ouvertes" lancée par l'AMRF le 8 décembre 2018 et les Cahiers de doléances
- ▶ Lancement du GDN (14 janvier 2019)
 - ▶ Réunions locales,
 - ▶ Plateforme du GDN, plateformes alternatives...
 - ▶ Cahiers citoyens

LES CAHIERS CITOYENS

Initiative lancée par l'Association des maires ruraux de France (Cahiers de doléances) et reprise par le gouvernement (Cahiers citoyens).

Ouverts dans les mairies :

- ▶ Pas de question mais remplis à partir de mi-janvier dans le contexte du Grand débat national
- ▶ Contributions manuscrites ou dactylographiées
- ▶ \cong 19247 Cahiers pour \cong 16500 communes / 34945 communes françaises

LA PLATE-FORME DU GDN

[HTTPS://GRANDDEBAT.FR/](https://granddebat.fr/)

- ▶ Un ensemble de questions à choix multiples et de 82 questions "ouvertes"
- ▶ les contributeurs :
 - ▶ des personnes
 - ▶ synthèses de réunion
- ▶ Le GDN comprend 4 thèmes
 - ▶ Transition écologique
 - ▶ Démocratie et citoyenneté
 - ▶ Fiscalité et dépenses publiques
 - ▶ Organisation de l'état et des services publics

EXEMPLE : LES QUESTIONS DU THÈME FISCALITÉ

1. Quelles sont toutes les choses qui pourraient être faites pour améliorer l'information des citoyens sur l'utilisation des impôts ?
2. Que faudrait-il faire pour rendre la fiscalité plus juste et efficace ?
3. Afin de réduire le déficit public de la France qui dépense plus qu'elle ne gagne, pensez-vous qu'il faut avant tout :
 - ▶ Réduire la dépense publique
 - ▶ Augmenter les impôts
 - ▶ Faire les deux en même temps
4. **Quels sont selon vous les impôts qu'il faut baisser en priorité ?**
5. Afin de baisser les impôts et réduire la dette, quelles dépenses publiques faut-il réduire en priorité ?
 - ▶ Les dépenses de l'État
 - ▶ Les dépenses sociales
 - ▶ Les dépenses des collectivités territoriales

LES AUTRES PLATES-FORMES



Entendre la France (EF, www.entendrelafrance.fr/) dirigé vers un public jeune

- ▶ Disponible sur les réseaux sociaux
- ▶ Mêmes thèmes que le GDN
- ▶ Simplification des questions (49 pour 82 (GDN))
- ▶ Possibilité de contribution libre hors thème

L'OBSERVATOIRE DES DÉBATS

<https://observdebats.hypotheses.org/>

"L'observation a ... porté ... sur les réunions locales déclarées.."

"Les observateur.rice.s bénévoles ont utilisé une grille d'observation commune afin d'homogénéiser autant que possible les observations."

"Un court questionnaire était ... proposé aux participants afin d'avoir des données socio-démographiques et de positionnement politique sur le public des débats."

QUELQUES CHIFFRES SUR LA TAILLE DES CORPUS

A faint, light-blue network diagram is overlaid on the slide. It consists of numerous nodes (represented by small grey squares) connected by thin lines, forming a complex web. Some nodes are circled in blue. The diagram is centered behind the main text and tables.

Les plateformes numériques

	Contributions (*10 ³)
Le Grand débat national	569
Entendre la France	55
Le vrai débat	25

Les Cahiers

	Contributions (*10 ³)
Cahiers de doléances/citoyens	225

DEUX QUESTIONS



1. Incidence du média sur le contenu des contributions et le profil des contributeurs et contributrices
2. Analyse du contenu des contributions et croisement avec des variables exogènes.

La question du média

- ▶ Le choix de la mairie, lieu d'organisation locale, (CC) ou celui d'une plate-forme ouverte pour le recueil des contributions pourrait avoir un impact différencié sur l'expression des contributeurs.
- ▶ le choix d'un cahier vs celui du numérique.
- ▶ Les profils des contributeurs pourraient être également hétérogènes.

Analyse du discours

- ▶ Le discours c'est un texte replacé dans son contexte d'énonciation
- ▶ Développement de méthodologies dérivées de théories en SHS (linguistique, politique, sociologie...), utilisation de méthodes statistiques et d'analyse des données sur les textes (lexicométrie, textométrie, logométrie)
 - ▶ Étude des mots, des thèmes, de l'argumentation...
 - ▶ Étude comparative des discours pour mettre en relief leur spécificité
 - ▶ ...

UN ÉTUDE TEXTOMÉTRIQUE DU CAHIER CITOYEN DE DOLE, JURA, MARION BENDINELLI

La détection de marqueurs linguistiques permet de conclure que

- ▶ les formats manuscrit ou dactylographié recrutent différents profils énonciatifs.
- ▶ les contributions manuscrites convoquent plus d'éléments relatifs à l'intime et à l'ancrage local des contributeurs (parcours de vie...), les contributions dactylographiées plus d'éléments factuels et une structure plus explicitement argumentative.

Sens et matérialités des contributions citoyennes. Analyse textométrique et discursive du cahier citoyen de Dole (Jura) -Marion Bendinelli (Université Bourgogne Franche-Comté, EA4661 ELLIADD, USR 3124 MSHE Claude-Nicolas Ledoux) & Anaïs Rico-Perrier (Université Bourgogne Franche-Comté)

QUELQUES OBJECTIFS DU POINT DE VUE SÉMANTIQUE ET COMPUTATIONNEL



- ▶ De quoi parle-t-on ? : repérage de mots et de thèmes et de leurs champs lexicaux
- ▶ Qu'est-ce qu'on en dit ? : repérage de profils de phrases, de constructions argumentales typiques
- ▶ Repérer des propositions originales (?)
- ▶ Contraster des sous-corpus et croiser les résultats avec d'autres variables : diachronie, spatialité, contributeurs

QUELQUES QUESTIONS SOULEVÉES D'UN POINT DE VUE SÉMANTIQUE : LIENS ENTRE FORMES D'EXPRESSION ET CONTENU

- ▶ La paraphrase et la synonymie : on peut dire des choses similaires de façon différente
 - ▶ *Longtemps je me suis couché de bonne heure ; Je me mettais au lit tôt, avant*
- ▶ La polysémie : un même mot peut avoir plusieurs sens suivant son contexte d'emploi, les habitudes des locuteurs, etc.
 - ▶ *Liberté d'expression* : possibilité d'exprimer un contenu sans entrave
 - ▶ *Liberté de langage* : franc-parler (forme et contenu)

IDÉE D'UTILISER LA PROXIMITÉ VECTORIELLE COMME MODÈLE DE LA SIMILARITÉ SÉMANTIQUE



Figuration des proximités,

<https://www.mn.uio.no/ifi/studier/masteroppgaver/lfg/evaluating-similarity-models.html>

La question du prétraitement

- ▶ Pour un corpus lemmatisé : le nombre des variantes des verbes est élevé par rapport aux autres catégories grammaticales. La lemmatisation augmente leur fréquence d'occurrence et de cooccurrence et la représentativité des verbes.
- ▶ Sur un corpus non lemmatisé : on tient compte des sens particuliers (relation contexte-variante)
 - ▶ travail : *travail de nuit* \neq *travaux de nuit*, *travail à domicile* \neq *travaux à domicile*, *réforme du travail*, *réforme des travaux* ? (*travail* : l'activité en elle-même)
 - ▶ travaux : *travaux publics*, *travail public* ?, *travaux de rénovation* \neq *travail de rénovation* (les réalisations particulières de cette activité)

Choisir une fenêtre

- ▶ Une fenêtre glissante de n mots (HAL, Glove)
- ▶ La phrase (moyenne des longueurs environ 15, dépend des langues)
- ▶ Le paragraphe
- ▶ Le document (LSA)

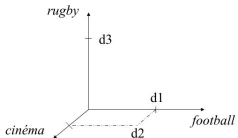
Espace avant réduction



Le modèle vectoriel

Visualisation

Chaque dimension correspond à un des mots du vocabulaire
Chaque document correspond à un point dans l'espace, la requête aussi



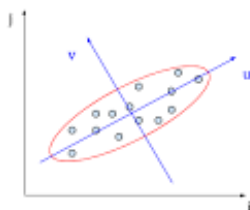
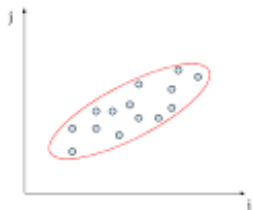


Réduction de matrices

- ▶ **L'idée** : changer de repère pour s'adapter à la forme du nuage de mots et fenêtres (phrases, paragraphes, documents) et ne garder que les dimensions les plus pertinentes
- ▶ **Méthode** : recherche les vecteurs et valeurs propres



Recherche des vecteurs propres



Principal Component Analysis - Représentation de données



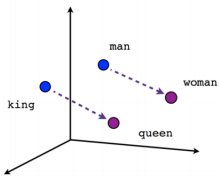
L'exemple de Word2vec

Deux architectures fondées sur le contexte d'emploi des mots

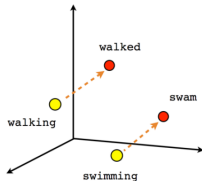
- ▶ *CBOW* (continuous bag-of-words) prédit le mot étant donné un contexte de mots
- ▶ *Skip-Gram* apprend le contexte étant donné un mot

Mikolov (2013) Efficient Estimation of Word Representations in Vector Space
<https://code.google.com/archive/p/word2vec/>

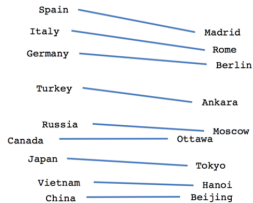
Exemples de résultats



Male-Female



Verb tense

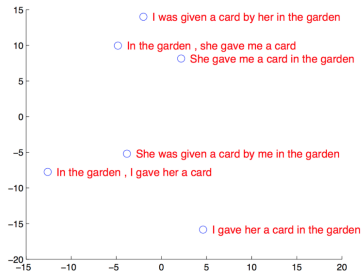
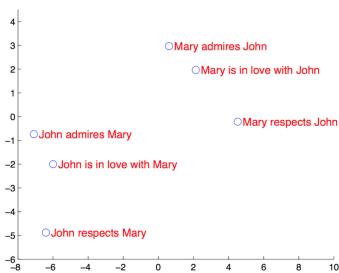


Country-Capital

Obtenus par projection à partir d'une analyse en composantes
principales *Word2vec*

Exemples de résultats avec un réseau LSTM

Après une réduction de l'espace une analyse en composantes principales



Sutskever et al. (2014) Sequence to Sequence Learning with Neural Networks

Créer des représentations pré-entraînées riches pour de multiples tâches

	Taille corpus (ordre de grandeur) (ordre de grandeur)	Nombre de paramètres (ordre de grandeur)
Elmo	10^7 à 10^8	10^8
Bert	$3,3 \times 10^9$	$3,45 \times 10^8$ (version Large)
GPT-3	10^{11}	$1,75 \times 10^{11}$

Nombre de lemmes d'une langue : environ 5×10^4 à 10^5
 nombre de flexions du français : environ 5×10^5 .

LA QUESTION DE LA POLYSÉMIE



Représenter des unités de sens plus fines que le sens des mots

- ▶ Vecteurs de contexte (BERT)
 - ▶ Yenicelik et al. (2020) *How does BERT capture semantics? A closer look at polysemous words*
 - ▶ Gari Soler et al. (2021) *Let's Play Mono-Poly: BERT Can Reveal Words' Polysemy Level and Partitionability into Senses*
- ▶ Modèles géométriques fondés sur les graphes de liens lexicaux

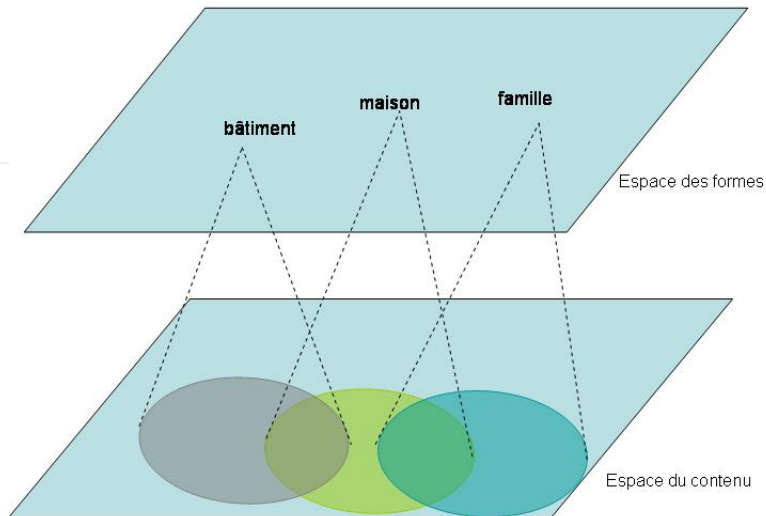
Le modèle géométrique des Atlas sémantiques



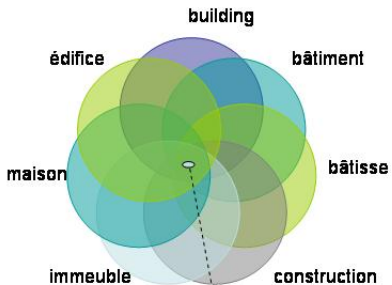
Les deux niveaux d'organisation

- ▶ celui des formes : mots
- ▶ celui d'un contenu représenté par un artefact : les cliques

Le modèle géométrique des Atlas sémantiques



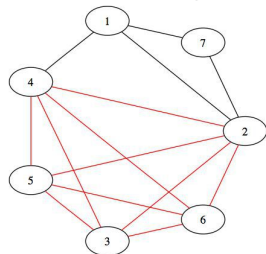
Les cliques



building, bâtiment, bâtisse, construction, immeuble, maison, édifice

Les cliques et la théorie des graphes

- ▶ Une clique est un sous-graphe maximal complet connexe
- ▶ Un algorithme de recherche de toutes les cliques maximales d'un graphe est coûteux





Continuité et topologie

- ▶ 21 *cruel, dur, impitoyable, implacable, inexorable, inflexible, insensible, sévère*
- ▶ 34 *dur, froid, impitoyable, implacable, insensible, sévère*
- ▶ 35 *dur, froid, inaccessible, indifférent, insensible*
- ▶ 39 *dur, impassible, indifférent, insensible, stoïque*
- ▶ 15 *calme, flegmatique, froid, impassible, imperturbable, insensible*
- ▶ 16 *calme, froid, inanimé, insensible*
- ▶ 63 *froid, inanimé, inerte, insensible*
- ▶ 83 *inanimé, inerte, insensible, mort*



Calcul de l'espace sémantique pour une unité

- ▶ Matrice *mots * cliques*
- ▶ Analyse factorielle des correspondances
 - ▶ normalisation : profils lignes et colonnes
 - ▶ réduction de matrice
 - ▶ projection des cliques et des mots sur un même espace

baraque
 bicoque
 boîte
 cabane
 cahute
 case
 cassine
 chaumière
 ... +14
 (Constellation 3 Français)

appartement
 bâtiment
 bâtisse
 bouge
 building
 chalet
 construction
 établissement
 ... +6
 (Constellation 4 Français)

abri
 asile
 bercail
 campagne
 château
 chez-soi
 clapier
 clinique
 ... +24
 (Constellation 5 Français)

domesticité
 gens
 monde
 serviteur

branche
 descendance
 domestique
 lignée
 maisonnée
 parents
 race
 suite

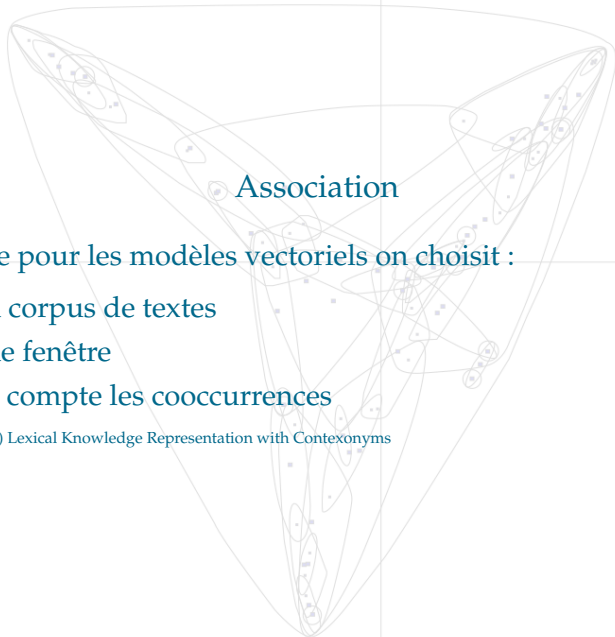
bas-lieu
 famille
 feu
 ménage
 naissance
 nom
 origine

maison



Tableau de transcription

- ▶ clique : point
- ▶ mot : enveloppe des cliques qui le contiennent
- ▶ valeur sémantique : région de l'espace qui correspond à un cluster obtenu par classification sur le nuage de points
- ▶ valeur sémantique générique si elle existe : cluster proche de l'origine des axes
- ▶ valeurs spécifiques : clusters excentrés
- ▶ proximité sémantique entre 2 mots : proximité sur la carte et recouvrement des enveloppes correspondantes



Association

Comme pour les modèles vectoriels on choisit :

- ▶ un corpus de textes
- ▶ une fenêtre
- ▶ on compte les cooccurrences

Ji et al. (2003) Lexical Knowledge Representation with Contextonyms

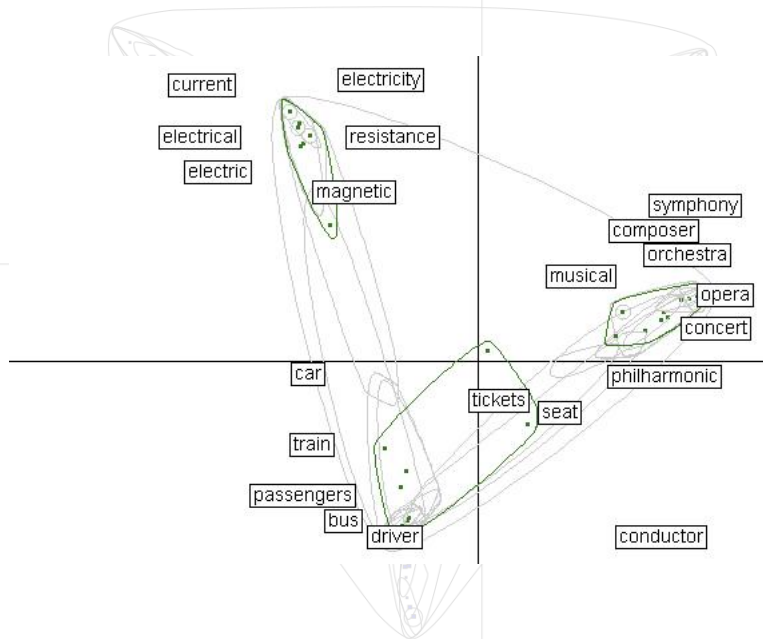


Algorithme

mots	co-occurrents sélectionnés	rejetés
m_i	c_1, c_2, \dots, c_k	c_{k+1}, \dots, c_n
c_1	g_1, g_2, \dots, g_l	g_{l+1}, \dots, g_m
..
c_k	h_1, h_2, \dots, h_q	h_{q+1}, \dots, h_p

Les co-occurrents sont ordonnés en fonction de leur fréquence de cooccurrence décroissante. On sélectionne une portion $1 \geq \alpha \geq 0$ des co-occurrents du mot initial m_i et une portion $1 \geq \beta \geq 0$ des co-occurrents des co-occurrents c_j de m_i .

- ▶ On calcule les cliques d'association à partir des co-occurrents sélectionnés
- ▶ on applique une AFC sur la matrice des *mots * cliques*



ANALYSE DES CORPUS DU GDN



- ▶ Étude de la participation
- ▶ Synthèses par question posée
- ▶ Classification, représentation des thèmes abordés
- ▶ Analyse du corpus à travers un thème (exemple : transport), analyse du discours...

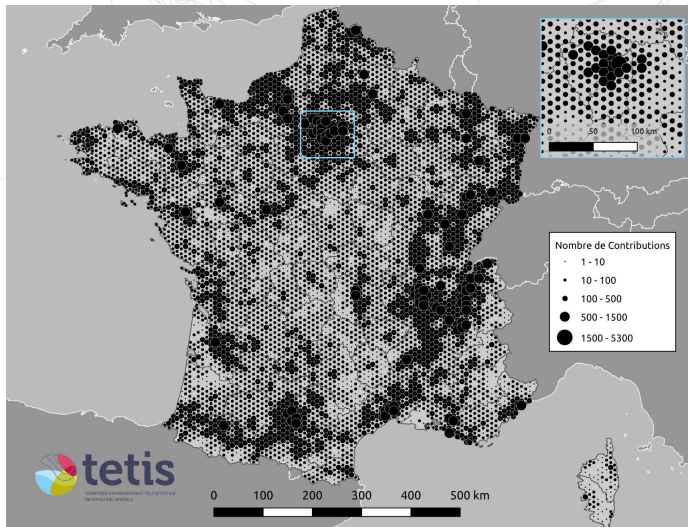
PARTICIPATION SUR LA PLATEFORME DU GDN

Les synthèses de la mission du Grand débat national :

- ▶ Le nombre de contributions issues des questionnaires et des propositions
- ▶ L'évolution du nombre de contributions par jour
- ▶ La répartition des répondants par région, par département, par type de commune (rurale, urbaine < 20000, 20000 < urbaine < 100000, urbaine > 100000, distance aux services publics)

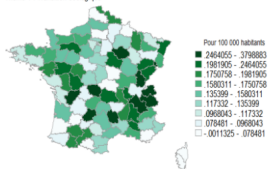
<https://granddebat.fr/media/default/0001/01/f73f9c2f64a8cf0b6efa24fdc80179e7426b8cc9.pdf>

Analyse de la participation

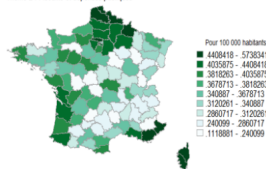


Nombre de contributions en ligne au Grand Débat National (volet Transition Écologique) en France métropolitaine 2019 Éric Chraïbi, et al.

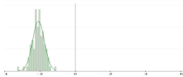
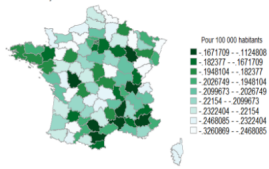
Thème 1 : Transition écologique



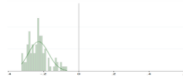
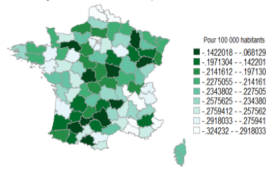
Thème 2 : Fiscalité et dépenses publiques



Thème 3 : Citoyenneté et démocratie



Thème 4 : Organisation de l'Etat et des services publics



SYNTHÈSES PAR QUESTION POSÉE

Synthèses de la Mission du GDN par thème par question

► **Fiscalité** : <https://granddebat.fr/media/default/0001/01/a95c01efe99bf13294aa6b574bf445a0925fb40b.pdf>

[//granddebat.fr/media/default/0001/01/a95c01efe99bf13294aa6b574bf445a0925fb40b.pdf](https://granddebat.fr/media/default/0001/01/a95c01efe99bf13294aa6b574bf445a0925fb40b.pdf)

► **Organisation de l'état et services publics** : <https://granddebat.fr/media/default/0001/01/6411a9e47f830b33c58e05acde5754f0aed9b631.pdf>

[//granddebat.fr/media/default/0001/01/6411a9e47f830b33c58e05acde5754f0aed9b631.pdf](https://granddebat.fr/media/default/0001/01/6411a9e47f830b33c58e05acde5754f0aed9b631.pdf)

► **Démocratie et citoyenneté** : <https://granddebat.fr/media/default/0001/01/7967bf7a5ea62fe6c284469196d9c829e26ac14a.pdf>

[//granddebat.fr/media/default/0001/01/7967bf7a5ea62fe6c284469196d9c829e26ac14a.pdf](https://granddebat.fr/media/default/0001/01/7967bf7a5ea62fe6c284469196d9c829e26ac14a.pdf)

► **Transition écologique** : <https://granddebat.fr/media/default/0001/01/b88758e8caa2733bec607a74b3b5371cc0a3b420.pdf>

[//granddebat.fr/media/default/0001/01/b88758e8caa2733bec607a74b3b5371cc0a3b420.pdf](https://granddebat.fr/media/default/0001/01/b88758e8caa2733bec607a74b3b5371cc0a3b420.pdf)



Le problème environnemental identifié comme le plus important

Question 1. Quel est aujourd'hui pour vous le problème concret le plus important dans le domaine de l'environnement ?

Une seule réponse possible



153 809
contributions

	% TOTAL	Type de communes				Distance aux services publics			
		Communes rurales	Communes issues d'unités urbaines de moins de 20.000 habitants	Communes issues d'unités urbaines de 20.000 à moins de 100.000 habitants	Communes issues d'unités urbaines de 100.000 habitants et plus	Très éloigné	Assez éloigné	Assez proche	Très proche
La pollution de l'air	20%	16%	18%	20%	22%	16%	18%	20%	22%
L'érosion du littoral	1%	1%	1%	2%	1%	1%	1%	1%	1%
Les dérèglements climatiques (crue, sécheresse)	29%	28%	29%	30%	29%	28%	29%	30%	29%
La biodiversité et la disparition de certaines espèces	22%	25%	23%	22%	21%	25%	24%	23%	21%
Autres	29%	31%	29%	28%	28%	31%	29%	28%	28%
Non réponses	7%	7%	6%	7%	6%	7%	6%	6%	7%



Les propositions pour répondre aux problèmes environnementaux identifiés

Question 1. Quel est aujourd'hui pour vous le problème concret le plus important dans le domaine de l'environnement ?

Autres citations (29% des réponses)

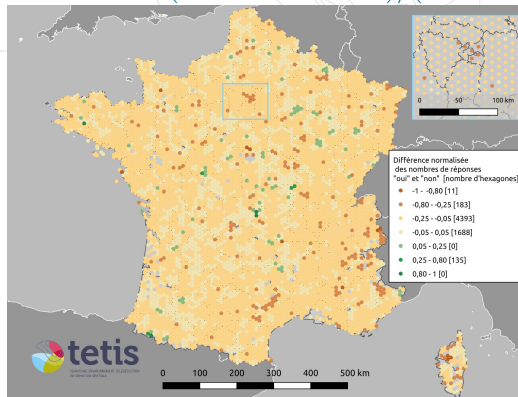
Question ouverte - Réponses non suggérées - Plusieurs réponses possibles



155 840 contributions

La pollution	6,4%
La pollution en général	4,8%
La pollution des sols	0,6%
La pollution émanant des transports	0,6%
L'environnement et l'écologie en général	2,8%
La dégradation de notre environnement en général	2,2%
L'écologie en général : tous les problèmes sont liés	0,5%
La gestion des déchets	2,4%
La gestion des déchets	2,2%
L'énergie et les ressources naturelles	2,2%
Le nucléaire, les déchets nucléaires	0,9%
La diminution des ressources naturelles causée par leur surexploitation	0,6%
La transition écologique au global doit être une priorité	0,4%
Les comportements et les actions	1,2%
Une agriculture irresponsable, intensive	0,7%
L'activité humaine non-maîtrisée, non-réglée	0,5%
Autres éléments	1,1%
L'impact sur la santé	0,9%
Les problèmes liés au climat	0,9%
Le climat	0,5%
Les émissions de gaz à effet de serre	0,4%
Les menaces sur la biodiversité	0,8%
La disparition des espèces	0,8%

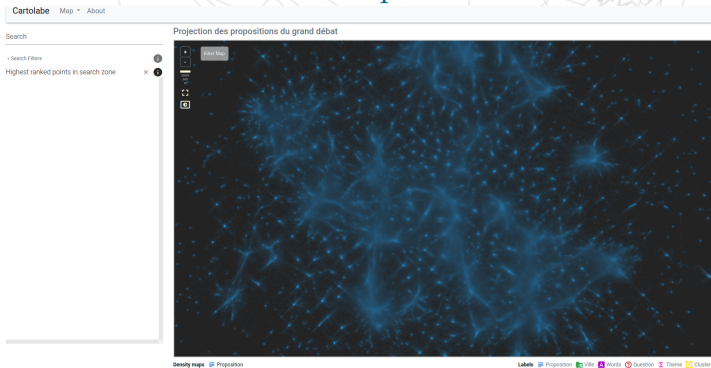
Question : Diriez-vous que votre vie quotidienne est aujourd'hui touchée par le changement climatique ? (Thème : Transition écologique. 134177 contributions, 131472 codes postaux exploitables) :
 64% : Oui, 29% : Non, 7% : Sans réponse.
 $Sentiment = (NbNon - NbOui) / (NbNom + NbOui)$



Le sentiment des contributeurs de l'impact du changement climatique sur leur vie quotidienne (contributions en ligne au Grand Débat National - volet Transition Écologique) 2019 [Éric Chraïbi et al.](#)

CLASSIFIER, CARTOGRAPHIER

Thèmes, Mots, Questions, Propositions, Moteur de recherche



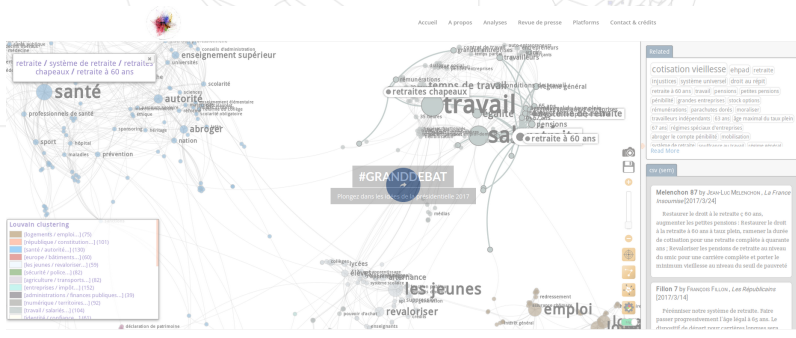
Developed by   

v2.4 © 2022 Cartolabe

[https:](https://cartolabe.fr/map/debat)

[//cartolabe.fr/map/debat](https://cartolabe.fr/map/debat)

CLASSIFIER, CARTOGRAPHIER



<https://politoscope.org/2019/01/granddebat/>

ÉMERGENCE DE THÈMES NON SOUMIS AU DÉBAT

- GDN santé [149178, $0,56 * 10^{-2}$] mental, prévention, pharmaceutique, honoraire, complémentaire, vieillesse, désert, nocif, généraliste, patient, dentaire, soignant, médecine, soin, médicament, tabac, infirmier, médecin, lunettes, couverture, dépendance, alcool, hôpital, optique, médical, éducation,
- EF santé [864, $0,82 * 10^{-3}$] soin, médecin, dangereux, médical, éducation, hôpital, domaine, secteur, sécurité, enseignement, professionnel, maison, justice, mauvais, environnement, coût, service, accès, manque
- VD santé [1773, $0,52 * 10^{-3}$] complémentaire, prévention, mutuel(le), éducation, environnement, alimentation, hôpital, professionnel, soin, sécurité, problème, assurance, établissement, maison, *ministère*, domaine, humain, bon, centre, maladie, *système*, protection, médecin, matière, meilleur, frais, *public*, dépense,

Liste des mots communs aux corpus des 3 plates-formes parmi les 200 mots les plus surreprésentés dans chacun des corpus par rapport au corpus de référence. Ces mots apparaissent dans chacun des 3 corpus GDN, VD et EF avec un rapport des fréquences normalisées supérieur à 26.

anti, plateforme, cyclable, isolation, défavorisé, pollueur, malus, assisté, drastiquement, optimisation, doublon, lambda, média, responsabiliser, migrant, TVA, taxer, réfugié, écologique, régalien, progressivité, taux, pesticide, polluant, obsolescence, raisonné, éolien, recyclage, stop, gaspillage, biodiversité, participatif, kérosène, taxe, impôt, citoyen, sécu, démuni, contraint, lobby, éolienne, niche, fraudeur, emballage, polluer, écologie.

GDN

incivilité [9,29]
 chaudière [5,12]
 bénévolat [3,98]
respect [2,52]
 habitation [2,38]
 isolation [2,37]
 dérèglement [2,22]
 régalien [2,13]
 incitatif [2,08]
associatif [2,07]
 bienveillant [2,04]
 incitation [1,990]
 piéton [1,95]
 solaire [1,93]
 sanctionner [1,90]
 entraide [1,90]
 collectivité [1,86]
 strate [1,86]
comptabiliser [1,82]
 commune [1,79]
 cyclable [1,79]
 fiable [1,77]
consultatif [1,76]
 allocation [1,75]
civique [1,75]
 foncier [1,74]
 pôle [1,74]
 pénaliser [1,72]
 territorial [1,70]

VD

OTAN [21,1]
Matignon [6,31]
Giscard [4,95]
 rémunération [4,68]
 conjoint [4,19]
 indexation [4,16]
 révocable [4,09]
 invalidité [3,70]
 dividende [3,68]
 instauration [3,44]
 pantouflage [3,32]
 détaché [3,32]
 autoroute [3,30]
 chaîne [3,28]
 rétablissement [3,19]
 additif [3,15]
 employeur [3,00]
Chirac [2,87]
PS [2,79]
député [2,78]
 rembourser [2,71]
 loyer [2,67]
uninominal [2,65]
 PMA [2,56]
 frais [2,54]
 réformer [2,44]
 prestation [2,41]
quorum [2,41]
 suppression [2,39]

EF

sexisme [9,64]
 lavable [9,37]
 végétarien [8,84]
discrimination [6,78]
 réutilisable [6,22]
racisme [5,87]
 migration [5,44]
 vrac [5,00]
faciès [4,47]
 réutiliser [4,38]
sexiste [4,30]
harcèlement [4,07]
 trier [4,02]
 alternative [3,84]
migrant [3,72]
migratoire [3,67]
 biodégradable [3,66]
 plastique [3,53]
 migrer [3,39]
réfugié [3,30]
 recycler [3,25]
 supermarché [3,24]
 sensibilisation [3,14]
 sensibiliser [3,08]
 emballage [3,06]
 paperasse [2,90]
 jetable [2,86]
 poubelle [2,82]
 ostentatoire [2,79]

Contextonymes par ordre alphabétique

▣ **transport** [\[définition\]](#) [↗](#) ;

▣ [bateau](#) [↗](#) ; ▣ [camion](#) [↗](#) ; ▣ [fer](#) [↗](#) ; ▣ [ferroutage](#) [↗](#) ; ▣ [fluvial](#) [↗](#) ; ▣ [fret](#) [↗](#) ; ▣ [kérosène](#) [↗](#) ; ▣ [longue](#) [↗](#) ; ▣ [marchandise](#) [↗](#) ; ▣ [maritime](#) [↗](#) ;
 ▣ [poids lourd](#) [↗](#) ; ▣ [rail](#) [↗](#) ; ▣ [routier](#) [↗](#) ; ▣ [taxation](#) [↗](#) ; ▣ [traverser](#) [↗](#) ;

▣ [aérien](#) [↗](#) ; ▣ [détriment](#) [↗](#) ; ▣ [développement](#) [↗](#) ; ▣ [distance](#) [↗](#) ; ▣ [doux](#) [↗](#) ; ▣ [express](#) [↗](#) ; ▣ [favoriser](#) [↗](#) ; ▣ [ferroviaire](#) [↗](#) ; ▣ [infrastructure](#) [↗](#) ;
 ▣ [mobilité](#) [↗](#) ; ▣ [moyen](#) [↗](#) ; ▣ [périurbain](#) [↗](#) ; ▣ **priviléger** [↗](#) ;

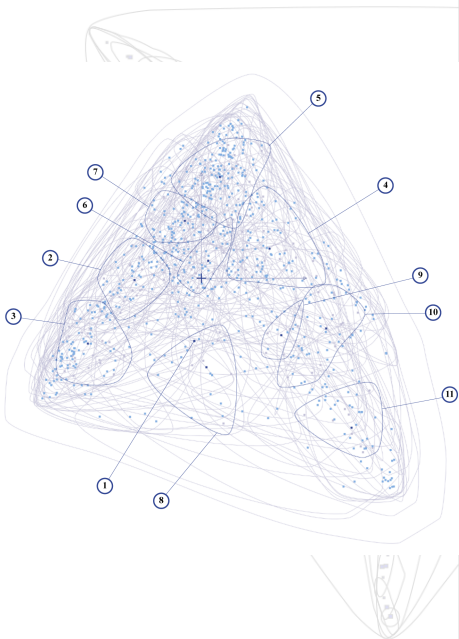
priviléger, favoriser, transport, doux
 privilégier, favoriser, transport, détriment
 privilégier, transport, doux, marche, commun, à_pied
 privilégier, transport, commun, marche, train
 privilégier, transport, marchandise, train, rail
 privilégier, favoriser, transport, marchandise, rail

▣ [à pied](#) [↗](#) ; ▣ [abonnement](#) [↗](#) ; ▣ [accès](#) [↗](#) ; ▣ [accessible](#) [↗](#) ; ▣ [adapté](#) [↗](#) ; ▣ [agglomération](#) [↗](#) ; ▣ [amélioration](#) [↗](#) ; ▣ [améliorer](#) [↗](#) ; ▣ [attractif](#) [↗](#) ;
 ▣ [banlieue](#) [↗](#) ; ▣ [billet](#) [↗](#) ; ▣ [bus](#) [↗](#) ; ▣ [carte](#) [↗](#) ; ▣ [centre](#) [↗](#) ; ▣ [collectif](#) [↗](#) ; ▣ [commun](#) [↗](#) ; ▣ [covoiturage](#) [↗](#) ; ▣ [cyclable](#) [↗](#) ; ▣ [demande](#) [↗](#) ;
 ▣ [desserte](#) [↗](#) ; ▣ [desservi](#) [↗](#) ; ▣ [développé](#) [↗](#) ; ▣ [développer](#) [↗](#) ; ▣ [domicile](#) [↗](#) ; ▣ [faciliter](#) [↗](#) ; ▣ [ferré](#) [↗](#) ; ▣ [fiable](#) [↗](#) ; ▣ [fréquence](#) [↗](#) ; ▣ [fréquent](#) [↗](#) ;
 ▣ [gare](#) [↗](#) ; ▣ [gratuit](#) [↗](#) ; ▣ [gratuité](#) [↗](#) ; ▣ [horaire](#) [↗](#) ; ▣ [ligne](#) [↗](#) ; ▣ [maillage](#) [↗](#) ; ▣ [marche](#) [↗](#) ; ▣ [métro](#) [↗](#) ; ▣ [métropole](#) [↗](#) ; ▣ [minute](#) [↗](#) ;
 ▣ [multiplier](#) [↗](#) ; ▣ [navette](#) [↗](#) ; ▣ [offre](#) [↗](#) ; ▣ [parisien](#) [↗](#) ; ▣ [parking](#) [↗](#) ; ▣ [partagé](#) [↗](#) ; ▣ [péage](#) [↗](#) ; ▣ [périphérie](#) [↗](#) ; ▣ [piéton](#) [↗](#) ; ▣ [piste](#) [↗](#) ;
 ▣ [province](#) [↗](#) ; ▣ [public](#) [↗](#) ; ▣ [rapide](#) [↗](#) ; ▣ [régional](#) [↗](#) ; ▣ [régulier](#) [↗](#) ; ▣ [relais](#) [↗](#) ; ▣ [RER](#) [↗](#) ; ▣ [réseau](#) [↗](#) ; ▣ [rural](#) [↗](#) ; ▣ [salariné](#) [↗](#) ; ▣ [sécurisé](#) [↗](#) ;
 ▣ [SNCF](#) [↗](#) ; ▣ [tarif](#) [↗](#) ; ▣ [ter](#) [↗](#) ; ▣ [train](#) [↗](#) ; ▣ [trajet](#) [↗](#) ; ▣ [tram](#) [↗](#) ; ▣ [tramway](#) [↗](#) ; ▣ [urbain](#) [↗](#) ; ▣ [usager](#) [↗](#) ; ▣ [vélo](#) [↗](#) ; ▣ [ville](#) [↗](#) ; ▣ [voie](#) [↗](#) ;

▣ [cyclable](#) [↗](#) ; ▣ [développer](#) [↗](#) ; ▣ [faciliter](#) [↗](#) ; ▣ [partagé](#) [↗](#) ; ▣ [piéton](#) [↗](#) ; ▣ [rapide](#) [↗](#) ; ▣ [salariné](#) [↗](#) ; ▣ [sécurisé](#) [↗](#) ; ▣ [urbain](#) [↗](#) ; ▣ [vélo](#) [↗](#) ;

▣ [à pied](#) [↗](#) ; ▣ [accès](#) [↗](#) ; ▣ [adapté](#) [↗](#) ; ▣ [agglomération](#) [↗](#) ; ▣ [amélioration](#) [↗](#) ; ▣ [centre](#) [↗](#) ; ▣ [collectif](#) [↗](#) ; ▣ [covoiturage](#) [↗](#) ; ▣ [développé](#) [↗](#) ;
 ▣ [domicile](#) [↗](#) ; ▣ [gratuit](#) [↗](#) ; ▣ [marche](#) [↗](#) ; ▣ [métropole](#) [↗](#) ; ▣ [multiplier](#) [↗](#) ; ▣ [parking](#) [↗](#) ; ▣ [périphérie](#) [↗](#) ; ▣ [piste](#) [↗](#) ; ▣ [public](#) [↗](#) ; ▣ [relais](#) [↗](#) ;
 ▣ [rural](#) [↗](#) ; ▣ [tramway](#) [↗](#) ; ▣ [usager](#) [↗](#) ; ▣ [ville](#) [↗](#) ; ▣ [voie](#) [↗](#) ;

▣ [abonnement](#) [↗](#) ; ▣ [accessible](#) [↗](#) ; ▣ [améliorer](#) [↗](#) ; ▣ [attractif](#) [↗](#) ; ▣ [banlieue](#) [↗](#) ; ▣ [billet](#) [↗](#) ; ▣ [bus](#) [↗](#) ; ▣ [carte](#) [↗](#) ; ▣ [commun](#) [↗](#) ; ▣ [demande](#) [↗](#) ;
 ▣ [desserte](#) [↗](#) ; ▣ [desservi](#) [↗](#) ; ▣ [ferré](#) [↗](#) ; ▣ [fiable](#) [↗](#) ; ▣ [fréquence](#) [↗](#) ; ▣ [fréquent](#) [↗](#) ; ▣ [gare](#) [↗](#) ; ▣ [gratuité](#) [↗](#) ; ▣ [horaire](#) [↗](#) ; ▣ [ligne](#) [↗](#) ;
 ▣ [maillage](#) [↗](#) ; ▣ [métro](#) [↗](#) ; ▣ [minute](#) [↗](#) ; ▣ [navette](#) [↗](#) ; ▣ [offre](#) [↗](#) ; ▣ [parisien](#) [↗](#) ; ▣ [péage](#) [↗](#) ; ▣ [province](#) [↗](#) ; ▣ [régional](#) [↗](#) ; ▣ [régulier](#) [↗](#) ;
 ▣ [RER](#) [↗](#) ; ▣ [réseau](#) [↗](#) ; ▣ [SNCF](#) [↗](#) ; ▣ [tarif](#) [↗](#) ; ▣ [ter](#) [↗](#) ; ▣ [train](#) [↗](#) ; ▣ [trajet](#) [↗](#) ; ▣ [tram](#) [↗](#) ;



1 chômeur

2 évasion, optimisation, fraude, taxation, fonds, groupe, effacement, fortement, mesure, concurrence, fraudeur...

3 GAFA, paradis, ISF, fiscal, bénéfice, multinational, fortune, niche, banque, capital, transaction, milliard, taxer, PME, rétablir, riche, français, CICE, partie, luxe, productif, imposé, résultat, actif, probant, concertation, biaisé

4 corruption, trafic, forces, guerre, ordre, arme, illégal, sanction, garantir, conflit, défense, prévention, pollution, crime, clandestin, passeur

5 chasse, biodiversité, espèce, interdiction, protection, protéger, soutenir, agir, soutien, climatique, priorité, plan, réchauffement, animal, pesticide, mener, sauvage, neuf, défendre, CO2, promouvoir, dérèglement, recours, urbain, humanité

6 lutter, lutte, abus, au noir, faux, gaspillage, se battre, légal, médical, désert

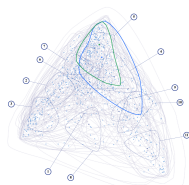
7 international, renforcer, européen, renforcement, contrôle, lobby, aérien, vente, commercial, renforcé, obsolescence, programmé...

8 inégalité, million, pauvre, injustice, chômage, voté, monter

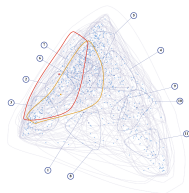
9 combattre, pauvreté, forme, dérive, extrême

10 racisme, discrimination, religieux, communautarisme, exclusion, haine, harcèlement, musulman, femme, blanc

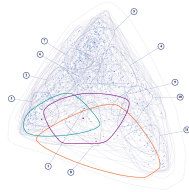
11 violence, anti, délinquance, drogue, délit, terrorisme, agression, incivilité, précarité, courant, manifestation, voter



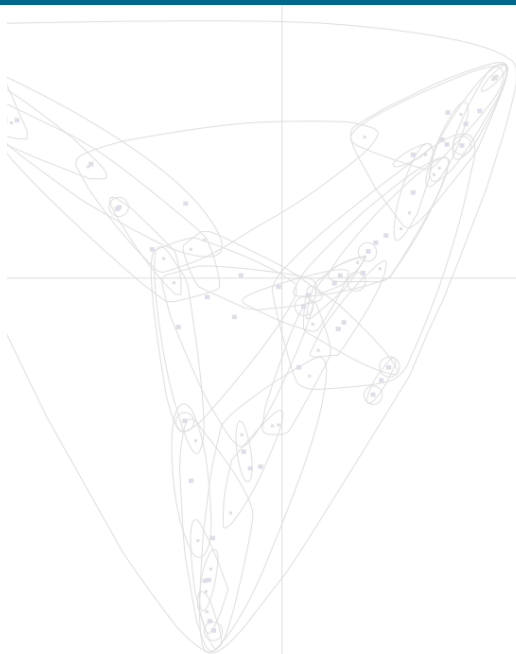
protection, protéger



mesure, effacement



injustice, pauvre, luxe



RELIER LA PAROLE À SON LIEU D'EXPRESSION



- ▶ Repérer des thèmes et des motifs par des modèles et méthodes en linguistique computationnelle
- ▶ Repérer leur emprise géographique
- ▶ Contraster ces thèmes et motifs
 - ▶ au sein des Cahiers (entre différentes aires géographiques...)
 - ▶ entre les différents corpus issus du GDN

LES CAHIERS CITOYENS, UN CORPUS GÉO-LOCALISÉ



Un corpus numérisé.
Cependant,

- ▶ des lexies non reconnues (de l'ordre de 46500 occurrences), des erreurs de transcriptions, des variantes orthographiques...
- ▶ un corpus non encore anonymisé

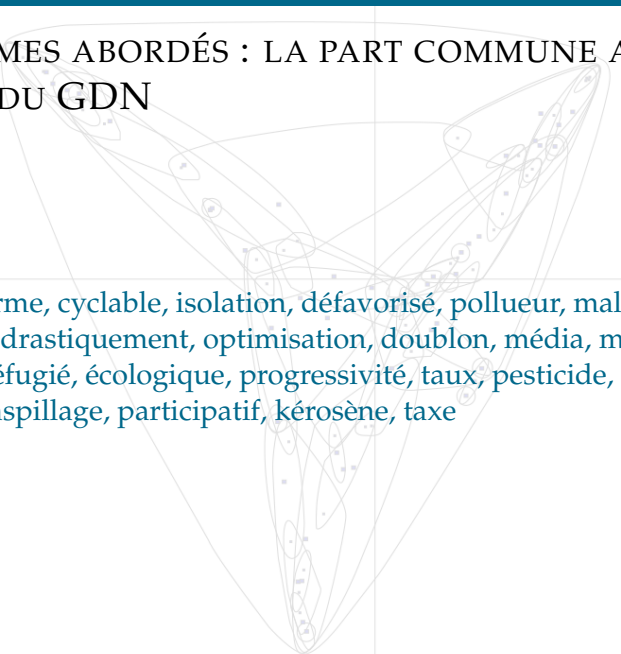
Un chantier préliminaire de consolidation des transcriptions et d'anonymisation est nécessaire

LES THÈMES ABORDÉS

Mots les plus surreprésentés par rapport à un corpus de référence

EDF, Élysée, Internet, ENA, millefeuille, assisté, nanti, indexation, défavorisé, réversion, gasoil, ONU, SNCF, relaxe, stop, UE, Bercy, HLM, PAC, isolation, retraité, OTAN, pollueur, ferroutage, optimisation, kérosène, fuel, taux, taxe, assistanat, taxer, plateforme, écologique, sénat, photovoltaïque, légalisation, doublon, ruralité, ralentisseur, votation, géothermie, drastiquement, RER, malus, réinvestir, indexer, fiscalisation, éolienne/éolien, cyclable, incivilité, décevement, participatif, carburant, migrant, désertification, dématérialisation

LES THÈMES ABORDÉS : LA PART COMMUNE AUX CORPUS DU GDN



plateforme, cyclable, isolation, défavorisé, pollueur, malus, assisté, drastiquement, optimisation, doublon, média, migrant, taxer, réfugié, écologique, progressivité, taux, pesticide, éolien, stop, gaspillage, participatif, kérosène, taxe

LES THÈMES ABORDÉS : LES TENDANCES PROPRES AUX CAHIERS

Mots sur-représentés dans les Cahiers vs. plateformes numériques

- ▶ **Format de lettre** : *votre, vous, nous, merci, souhaiter, exprimer, demander, comment, quel*
- ▶ **Revenus et difficultés** : *retraité, pension de retraite, veuve, actif, revalorisation, inflation, salaire, indexation*
- ▶ **Ruralité et carburant**
- ▶ **Élus** : *sénateur, maire, député*
- ▶ **Privilèges et injustices**

LES THÈMES ABORDÉS : LES TENDANCES PROPRES AUX CAHIERS

Mots sous-représentés dans les Cahiers vs. plateformes numériques

- ▶ Écologie des pratiques : *emballage, surconsommation, compostage, déchet, recycler, poubelle*
- ▶ Discrimination : *sexiste, discrimination, mixité*

LES THÈMES ABORDÉS : CONTRASTER LES AIRES GÉOGRAPHIQUES

- ▶ dans les communes (*pop.* > 100000) :
 - ▶ Le thème de la criminalité et le sentiment de crainte *criminel, relaxe, agression, non-lieu, traumatisme, crime, correctionnel, crainte*
 - ▶ Questions relatives à la ville : tram, propreté, trottoir, bibliothèque, immeuble, éducateur
- ▶ dans les communes (*pop.* < 30000) :
 - ▶ Le thème du déploiement de la fibre *fibre, déploiement, Internet, ADSL...*
 - ▶ Ruralité et commune : ruralité, village, clôture, assainissement, préemption, Regroupement pédagogique intercommunal (RPI) et remédiation
 - ▶ Entreprise : artisan, entrepreneur, liquidateur, SARL, endetter,
 - ▶ Sentiments exprimés : insoutenable, dramatique, immuable

LES RÉSEAUX LEXICAUX : CONTRASTER LES CORPUS DU GDN

Cahiers citoyens

Contre

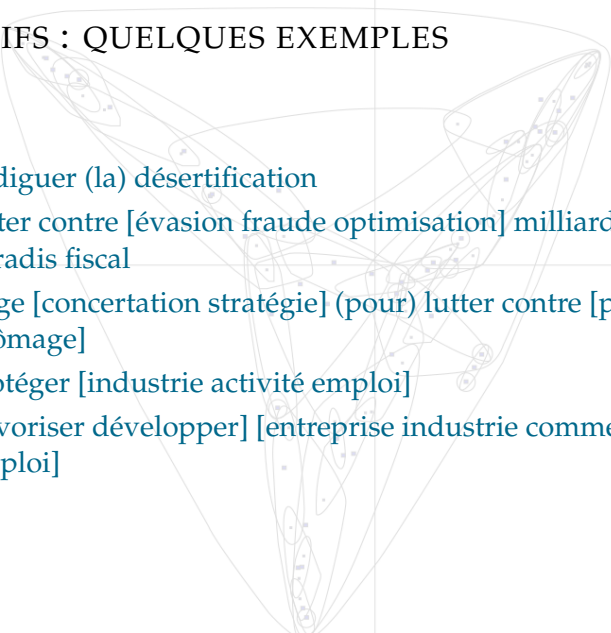
lutter/lutte (1) *évasion* (3) *fraude*(4)
réchauffement (5) *incivilité* (7) *crime* (8)
terrorisme (12) **désertification** (14) *pauvreté*
 (17) **désert** (18)...

Plateforme GDN

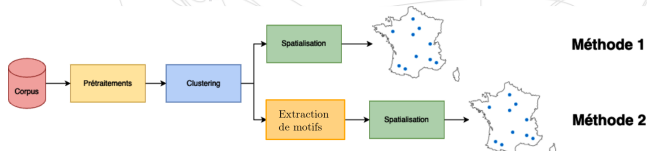
Contre

lutter/lutte (1) expérimentation (4) *évasion*
 (5) *fraude* (9) passeur (11) *terrorisme* (13)
crime (16) *réchauffement* (17) *pauvreté* (21)
 communautarisme (22) fake_news (24)
désertification (28) **désert** (41)...

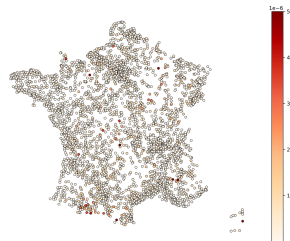
LES MOTIFS : QUELQUES EXEMPLES

- 
- ▶ endiguer (la) désertification
 - ▶ lutter contre [évasion fraude optimisation] milliard (vers) paradis fiscal
 - ▶ large [concertation stratégie] (pour) lutter contre [pauvreté chômage]
 - ▶ protéger [industrie activité emploi]
[favoriser développer] [entreprise industrie commerce emploi]
 - ▶ ...

EMPREINTE GÉOGRAPHIQUE DE CLUSTERS DE PHRASES ET DE MOTIFS



Rétablir l'ISF : [[rétablir,
ré-instaurer, remettre,
réintroduire...] {[rétablissement,
réintroduction...] de}] l'[ISF,
impôt (solidaire) sur la fortune]



Salomé Chandora (2023), Fouille sémantique et spatiale dans le corpus Cahiers Citoyens : comparaison de méthodes symbolique et numérique, Master 2, *Traitement automatique des langues*



PERSPECTIVES

- ▶ Compléter des analyses linguistiques plus ciblées : analyse des émotions, prise en compte de la modalité, de la forme (négative, affirmative)...
- ▶ Croiser des analyses textuelles à des données géographiques, sociologiques, démographiques, historiques : taille des communes, distance au centre-ville, éloignement du lieu de travail, isolement rural...
- ▶ Permettre l'interrogation du corpus
 - ▶ calculer l'emprise géographique d'un thème
 - ▶ contraster des aires géographiques
 - ▶ ...
- ▶ Être intégrée dans une plateforme d'interrogation, de valorisation et de restitution