



**HAL**  
open science

## Extraction and spatial contextualisation of transport proposals in the Great National Debate

Jacques Fize, Lucile Sautot, Martin Lentschat, Laurence Dujourdy, Ludovic Journaux, Mohamed Hilal

► **To cite this version:**

Jacques Fize, Lucile Sautot, Martin Lentschat, Laurence Dujourdy, Ludovic Journaux, et al.. Extraction and spatial contextualisation of transport proposals in the Great National Debate. *Revue Internationale de Géomatique*, 2022, 31 (3-4), pp.329-354. 10.3166/rig31.329-354 . hal-04210914

**HAL Id: hal-04210914**

**<https://institut-agro-dijon.hal.science/hal-04210914v1>**

Submitted on 20 Dec 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

---

# Extraction et mise en contexte spatial des propositions relatives au transport dans le Grand Débat National

Jacques Fize<sup>1</sup>, Lucile Sautot<sup>2</sup>, Martin Lentschat<sup>3</sup>,  
Laurence Dujourdy<sup>4</sup>, Ludovic Journaux<sup>5</sup>,  
Mohamed Hilal<sup>6</sup>

1. Laboratoire ERIC - Université Lumière Lyon 2 - Université de Lyon -  
jacques.fize@insa-lyon.fr

2. AgroParisTech - UMR TETIS -  
lucile.sautot@agroparistech.fr

3. Université Grenoble Alpes - LIG - SIGMA  
martin.lentschat@univ-grenoble-alpes.fr

4. L'Institut Agro Dijon, UBFC - Cellule d'appui à la recherche en sciences  
des données -  
laurence.dujourdy@agrosupdijon.fr

5. L'Institut Agro Dijon, UBFC - Lab. Info. Bourgogne EA 7534-  
ludovic.journaux@agrosupdijon.fr

6. INRAE, L'Institut Agro Dijon, UBFC - UMR CESAER -  
mohamed.hilal@inrae.fr

---

RÉSUMÉ. Le Grand Débat National, lancé début 2019 par Emmanuel Macron, Président de la république, début 2019 pour répondre au mouvement social des Gilets Jaunes, a permis de collecter les contributions de citoyens sur la transition écologique via une plateforme en ligne. Dans cet article, nous exploitons le corpus constitué par ces contributions pour identifier des zones où les participants demandent le développement de pistes cyclables et d'équipements ferroviaires. Pour cela, nous avons créé un modèle de classification permettant d'identifier les contributions traitant de la thématique du transport et proposé une méthode d'extraction de contributions traduisant les propositions des contributeurs. A l'aide des codes postaux donnés par les contributeurs, nous avons cherché à expliquer la fréquence spatiale des demandes, en lien avec leur contexte résidentiel (urbain, périurbain ou rural) et l'accessibilité des commerces et services.

ABSTRACT. The Great National Debate, launched by Emmanuel Macron in early 2019

to respond to the "Gilets Jaunes" social movement, allowed the collection of citizens' contributions on the ecological transition via an online platform. In this article, we use the corpus constituted by these contributions to identify locations where participants are asking for the development of bicycle paths and railway facilities. For this purpose, we have created a classification model to identify answers related to the theme of transportation and proposed a method for extracting contributions that reflect the contributors' proposals. We then sought to explain the observed spatial frequency of requests in relation to information describing the area of residence (urban, peri-urban, rural) and accessibility to shops and services.

MOTS-CLÉS : Grand Débat National, opinion-mining, transport

KEYWORDS: Grand Débat National, opinion-mining, transport

---

DOI:10.3166/RIG.XX.1-?? © 2023 Lavoisier

## 1. Introduction

En octobre 2018, une hausse du prix du carburant automobile a déclenché en France un mouvement spontané de protestations sociales dénommé "mouvement des gilets jaunes". Le mouvement s'exprime par des blocages de routes, de ronds-points et de nombreuses manifestations. La première manifestation réunit 280 000 personnes le 17 novembre 2018 (acte 1) et surprend par son ampleur (Boyer *et al.*, 2020). Elle sera suivie par de multiples formes de mobilisations accompagnées de revendications sociales et politiques. En décembre 2018, face à ce mouvement, le gouvernement français saisit la Commission nationale du débat public<sup>1</sup> (CNDP) afin de lancer une consultation citoyenne : le Grand Débat National. Cette consultation s'est tenue du 15 janvier au 15 mars 2019 autour des quatre grands thèmes suivants : transition écologique, fiscalité, démocratie et citoyenneté, organisation de l'État et des services publics. Différentes modalités de participations sont proposées : réunions publiques locales, cahiers de doléances en mairie, conférences avec des citoyens tirés au sort, contributions en lignes sur une plateforme dédiée, etc.

L'analyse et la synthèse des contributions en ligne proposent plusieurs défis. Tout d'abord, elles sont nombreuses. Sur la plateforme en ligne, pour le questionnaire de la Transition Écologique, plus de 134 000 personnes ont rempli 12 champs de texte libre, soit un total de 1,6 millions de textes potentiels<sup>2</sup> à analyser. Les textes recueillis sont de différentes longueurs et présentent une grande diversité (style, syntaxe, orthographe, grammaire, lexique). À la demande du gouvernement français, une première analyse des contributions en ligne est ef-

---

1. <https://www.debatpublic.fr/>

2. Les contributeurs ne répondent pas nécessairement à l'ensemble des questions.

fectuée par l'entreprise OpinionWay<sup>3</sup> immédiatement après la clôture du débat (Opinion Way, 2019). Les synthèses proposées offrent une vue d'ensemble des contributions, mais présentent plusieurs limitations. Premièrement, la méthodologie utilisée n'est pas complètement documentée (Bellet *et al.*, 2021), ce qui rend difficile sa reproductibilité et la comparaison avec d'autres travaux. Deuxièmement, selon les questions analysées, un nombre important de réponses (entre 10 % et 30 %) est écarté dans les synthèses OpinionWay<sup>4</sup>. Enfin, la dimension territoriale du mouvement social à l'origine de la consultation qu'aux thématiques de la consultation a été très peu abordée.

Des travaux scientifiques ont été publiés sur l'analyse de ces contributions. Parmi ces communications, Ploux *et al.* (2021) ont proposé une analyse automatique sur la base de motifs et d'une comparaison du lexique employé dans différentes consultations, incluant celles du Grand Débat National. Messer *et al.* (2019) ont également proposé une approche qualitative des contributions relatives à la mobilité. Ce rapport, très riche sur la description des propositions formulées par les contributeurs, ne porte pas sur la dimension spatiale.

Dans nos travaux, nous souhaitons analyser quel est l'apport de ces contributions à la thématique des mobilités qui est à la croisée d'enjeux écologiques et d'inclusion sociale. En effet, les habitants d'un territoire ont des contraintes, obligations et besoins quotidiens qui ne se déroulent pas à leur adresse ou, voire même, dans leur commune de résidence. Pour travailler, s'approvisionner en produits domestiques divers et variés, se soigner, s'instruire, se cultiver, se distraire, etc. ils sont dans l'obligation de se rendre dans différents lieux. La majorité des déplacements quotidiens s'exercent à l'intérieur d'un espace, structuré à partir des mobilités courantes. Les individus privilégient des déplacements courts et proches de leurs lieux de résidence pour l'accès aux commerces et services. Les mobilités peuvent être plus étendues pour rejoindre un lieu de travail ou des équipements et services non disponibles à proximité du lieu de résidence. La fréquence et les modalités des déplacements diffèrent selon les territoires (Brutel, Pages (2021)). Les distances parcourues quotidiennement sont plus courtes dans les centres urbains, avec une utilisation importante des transports en commun. Les distances sont plus longues dans les espaces périurbains et ruraux, avec une forte dépendance à l'automobile et une très faible utilisation des transports en commun et des modes actifs comme la marche ou le vélo.

Les mobilités ont un impact direct sur le climat et la qualité de vie (pollution, bruit, etc.). Elles constituent un enjeu important pour les pouvoirs publics

---

3. Les synthèses OpinionWay sont disponibles ici : <https://granddebat.fr/pages/syntheses-du-grand-debat>

4. La documentation d'OpinionWay mentionne dans les résultats un pourcentage "Autres contributions trop peu citées ou inclassables" pour les questions ouvertes sans que la méthodologie fournie décrive précisément comment une contribution est assignée à cette classe. Voir (Opinion Way, 2019)

afin d'accélérer la transition écologique. Les acteurs territoriaux (collectivités, opérateurs publics ou privés) tentent d'imaginer des solutions de mobilités durables et inclusives pour diminuer les émissions de gaz à effet de serre, tout en facilitant la mobilité quotidienne dans les territoires urbains, périurbains et ruraux. Parmi les solutions, proposant des alternatives à l'automobile, comme les transports en commun, l'autopartage, les points multimodaux, la logistique urbaine, etc., la desserte ferroviaire et le développement des pistes cyclables ont attiré notre attention.

Dans cet article, nous présentons une analyse des réponses associées au transport dans les contributions de la partie sur la transition écologique. Nous proposons une approche permettant d'identifier et cartographier les opinions des contributeurs sur cette thématique. Le processus proposé se divise en plusieurs étapes : (i) identification des réponses associées aux transports à l'aide d'un modèle de classification, (ii) extraction de contributions dans les réponses identifiées, puis (iii) la cartographie des contributions extraites à l'aide des méta-données des contributeurs. Enfin, nous proposons une analyse comparative de la présence de contributions avec des variables socio-économiques. Les données mises à disposition dans le cadre du grand débat ne contenant pas de description socio-économique des contributeurs, nous nous appuyons sur les données socio-économiques des communes pour formuler des hypothèses sur les contributeurs exprimant les opinions extraites.

Cet article s'organise de la manière suivante. La section 2 présente les données et méthodes utilisées pour classer les réponses et extraire les propositions des contributeurs, ainsi que pour les mettre en relation avec leur contexte résidentiel. La section 3 présente les résultats obtenus, c'est-à-dire les propositions extraites selon leurs occurrences et leur localisation, et une première caractérisation des espaces dans lesquelles ces propositions ont été formulées. Enfin, nous discutons en section 4 des limitations des résultats obtenus, avant de conclure en section 5.

## 2. Matériel et méthodes

Dans ces travaux, nous réalisons une analyse des opinions sur la question du transport dans les réponses de la *transition écologique*. Ce processus s'effectue en deux étapes : la première (*c.f.* Section 2.2) identifie les réponses mentionnant le transport ; la seconde (*c.f.* Section 2.3) en extrait les contributions en s'appuyant sur des règles d'extraction et plusieurs lexiques. Dans la Section 2.4, nous présentons les informations contextuels associées (et manquantes) pour chaque contributeur dans les données. Le processus est illustré dans la Figure 1. Le code source pour reproduire les méthodes décrites dans l'article est accessible à l'adresse suivante : <https://github.com/Make-the-Debat-Great-Again/RIG>

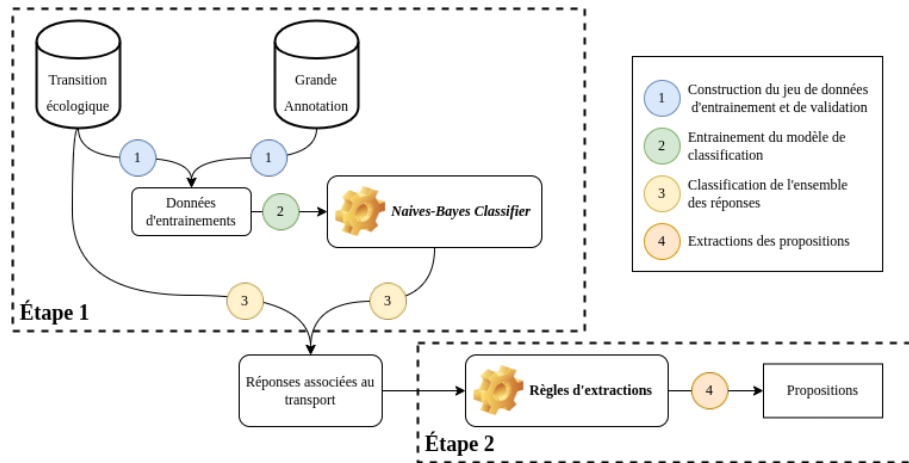


FIGURE 1 – Processus d'extraction des motifs d'opinions

### 2.1. Données

Dans ce travail, nous utilisons les réponses récoltées sur la plateforme web du Grand Débat National<sup>5</sup>. La plateforme proposait deux modalités de contribution : libre ou questionnaire à choix multiples. Seules les contributions libres sont exploitées dans cet article. Plus particulièrement, nous avons uniquement retenu les contributions au volet portant sur la transition écologique.

Concernant l'analyse du contexte résidentiel, nous exploitons le code postal indiqué par chaque contributeur, sachant que 97 % de ces codes sont exploitables. Les 3% inexploitable correspondent à des codes postaux incomplets (e.g. 34). Le dictionnaire, qui permet de faire le lien avec les informations communales et leurs coordonnées géographiques (latitude, longitude), est extrait de la base officielle des codes postaux<sup>6</sup>.

Les réponses relatives à la demande de transport sont géographiquement contextualisées à l'aide de données tirées d'autres sources. Depuis OpenStreet-Map, nous récupérons le tracé des pistes cyclables, des lignes ferroviaires et la position des gares ferroviaires. Les codes postaux sont utilisés pour décrire le contexte résidentiel des contributeurs. Outre la densité de population, nous mobilisons deux nomenclatures statistiques spatiales qui permettent de saisir le degré d'urbanisation des lieux de résidence et la position de ces lieux dans la hiérarchie urbaine. La base des aires urbaines de l'INSEE<sup>7</sup> permet de distinguer les répondants selon qu'ils résident dans un pôle urbain, une commune

5. <https://granddebat.fr/pages/donnees-ouvertes>

6. <https://www.data.gouv.fr/fr/datasets/base-officielle-des-codes-postaux/>

7. <https://www.insee.fr/fr/information/2115011>

périurbaine, une petite ou moyenne aire urbaine ou une commune rurale située à l'écart des influences urbaines. Les commerces et services publics et privés sont indispensables pour répondre aux besoins des habitants d'un territoire. En complément du logement et de l'emploi, ils constituent une ressource primordiale du quotidien et offrent la possibilité de s'approvisionner en produits divers et variés, se soigner, s'instruire, se cultiver, se distraire, etc. Cette offre n'est pas homogène sur le territoire : les services, commerces et équipement ne sont pas de même nature et n'ont pas la même concentration spatiale. INRAE<sup>8</sup> et l'ANCT<sup>9</sup> identifient quatre niveaux de centres d'équipements et de services (Hilal *et al.*, 2020) : les centres locaux, de premier niveau, offrent une douzaine de commerces et services du quotidien, les centres intermédiaires proposent une vingtaine d'équipement et de services complémentaires (santé, éducation, supermarché, etc.), les centres structurants ont une offre plus élargie (lycée, équipements culturels et sportifs, etc.) et, enfin, les centres majeurs offrent un gamme complète de commerces et services. Dans notre article, nous retenons comme indicateurs les temps de trajet nécessaires pour se rendre dans les centres les plus proches de son domicile pour les quatre niveaux de centralité. La Table 1 l'ensemble des données utilisées et leur source.

TABLE 1 – Liste des indicateurs de contexte spatial

Indicateurs	Source
Présence de pistes cyclables	OpenStreetMap
Présence d'une gare de voyageurs	OpenStreetMap
Présence d'une ligne ferroviaire	OpenStreetMap
Code postal	La Poste
Densité de population en 2017	INSEE
Catégorie de la commune dans le zonage en aires urbaines	INSEE
Temps de trajet pour accéder au centre majeur le plus proche (minutes)	INRAE, ANCT
Temps de trajet pour accéder au centre structurant le plus proche (minutes)	INRAE, ANCT
Temps de trajet pour accéder au centre intermédiaire le plus proche (minutes)	INRAE, ANCT
Temps de trajet pour accéder au centre local le plus proche (minutes)	INRAE, ANCT

## 2.2. Classification des réponses liées aux transports

Dans ce travail, nous nous concentrons sur la thématique du transport dans le cadre de la transition écologique. Pour identifier les réponses associées au

8. Institut national de recherche pour l'agriculture, l'alimentation et l'environnement

9. Agence nationale de la cohésion des territoires

transport, nous entraînons un modèle de classification à l'aide d'apprentissage automatique. Pour cela, nous isolons les réponses liées à ce thème en utilisant les annotations réalisées, parallèlement à l'appel à contribution du grand débat, par le collectif nommé Grande Annotation<sup>10</sup>. Les réponses du Grand Débat ont été annotées à l'aide d'une liste de *tags*. Chaque réponse à une question peut être associée avec plusieurs tags. Au moment du téléchargement des données (31 mai 2020), 280 321 réponses étaient annotées.

Pour entraîner ce modèle, nous avons construit un jeu de données composé de deux types de réponses, l'une évoquant le transport et l'autre non. L'identification des réponses annotées traitant de transport s'opère en sélectionnant celles ayant un des *tags* faisant partie de la liste présentée dans la Table 2. La Table 6 indique la distribution des réponses associées pour chacun des tags choisis

Plusieurs algorithmes de classification sont évalués. Parmi les algorithmes existants, nous avons utilisé : Support Vector Machine (SVM) (Vapnik, 2013); classifieurs Naïf-Bayésien (Manning *et al.*, 2010); K plus proches voisins (Altman, 1992); les Arbres de décisions (Loh, 2011) et Random Forest (Breiman, 2001). De plus, nous ajoutons des approches d'*apprentissage profond* : Bi-LSTM et BERT (Devlin *et al.*, 2019).

TABLE 2 – Tags de la Grande Annotation associée à la thématique du transport

Transports doux ou moins nombreux	Pistes cyclables, sécurisées	Commerce de proximité, circuit court
Déménagement ou changement travail	Obligation, incitation au transport collectif	Transport commun gratuit
Télétravail	Transport commun plus fiable, rapide	Covoiturage, autopartage, stop
Pistes cyclables	Transport commun moins cher	Transport commun plus sûr
Transports en commun meilleurs	Télétravail, horaires flexibles	Parkings (vélo, auto) à proximité des gares
Transports en commun moins chers	Transport commun plus proche	Développer le réseau ferroviaire
Transport à la demande	Transport commun plus fréquent	Ramassage scolaire
Vélo, trottinette (électrique)	Voiture électrique	Transports en commun (bus, tram)
Voiture plus propre	Marche à pied	Voiture à hydrogène
Voiture autonome	Train	Voiture hybride
A Déplaçons-nous moins	Transport fluvial Biocarburant (E85, etc)	

10. <https://grandeannotation.fr>



Les données sont représentées en *bag of words* (sac de mots)(Manning *et al.*, 2010) excepté pour les méthodes d'apprentissage profond. Une représentation *bag of words* numérise un texte sous forme de vecteur  $v = \{v_0, \dots, v_N\}$  de dimension  $N$ , où  $N$  correspond au nombre de mots uniques présents dans le corpus. Ici, chaque valeur  $v_i$  correspond au nombre d'occurrences d'un mot pondéré par la fréquence inverse de document (ou TF-IDF)(Salton, McGill, 1983). Plus un mot sera présent dans l'ensemble du corpus de texte, plus il sera pénalisé. Avant d'être encodés, les textes sont lemmatisés. La lemmatisation remplace chaque mot par sa forme canonique, son lemme, par exemple, "aimons" sera remplacé par "aimer". Ce traitement permet de diminuer la taille du vocabulaire sur lequel s'appuie l'encodage du texte et donc facilite la comparaison des contributions.

### 2.3. *Extraction de contributions associés à une opinion*

Dans ces travaux, nous nous appuyons sur la définition d'une opinion de (Liu, Zhang, 2012). Une opinion  $o_i$  est définie selon le quintuplet  $\{o_e, o_p, o_c, o_a, o_d\}$ ;  $o_i \in O$ , où  $O$  correspond à l'ensemble des opinions présentes dans une unité de texte (phrase, paragraphe, etc.),  $o_e$  correspond à l'émetteur de l'opinion,  $o_p$  est le polarisateur,  $o_c$  la cible de l'opinion,  $o_a$  est l'attribut de la cible de l'opinion et  $o_d$  est la date à laquelle l'opinion est émise. Par exemple, dans la phrase, "*Je trouve que le prix du billet de train est hors de prix.*", l'opinion  $o_1$  est {"Je", "hors de prix", "billet de train", "prix",  $\emptyset$  }. Dans ces travaux, nous ajoutons à ce quintuplet la localisation géographique de l'émetteur,  $o_l$ .

Dans ces travaux, nous nous appuyons sur une forme particulière de contributions : **la proposition**. Reprenant le quintuplet de (Liu, Zhang, 2012), l'émetteur  $o_e$  correspond à l'auteur de la contribution. Le polarisateur  $o_p$  est porté par un verbe (modifier, développer, augmenter, etc.) ou adverbe (plus, moins, etc.) de changement. La cible de l'opinion  $o_c$  correspond à un objet appartenant à la catégorie de transport (moyen de locomotion, infrastructure, etc.). Si celle-ci peut être absente, l'attribut  $o_a$  de l'objet cible  $o_c$  est caractérisé par un nom (*e.g.* prix, fonctionnement, rapidité, etc.). La *date d'émission*  $o_d$  correspond à la date de publication sur la plateforme. La *localisation*  $o_l$  est indiquée par le code postal donné par le contributeur.

Afin de traiter de la thématique du transport, les objets cibles  $o_c$ , les attributs  $o_a$  mais aussi les polarisateurs  $o_p$  sont répertoriés dans une ontologie construite sur mesure et illustrée dans la Figure 12 (en annexe). Cette ontologie<sup>11</sup> est composée de termes liés aux infrastructures, moyens de locomotion, entreprises de transport, les caractéristiques, etc.

---

11. L'ontologie est accessible au format OWL à cette adresse : <https://doi.org/10.5281/zenodo.7774988>.

L'ontologie est composée de 283 concepts recouvrant 370 termes de vocabulaire, chaque concepts possédant un label privilégié (skos:prefLabel) et, le cas échéant, un ou plusieurs synonymes exactes et acronymes (skos:altLabel) dudit concept. Ces concepts sont structurés par 294 relations de subsumption (i.e. relation subClassOf). Certain concepts possèdent également des relations décrivant leurs propriétés, comme le carburant consommé par un véhicule (e.g. un train roule à l'électricité ou aux hydrocarbures) ou l'utilisation des infrastructures (e.g. une gare ferroviaire accueil des trains). L'utilisation d'une ontologie nous offre ici une base de vocabulaire structurée permettant de reconnaître les termes spécifiques liées aux transports, via l'identification des PrefLabel et Alt-Label. Le plein potentiel d'une ontologie s'exprime dans les propriétés y étant décrites, celles-ci ne sont pas exploitées pour le moment mais permettraient à de futures études de recouper les propositions portant sur un même sujet général (e.g. bus et train sont tous deux des transports collectifs) ou de poser des questions plus spécifiques (e.g. quelle est la polarité des contributions mentionnant un véhicule électrique?).

Pour identifier les contributions de type proposition, nous utilisons un ensemble de règle d'extractions s'appuyant sur l'arbre de dépendances syntaxiques dans une phrase. Par exemple, la phrase "Il faut augmenter le nombre de bus en ville." donne l'arbre illustré dans la Figure 2. Précisément, ces règles s'appuient sur l'existence d'un sous-graphe formé par des dépendances sélectionnées. Les règles définies sont les suivantes :

- Adverbe  $\xrightarrow{advmod}$  Groupe nominale  
*ex. Plus de train*
- Verbe  $\xrightarrow{obj}$  Groupe nominale  
*ex. Augmenter les bus dans ma ville*
- Verbe  $\xrightarrow{xcomp}$  Verbe  $\xrightarrow{obj}$  Groupe nominale  
*ex. Il faut augmenter le nombre de train*
- Groupe nominale  $\xleftarrow{nsubj}$  Verbe  $\xrightarrow{xcomp}$  Verbe  $\xrightarrow{obj}$  Groupe nominale  
*ex. Les entreprises doivent faire des efforts*

Pour faire cette extraction sur ces sous-graphes, nous utilisons la bibliothèque (library) Spacy<sup>12</sup> et son module Dependency Matcher. L'ensemble des relations syntaxiques disponibles sont définies par l'*Universal Dependency Relation*<sup>13</sup>. Pour définir, les relations syntaxiques utilisées, nous utiliserons les notations A et B pour définir la source et la destination de la relation. La relation *advmod* correspond à une modification du mot B par l'adverbe A. La relation *obj* indique l'objet B visé par le verbe A. La relation *nsubj* indique le

12. <https://spacy.io/>

13. <https://universaldependencies.org/u/dep/>

sujet B permettant de conjuguer le verbe A. La relation *xcomp* indique que B est un complément du verbe A.

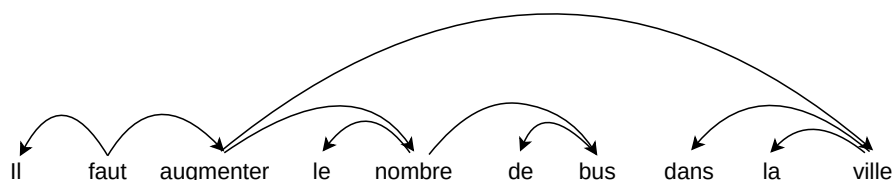


FIGURE 2 – Arbre de dépendances syntaxiques pour la phrase "Il faut augmenter le nombre de bus en ville."

Dans la section suivante, nous présentons les résultats de l'extraction de contributions obtenus en s'appuyant sur notre méthode et les données utilisées.

#### 2.4. Contexte résidentiel des contributeurs

Les données ne contiennent aucune information décrivant les caractéristiques individuelles (âge, genre, diplôme, statut matrimonial, etc.), sociales (profession ou catégorie socioprofessionnelle, etc.) et économiques (niveau de revenu, etc.). De fait, il est impossible de caractériser individuellement les contributeurs porteurs des opinions d'intérêts comme la demande de développement des pistes cyclables ou d'équipements ferroviaires. La seule information disponible étant le code postal de la commune de résidence, nous mobilisons à titre exploratoire quelques informations relatives aux espaces de résidence de ces contributeurs.

Réalisée à l'échelle des codes postaux, l'analyse sera supracommunale dans le cas des contributeurs habitants hors des grandes villes, où chaque zone est constituée de plusieurs communes. À l'inverse, dans les très grandes métropoles, découpées en plusieurs zones postales, nous utiliserons le maillage communal, car nous ne disposons pas nécessairement de données infracommunales. Lors d'un pré-traitement, données communales sont agrégées à l'échelle des zones postales<sup>14</sup>.

Les contributions sont ensuite liées aux données décrivant le contexte résidentiel via une jointure portant sur les codes postaux. Les contributeurs n'ayant pas fourni de code postal ou ayant fourni un code postal incomplet sont exclus des analyses.

Après cette mise en lien, nous dénombrons par zone postale le nombre totale de propositions ainsi que le nombre de celles qui concernent un développement des équipements ferroviaires et des pistes cyclables. Quelques statistiques

14. Cette opération est décrite et réalisée avec le code R disponible ici : <https://doi.org/10.5281/zenodo.6683741>

descriptives statistiques descriptives sont ensuite produites pour comparer la fréquence d'apparition des demandes.

### 3. Résultats

#### 3.1. Identification des réponses associées au transport

Dans la Table 3, nous présentons les valeurs d'exactitudes (*accuracy*) obtenues par les différents algorithmes de classifications sélectionnés. L'exactitude est calculée entre les labels prédits  $\hat{y}$  et attendus  $y$  par la formule suivante :

$$Exactitude(y, \hat{y}) = \frac{1}{|y|} \sum_{i=1}^{|\hat{y}|} x = \begin{cases} 1 & y_i == \hat{y}_i \\ 0 & sinon \end{cases}$$

Si les valeurs d'exactitude sont proches, ce sont les algorithmes NAIVE-BAYES et BERT qui présentent les scores les plus élevés. Les modèles comme BERT étant très chers en ressources et temps de calcul, les résultats des prochaines étapes s'appuieront sur les sorties du modèle entraîné avec l'algorithme NAIVE-BAYES. Dans la Table 4, nous indiquons le pourcentage de réponses à une question traitant du transport. Nous observons que les pourcentages les plus élevés se trouvent dans les réponses aux questions indiquant clairement la notion de transport, par exemple, "*Si non, quelles sont les solutions de mobilités alternatives que vous souhaiteriez pouvoir utiliser ?*". Toutefois, nous observons un pourcentage relativement élevé sur la question générale : "*Que faudrait-il faire selon vous pour apporter des réponses à ce problème ?*"

TABLE 3 – Résultat de classification selon différents algorithmes

Modèle	Paramètres	Exactitude
SVM	g=0.1 & kernel=rbf	0.86
NAIVE-BAYES	alpha = 1	<b>0.87</b>
SGD	loss=log & penalty = L1	0.86
KNN	n = 5 & p =2	0.76
DECISION TREE	criterion = gini	0.83
RANDOM FOREST	g=100 & kernel=rbf	0.84
Bi-LSTM	32 LSTM cells	0.69
BERT	BERT multilingual base model (cased)	<b>0.87</b>

#### 3.2. Répartition des propositions identifiées

La distribution des propositions détectées selon le polarisateur (le verbe utilisé) et l'objet ciblé sont illustrées dans la Figure 3. Dans les propositions les plus fréquentes, nous observons une forte demande de développement des infrastructures de transports publics et leur accès sur le plan financier. De plus, il

TABLE 4 – Pourcentage de réponses associées au transport par question

Question	Pourcentage
Si non, quelles sont les solutions de mobilité alternatives que vous souhaiteriez pouvoir utiliser ?	59,51
Si oui, que faudrait-il faire pour vous convaincre ou vous aider à utiliser ces solutions alternatives ?	34,08
Qu'est-ce qui pourrait vous inciter à changer vos comportements comme par exemple mieux entretenir et régler votre chauffage, modifier votre manière de conduire ou renoncer à prendre votre véhicule pour de très petites distances ?	28,79
Si oui, que faites-vous aujourd'hui pour protéger l'environnement et/ou que pourriez-vous faire ?	27,71
Quelles seraient pour vous les solutions les plus simples et les plus supportables sur un plan financier pour vous inciter à changer vos comportements ?	17,54
Que faudrait-il faire selon vous pour apporter des réponses à ce problème ?	10,36
Et qui doit selon vous se charger de vous proposer ce type de solutions alternatives ?	9,30
Y a-t-il d'autres points sur la transition écologique sur lesquels vous souhaiteriez vous exprimer ?	5,40
Que pourrait faire la France pour faire partager ses choix en matière d'environnement au niveau européen et international ?	3,75
Si oui, que faudrait-il faire pour vous convaincre ou vous aider à changer de mode de chauffage ?	1,42
Si oui, de quelle manière votre vie quotidienne est-elle touchée par le changement climatique ?	1,05
Quel est aujourd'hui pour vous le problème concret le plus important dans le domaine de l'environnement ?	0,59

est intéressant de constater que la gratuité des transports est mentionnée avec le verbe *rendre*. La deuxième tendance centrale repose sur le pouvoir d'achat par une demande de la baisse des prix des moyens de transport ainsi que de la limitation de la consommation de certains produits. À l'inverse du développement des transports publics, nous voyons des propositions proposant la diminution de l'utilisation des voitures à moteur thermique.

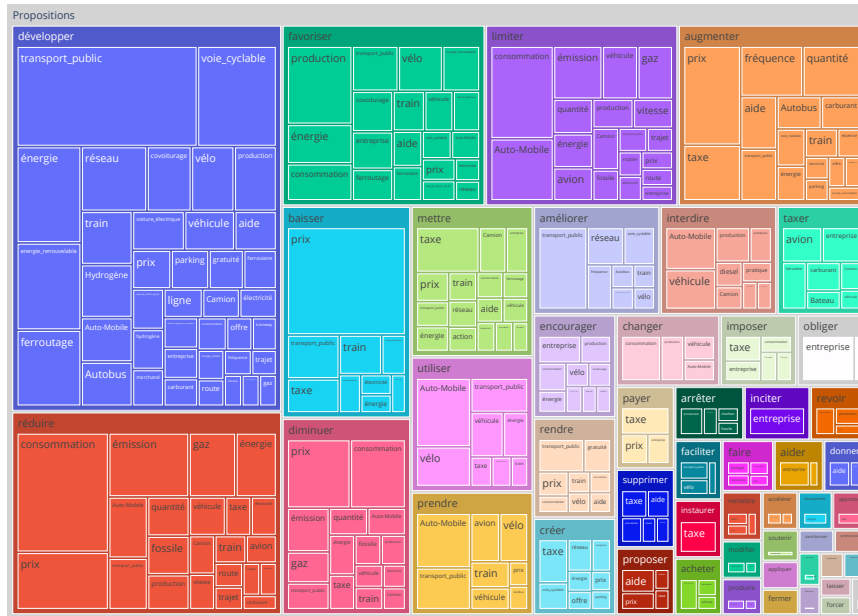


FIGURE 3 – Résumé des propositions identifiées dans le corpus de réponse sur le thème de la transition écologique

### 3.3. Répartition spatiale des contributions : le cas du développement des trains et des pistes cyclables

En France, le nombre de personnes habitant en ville a fortement augmenté depuis les années 1950. Aujourd’hui, c’est le cas de plus de 80 % de Français. Cette urbanisation s’est accompagnée de l’étalement des villes sur les campagnes et de la création d’aires urbaines formées d’un pôle urbain et de couronnes périurbaines. Le développement des espaces périurbains s’accompagne d’une multiplication des constructions et d’un remplacement des terres agricoles par des lotissements de maisons individuelles et petites résidences. La multiplication des voies rapides, reliant les espaces périurbains au centre-ville et la construction de vastes zones commerciales, transforment les paysages. Très intense autour des grands pôles urbains, ce phénomène d’étalement se retrouve également autour des villes moins peuplées tandis que de très nombreuses communes rurales restent à l’écart de ce mouvement.

Cette distinction se retrouve grâce au découpage des aires urbaines de l’INSEE, à partir duquel nous avons constitué quatre catégories d’espace : les pôles urbains (des grandes aires urbaines), le périurbain (de ces aires), les petites ou

moyennes aires urbaines, enfin les espaces hors influence des pôles. Chaque catégorie correspond à un contexte résidentiel offrant un cadre et des conditions de vie spécifiques, plutôt urbains pour les deux premières catégories et plutôt rurales pour les deux dernières.

TABLE 5 – Répartition des contributeurs par catégories d'espace

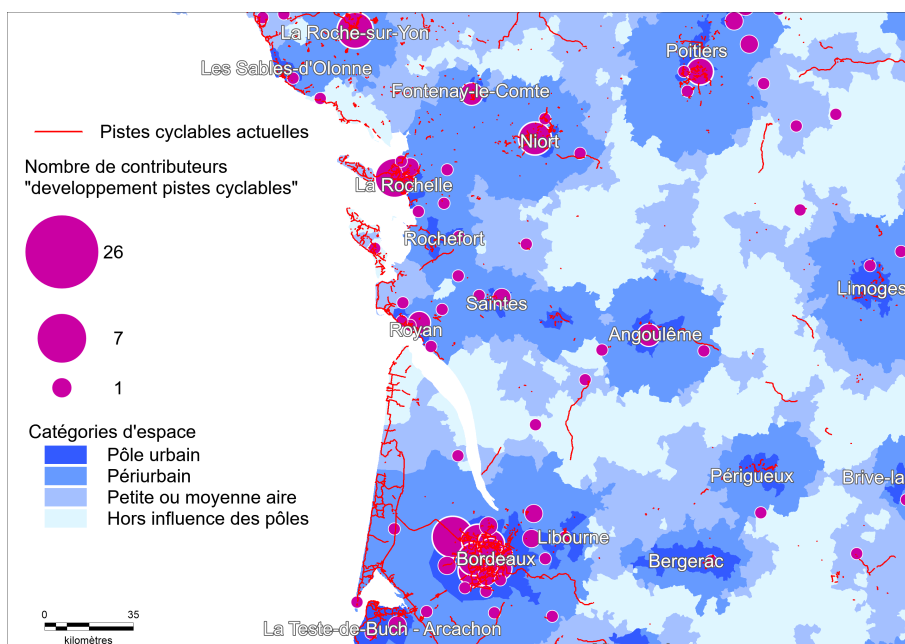
Catégorie	Contributeurs	dont train # (%)	dont vélo # (%)	Somme # (%)
Pôles urbains	60 658	2 803 (4.6)	1 613 (2.7)	4416 (7.2)
Périurbain	25 295	1 243 (4.9)	481 (1.9)	4416 (6.8)
Petites ou moyennes aires	13 496	654 (4.8)	201 (1.5)	855 (6.3)
Hors influence des pôles	4 641	219 (4.7)	46 (1.0)	265 (5.7)

Sources: INSEE, GDN, La Poste

La table 5 dénombre le nombre total de contributeurs ainsi que le nombre et la part de ceux demandant un développement du réseau ferroviaire et des pistes cyclables. Les contributeurs ayant fait cette demande sont peu nombreux : moins de 5 000 pour le train et moins de 2 500 pour les pistes cyclables. Les demandes de train se retrouvent chez 4.6 à 4.9 % des contributeurs. Les écarts observés entre catégories d'espace sont statistiquement significatifs (test de Kruskal-Wallis, p-value < 2.2e-16) : les habitants du périurbain et des petites et moyennes aires sont proportionnellement plus nombreux à demander le développement du train. La demande des populations rurales (hors influence des pôles) est supérieure à celle des habitants des pôles urbains. La plus forte demande, constatée pour les contributeurs résidant dans le périurbain, les petites et moyennes aires ou le rural, est à mettre en relation avec une plus faible desserte ferroviaire de ces espaces et le recours quasi généralisé à l'automobile pour effectuer les déplacements. Les demandes de pistes cyclables sont bien moins nombreuses. Elles sont plus fréquentes chez les contributeurs résidant dans les pôles urbains et moins fréquentes dans le rural. Entre ces deux catégories, la fréquence des demandes augmente progressivement en passant des contributeurs des petites et moyennes aires à ceux du périurbain.

Les Figures 4 et 5 permettent de localiser le nombre de contributeurs demandant un développement, respectivement, des pistes cyclables et du train. Sur le fond sont figurés nos quatre catégories d'espace, avec en supplément le tracé des infrastructures recensées par OpenStreetMap (pistes cyclables, voies ferrées et gares ferroviaires). Cette article est une extension d'une contribution faites pour la conférence SAGEO. L'évènement s'étant déroulé à La Rochelle, nous avons choisi, à titre d'exemple, de présenter les résultats obtenus sur le centre-ouest de la France métropolitaine. Dans le département de Charente-Maritime, par exemple, des contributeurs demandant le développement de pistes cyclables sont présents dans les aires urbaines de La Rochelle,

Saintes et Saint Jean d'Angély, ainsi que dans les communes de l'axe Rochefort/Saintes/Cognac. Concernant les demandes de développement du réseau ferroviaire et/ou une augmentation de la fréquence des trains, les besoins sont exprimés au centre des aires urbaines de Sainte et de La Rochelle. On identifie également une demande dans les communes isolées au nord de Saint Jean d'Angély, entre les gares de Saint Jean d'Angély et de Villeneuve-la-Comtesse.

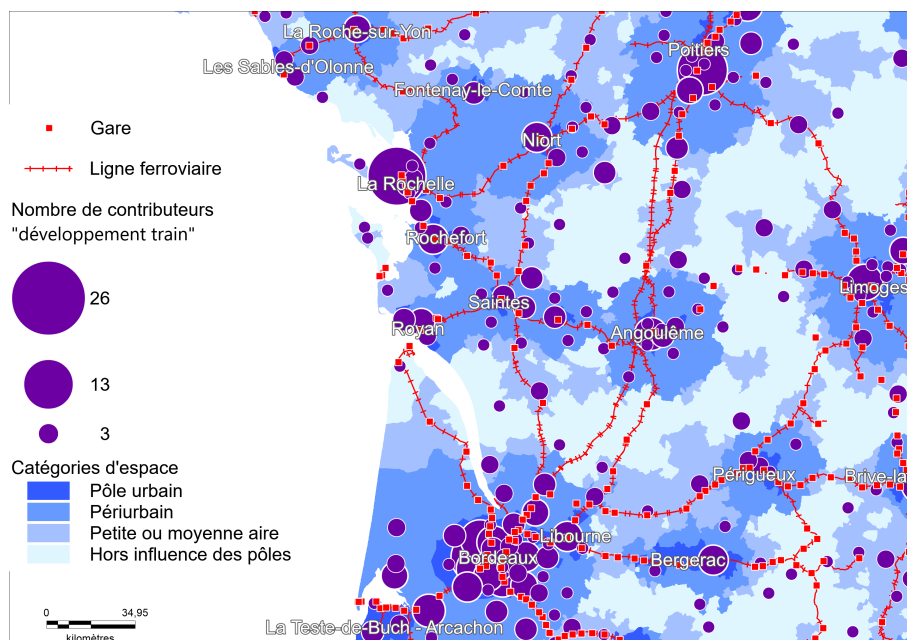


Sources: IGN, INSEE, OSM, GDN, La Poste

FIGURE 4 – Contributeurs demandant un développement des pistes cyclables en *Charente-Maritime* et en *Charente*

Le tableau et les cartes produites tendent à montrer que les fréquences associées aux demandes de train et de pistes cyclables se différencient selon les catégories d'espace, mais également, sur les cartes, avec une répartition spécifique des points. Afin de conforter cette analyse de la répartition de la demande de ces deux équipements visant à améliorer les déplacements du quotidien (accès aux commerces, établissements scolaires, services publics, services médicaux, sanitaires ou sociaux, etc.), nous avons choisi cinq indicateurs supplémentaires de caractérisation du contexte résidentiel des contributeurs. Le premier, la densité de population, va nous permettre d'explorer la variation de la demande entre les espaces les plus denses et les moins denses. Les quatre suivants mesurent l'accessibilité des commerces et services présents dans les communes de résidence des contributeurs. Plutôt que de dérouler les équipements l'un après l'autre, nous exploitons les données de temps de trajet (Hilal *et al.*,



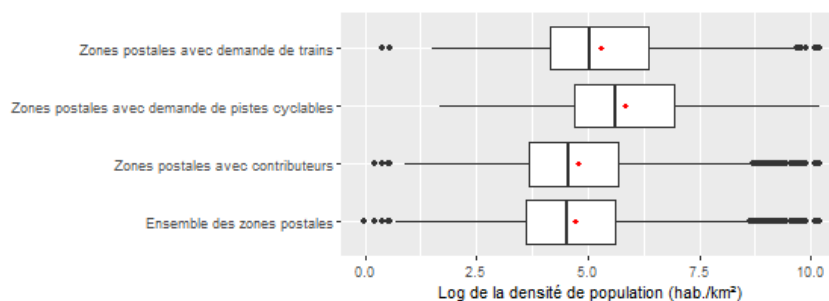


Sources: IGN, INSEE, OSM, GDN, La Poste

FIGURE 5 – Contributeurs demandant un développement du réseau ferroviaire et/ou une augmentation de la fréquence des trains en Charente et Charente Maritime

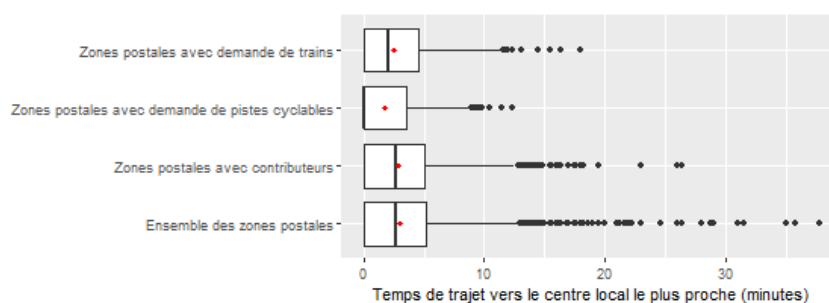
2018) pour rejoindre un des 10 774 centres d'équipement et de services (Hilal *et al.*, 2020), classés en quatre niveaux dénommés centres locaux, intermédiaires, structurants et majeurs<sup>15</sup>. Les centres locaux se caractérisent par la présence de services et équipements de « proximité » comme, par exemple, un salon de coiffure, une boulangerie, une école élémentaire, un ou une infirmière, etc. Plus diversifiée, l'offre des centres intermédiaires vient y ajouter plus de professionnels de santé, de services aux particuliers, de commerces, une école maternelle et un collège. Dans les centres structurants, une quarantaine de commerces et services viennent compléter la palette des équipements des centres intermédiaires. On y trouve des équipements spécialisés dans le commerce et les services aux particuliers, des services liés à la santé et à la médecine de ville, le lycée et quelques équipements de sports, loisirs et culture. Enfin, outre les commerces et services précédents, les centres majeurs offrent la quasi-totalité des services de santé, des commerces de la grande distribution ou alimentaires spécialisés, plusieurs services aux particuliers supplémentaires par rapport aux autres niveaux, notamment des services de l'administration judiciaire. Les centres majeurs se

15. <https://www.observatoire-des-territoires.gouv.fr/niveau-de-centres-dequipements-et-de-services-des-communes>



Sources: INSEE, GDN, La Poste

FIGURE 6 – Répartition des contributeurs selon la densité de population (échelle log) de leur zone postale de résidence

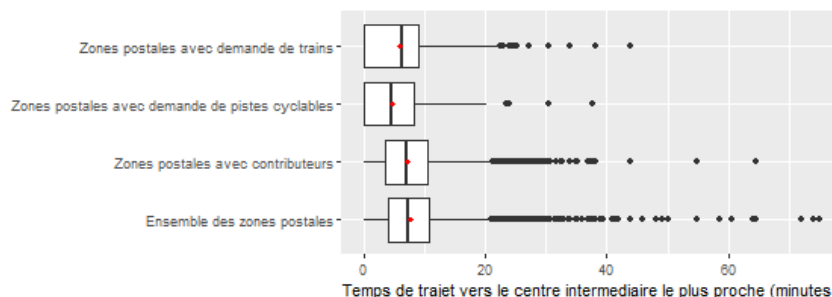


Sources: INSEE, INRAE/ANCT, GDN, La Poste

FIGURE 7 – Répartition des contributeurs selon le temps de trajet pour rejoindre le centre local le plus proche

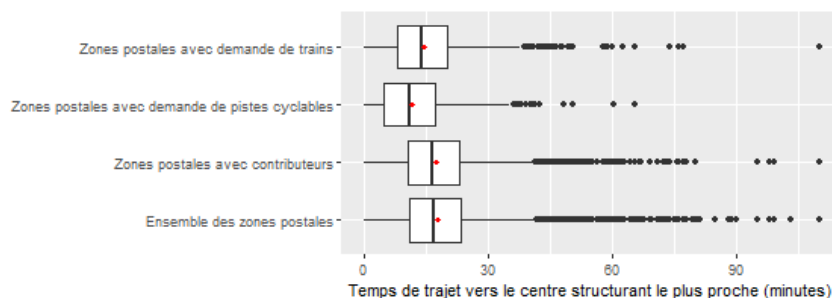
caractérisent également par des services et équipements généralement présents dans les plus grandes villes comme des services de police, de l'administration fiscale, des établissements scolaires et universitaires, des équipements de sports, loisirs et culture.

L'effet de la densité de population semble très net comme le montre la Figure 6. En moyenne, la demande de développement de pistes cyclables est plus fréquentes dans les espaces les plus denses. L'accessibilité des centres d'équipement et de services semblent également avoir un effet puisque les contributeurs résidant à proximité d'un centre expriment plus souvent une demande relative au développement de pistes cyclables et de façon un peu moins marquée du train (Figures 7 à 10).



Sources: INSEE, INRAE/ANCT, GDN, La Poste

FIGURE 8 – Répartition des contributeurs selon le temps de trajet pour rejoindre le centre intermédiaire le plus proche



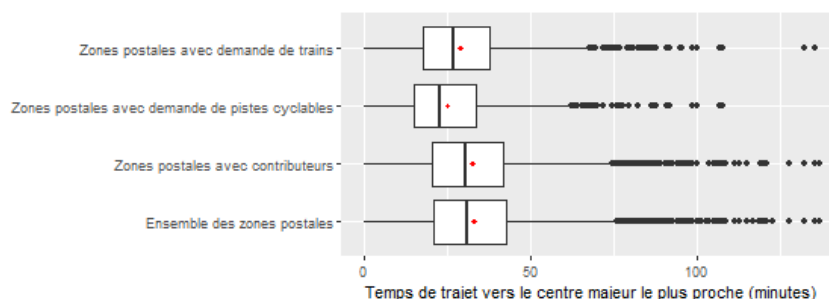
Sources: INSEE, INRAE/ANCT, GDN, La Poste

FIGURE 9 – Répartition des contributeurs selon le temps de trajet pour rejoindre le centre structurant le plus proche

## 4. Discussions

### 4.1. Représentativité des contributeurs

Les contributions du Grand Débat National ne sont pas distribuées aléatoirement dans l'espace. Leur nombre varie selon les caractéristiques économiques et sociales des départements avec, en particulier, un effet significatif du niveau de vie médian et du taux de diplôme sur la participation globale (Bennani *et al.*, 2020). De plus, nos résultats reflètent l'opinion des contributeurs, mais ne peuvent en aucun cas être considérés comme un reflet fidèle de l'opinion de la population. La méthode d'enquête de ce débat le rend plus proche de la production participative (crowdsourcing) ou du réseau social, que du sondage. La différence entre les statistiques classiques et l'analyse des réseaux sociaux (et plus généralement l'analyse de contenu généré par les utilisateurs) est bien



Sources: INSEE, INRAE/ANCT, GDN, La Poste

FIGURE 10 – Répartition des contributeurs selon le temps de trajet pour rejoindre le centre majeur le plus proche

connue, mais son importance a néanmoins été reconnue même en présence de différents types de biais (Morstatter, Liu, 2017). Ainsi, les caractéristiques sociologiques des contributeurs (âge, sexe, catégorie socio-professionnelle, etc) n'ont pas été demandées et sont absentes dans le jeu de données. Il n'est donc pas possible de s'assurer que les contributeurs constituent un échantillon représentatif de la population. Seul le code postal, renseigné par le contributeur, indique la répartition spatiale des contributeurs. La participation n'a pas été homogène sur le territoire national (Bennani *et al.*, 2020). On constate par exemple une sous-représentation des grands pôles urbains (58,4 % de la population contre seulement 27 % des contributeurs) et une surreprésentation du périurbain (24,6 % contre 47,6 %) et des autres communes (17 % contre 25,5 %). De plus, il faut noter que les cartes qui ont été présentées ne montrent pas le nombre de contributions dans lesquelles les motifs ont été détectés.

#### 4.2. Couverture des motifs détectés dans les contributions

Plusieurs limitations importantes du travail proposé dans cet article se rapportent à la couverture de la méthode de détection de motifs. Premièrement, nous nous appuyons sur un échantillon de règles restreint. Deuxièmement, ces règles s'appuient sur l'extraction des relations syntaxiques, elles-mêmes issues d'un modèle prédictif. Une des grandes limitations de ce modèle repose sur les données utilisées pour l'apprentissage issue de source non-bruité comme Wikipedia, documents médicaux ou encore venant de la presse Candito, Seddah (2012). Dans le cadre du Grand Débat, ce sont des textes plus bruts (fautes d'orthographe, oublis de ponctuation, etc.), ce qui peut perturber la prédiction du modèle. Une des perspectives de ce travail est de produire un modèle automatique en s'appuyant sur des annotations expertes qui indiquent les propositions et leur composantes.

### **4.3. *Biais associés aux modes de transport étudiés***

Dans cet article, nous avons étudié spécifiquement les contributions concernant les pistes cyclables et les équipements ferroviaires. Or, comme le montre l'étude de Brutel, Pages (2021), le mode de transport dépend du type d'espace habité et de l'activité de l'utilisateur. Ainsi, le vélo est plus pratiqué dans les pôles urbains que dans les espaces ruraux. Pour avoir une couverture plus importante des types d'utilisateurs et des types d'usages, il faudrait comparer la répartition spatiale des opinions concernant différents modes de transports, en prenant notamment en compte le transport routier. Un complément extrêmement intéressant aux cartes produites dans cet article serait une analyse similaire sur les prix des autoroutes et des carburants.

Néanmoins, nous disposons de peu de contributions (toutes thématiques confondues) hors des pôles urbains. On pourrait penser que les problèmes spécifiques aux territoires ruraux risquent donc d'être très peu présentes au sein du jeu de données, mais il semble que ce ne soit pas le cas, certains travaux ayant spécifiquement traité cette question (Flipo (2021)).

### **4.4. *Utilisation des contributions au Grand Débat National au niveau opérationnel***

Au delà de l'intérêt scientifique qui peut être porté aux contributions au Grand Débat National, nous pensons que les collectivités territoriales, à un échelon supra communal, peuvent utiliser ces données. Il semblerait que des tentatives aient été proposées (Messer *et al.* (2019) pour la ville de Paris, ou Flipo (2021) pour la Drôme et l'Ardèche) mais quelles sont en nombre limité. En dépit des nombreux biais décrits dans cet article et plus généralement dans la littérature dont est affectée cette consultation, le nombre de participants, la diversité des thématiques abordées et l'utilisation systématique de réponses en texte libre rendent ce jeu de données très riche d'information concernant l'opinion des contributeurs sur l'état des services de transports actuels mais également sur les services futurs désirés ou désirables. Il serait intéressant d'étudier dans quelle mesure les connaissances extraites de ces contributions peuvent ensuite être réutilisées par les collectivités territoriales, notamment sur la thématique des transports et de la mobilité.

## **5. Conclusion**

Dans cet article, nous présentons une approche d'extraction de propositions sur les contributions en ligne au Grand Débat National et leur mise en contexte spatial. Les propositions extraites, couplées à des données publiques, permettent d'identifier des espaces dans lesquels les contributeurs demandent le développement d'équipements et d'infrastructures. Les résultats montrent que ces demandes se différencient spatialement sur le gradient urbain-périurbain-

rural et selon l'accessibilité des centres d'équipements et de services. Ces premiers résultats sont encourageants et montrent le potentiel des contributions au Grand Débat National comme une source d'information pour les acteurs publics et privés (collectivités territoriales, opérateurs, etc.) en vue d'améliorer le cadre et les conditions de vie des populations.

Néanmoins, en raison de la non-représentativité et du faible nombre de contributeurs à une échelle locale, les conclusions tirées de ces données doivent être prises avec prudence : si les cartes produites permettent de localiser des demandes ponctuelles d'équipements, la proportion de la population formulant cette demande reste très faible.

De plus, l'absence d'informations individuelles sur les contributeurs et leur localisation approximative ne permettent pas de produire des analyses sur des populations spécifiques. L'utilisation de données décrivant le contexte géographique (densité de population, zonage en aires urbaines, temps d'accès aux centres d'équipements et de service) permet d'obtenir des analyses globales pertinentes.

Les futurs travaux porteront en premier lieu sur l'amélioration de l'extraction d'opinions, à l'aide de modèle automatique entraîné à l'aide d'algorithme d'apprentissage supervisé. En second lieu, l'analyse territoriale peut être poursuivie, en mettant en lien de nouvelles variables contextuelles, mais également en utilisant des extractions de propositions portant sur d'autres sujets (équipements routiers par exemple). Enfin, la géovisualisation des résultats constitue un troisième axe de développement.

## Bibliographie

- Altman N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, vol. 46, n° 3, p. 175–185. Consulté sur <https://www.tandfonline.com/doi/pdf/10.1080/00031305.1992.10475879>
- Bellet A., Denis P., Gilleron R., Keller M., Vauquier N. (2021). Pour plus de transparence dans l'analyse automatique des consultations ouvertes : leçons de la synthèse du Grand Débat National. *Statistique et Société*, vol. 9, n° 1-2, p. 147-168. Consulté sur <https://hal.inria.fr/hal-02860659>
- Bennani H., Gandré P., Monnery B. (2020). Les déterminants locaux de la participation numérique au grand débat national: une analyse économétrique. *Revue économique*, vol. 71, n° 4, p. 715–738. Consulté sur <https://www.jstor.org/stable/26956222>
- Boyer P. C., Delemotte T., Gauthier G., Rollet V., Schmutz B. (2020). Les déterminants de la mobilisation des Gilets jaunes. *Revue économique*, vol. 71, n° 1, p. 109–138. Consulté sur <https://doi.org/10.3917/reco.711.0109>
- Breiman L. (2001). Random forests. *Machine learning*, vol. 45, n° 1, p. 5–32. Consulté sur <https://link.springer.com/content/pdf/10.1023/A:1010933404324.pdf>
- Brutel C., Pages J. (2021). La voiture reste majoritaire pour les déplacements domicile-travail, même pour de courtes distances. *Insee Première*, n° 1835. Consulté sur <https://www.insee.fr/fr/statistiques/5013868>
- Candito M., Seddah D. (2012, juin). Le corpus sequoia : annotation syntaxique et exploitation pour l'adaptation d'analyseur par pont lexical (the sequoia corpus : Syntactic annotation and use for a parser lexical domain adaptation method) [in French]. In *Proceedings of the joint conference jep-taln-recital 2012, volume 2: Taln*, p. 321–334. Grenoble, France, ATALA/AFCP. Consulté sur <https://aclanthology.org/F12-2024>
- Devlin J., Chang M.-W., Lee K., Toutanova K. (2019). *Bert: Pre-training of deep bidirectional transformers for language understanding*. Consulté sur <https://arxiv.org/pdf/1810.04805.pdf>
- Flipo A. (2021). Que pensent les habitants des territoires peu denses de la mobilité ? une exploitation textuelle des données du grand débat national. *Flux*, vol. 1, n° 123, p. 54-73. Consulté sur <https://doi.org/10.3917/flux1.123.0054>
- Hilal M., Fugerey-Scarbel A., Gojard S. (2018). *Odomatrix : Mesurer des temps de trajet pour améliorer les conditions de vie*. Research Report. INRA. Consulté sur <https://hal.archives-ouvertes.fr/hal-02068383>
- Hilal M., Moret D., Piguët V., Bouscasse H., Cavailles J., Diallo A. *et al.* (2020, juin). *Centralités : comment les identifier et quels rôles dans les dynamiques locales et intercommunales ?*. Rapport technique. Agence nationale de la cohésion des territoires. Consulté sur <https://hal.archives-ouvertes.fr/hal-03064655>
- Liu B., Zhang L. (2012). A survey of opinion mining and sentiment analysis. In C. C. Aggarwal, C. Zhai (Eds.), *Mining text data*, p. 415–463. Boston, MA, Springer US. Consulté sur [https://doi.org/10.1007/978-1-4614-3223-4\\_13](https://doi.org/10.1007/978-1-4614-3223-4_13)

- Loh W.-Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, n° 1, p. 14–23. Consulté sur <https://doi.org/10.1002/widm.8>
- Manning C., Raghavan P., Schütze H. (2010). Introduction to information retrieval. *Natural Language Engineering*, vol. 16, n° 1, p. 100–103. Consulté sur <https://link.springer.com/article/10.1007/s10791-009-9115-y>
- Messer M. A., Utz S., Moreau D. (2019). *Analyse exploratoire des contributions au grand débat national, volet mobilité*. techreport. Lausanne (Suisse), Forum Vies Mobiles, Paris (France). Consulté sur [https://fr.forumviesmobiles.org/sites/default/files/editor/granddebat\\_rapport-final.pdf](https://fr.forumviesmobiles.org/sites/default/files/editor/granddebat_rapport-final.pdf)
- Morstatter F., Liu H. (2017, 1<sup>er</sup> 6). Discovering, assessing, and mitigating data bias in social media. *Online Social Networks and Media*, vol. 1, p. 1–13. Consulté sur <https://doi.org/10.1016/j.osnem.2017.01.001>
- Opinion Way. (2019). *Traitement des données issues du grand débat national*. Rapport technique. Paris, France : Auteur. Consulté sur <https://granddebat.fr/pages/syntheses-du-grand-debat>
- Ploux S., Genay M., Ploux-Chillès L. (2021, décembre). Les mots du Grand Débat national : les réseaux lexicaux des contributions déposées sur trois plateformes. *Humanités numériques*, n° 4. Consulté sur <https://journals.openedition.org/revuehn/2655>
- Salton G., McGill M. (1983). *Introduction to modern information retrieval*. New York, McGraw-Hill.
- Vapnik V. (2013). *The nature of statistical learning theory*. Berlin, Springer science & business media.



**Annexe A. La grande Annotation**

Categorie	% (total = 6459 )
Transports doux ou moins nombreux	22.51
Marche à pied	14.77
Voiture électrique	6.87
Transports en commun meilleurs	6.72
Vélo, trottinette (électrique)	5.91
Télétravail	5.14
Train	4.32
Pistes cyclables, sécurisées	3.93
Transport commun plus fréquent	3.28
Transport commun gratuit	3.28
Pistes cyclables	3.20
Transport commun plus proche	2.62
Transport commun plus fiable, rapide	2.46
Covoiturage, auto-partage, stop	2.45
Transports en commun moins chers	2.09
Transports en commun (bus, tram)	1.95
Transport commun moins cher	1.66
Voiture plus propre	1.41
Voiture à hydrogène	0.81
Commerce de proximité, circuit-court	0.67
Développer le réseau ferroviaire	0.56
Parkings (vélo, auto) à proximité des gares	0.51
Voiture hybride	0.45
Voiture autonome	0.43
A Déplaçons-nous moins	0.39
Télétravail, horaires flexibles	0.31
Transport à la demande	0.29
Déménagement ou changement travail	0.23
Obligation, incitation au transport collectif	0.22
Transport commun plus sûr	0.22
Biocarburant (E85, etc.)	0.17
Transport fluvial	0.15
Ramassage scolaire	0.02

TABLE 6 – Distribution de réponses associés aux tags sélectionnés

Annexe B. Exemple d'extraction

Negation	Acteur	Action de l'acteur (verbe)	Verbe	Objet de la proposition	Concept(s) identifié(s)	pattern_name
0			renforcer	les transports en commun en ruralité par vidéo surveillance (utile contre le jet de déchet en voiture)	transport_public	Sujet - Verbe
0			rendre	(CO-voiturage, transport en commun, déplacement en vélo)	avion Auto-Mobile	Sujet - Verbe
0			encourager	la consommation de produits biologiques	transport_public vélo	Sujet - Verbe
0			favoriser	les parkings à vélo	consommation	Sujet - Verbe
0			investir	professionnels qui majoreront leurs tarifs ou renforceront encore l'assistant	parking vélo	Sujet - Verbe
1			mettre	la voiture mais les TC quand ils existent	aide prix	Sujet - Verbe
1			utiliser	de liaisons simples avec les aéroports régionaux	Auto-Mobile	Sujet - Verbe
1			offrir	nouveaux locaux tant pour favoriser la pratique du sport pour venir au travail que pour le pratiquer à la pause déjeuner	aéroport	Sujet - Verbe
1			voir	l'élargissement des routes	pratique	Sujet - Verbe
1			promouvoir	la manière de conduire et de réapprendre au gens à rouler de manière économique	route	Sujet - Verbe
0	régulièrement à tout le monde	permettre	influencer	l'efficacité de la combustion	rouler	Proposition effectuée par un acteur
0	cette technologie	permettre	augmenter	des trains plutôt que	efficacité	Proposition effectuée par un acteur
0	l'état	devoir	remettre	conditionner agriculteurs	train	Proposition effectuée par un acteur
0	L'implantation de haies	devoir	prendre	les transports en commun	aide	Proposition effectuée par un acteur
0	Les collectivités locales et l'Etat	devoir	financer	un réseau de transport ou la sécurisation des routes départementales qui le reliant aux autres communes	transport_public	Proposition effectuée par un acteur
1	un petit village	pouvoir	retarder	l'utilisation d'énergies propres,	réseau transport route	Proposition effectuée par un acteur
1	Le nucléaire	servir	prendre	la voiture	énergie	Proposition effectuée par un acteur
1	Personne	envisager	valoriser	la richesse individuelle ni l'usage distinctif de biens matériels ou de pratiques anti écologiques	Auto-Mobile	Proposition effectuée par un acteur
1	Le nouveau modèle de société à bâtir	pouvoir	proposer	de voitures roulant diesel	pratique	Proposition effectuée par un acteur
1	les entreprises en France	devoir			Auto-Mobile rouler diesel	Proposition effectuée par un acteur

FIGURE 11 – Exemple de données extraites autour de chaque proposition identifiée

Annexe C. Aperçu de l'ontologie descriptives des transports

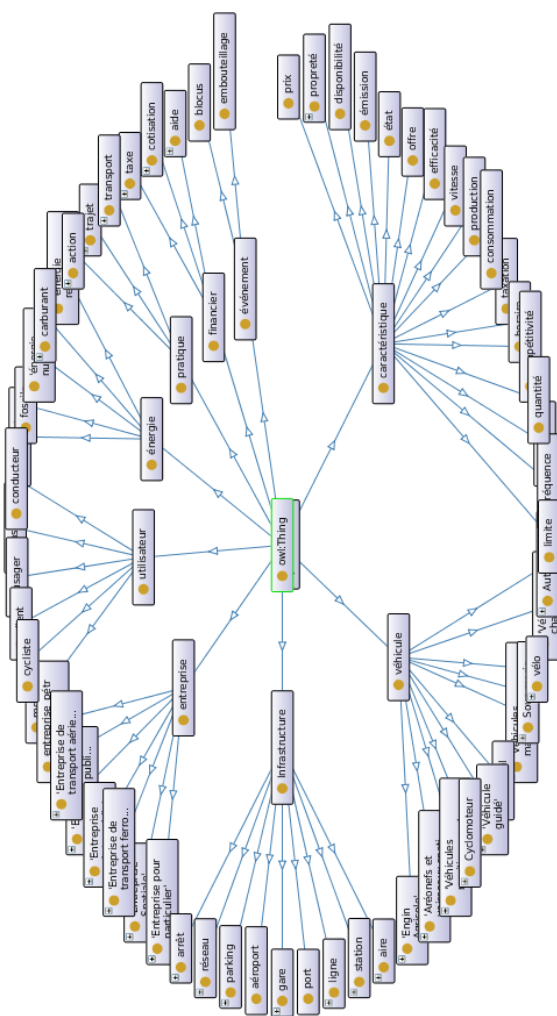


FIGURE 12 – Ontologie utilisée dans l'analyse des contributions