



**HAL**  
open science

## Les retours d'une exploration méthodologique croisant données Twitter, recrutement via Facebook et questionnaires web. La mobilité des jeunes dans et par les réseaux sociaux

Emmanuel Ravalet, Jean-François Lucas, Antoine Lohou

### ► To cite this version:

Emmanuel Ravalet, Jean-François Lucas, Antoine Lohou. Les retours d'une exploration méthodologique croisant données Twitter, recrutement via Facebook et questionnaires web. La mobilité des jeunes dans et par les réseaux sociaux. NETCOM: Réseaux, communication et territoires / Networks and Communications Studies, 2017, Mobilités et (r)évolutions numériques. Première partie. Le numérique, source d'innovations méthodologiques pour analyser la mobilité?, 31 (3/4), pp.309-334. 10.4000/netcom.2734 . hal-01941783

**HAL Id: hal-01941783**

**<https://hal.science/hal-01941783v1>**

Submitted on 2 Dec 2018

**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.

**LUCAS Jean-François**

Sociologue (PhD), expert ville numérique / Smart City chez Chronos, cabinet de conseil et de prospective.  
Chercheur associé au Laboratoire de sociologie urbaine (LaSUR), École polytechnique fédérale de Lausanne (EPFL)

---

# **Les retours d'une exploration méthodologique croisant données Twitter, recrutement via Facebook et questionnaires web**

**La mobilité des jeunes dans et par les réseaux sociaux**

2017

---

**Attention, ce texte est un document de travail.**

Il peut comporter des différences avec la version définitive du texte.

Merci de citer et de se référer à la version définitive :

**Emmanuel Ravalet, Jean-François Lucas et Antoine Lohou**, « Les retours d'une exploration méthodologique croisant données Twitter, recrutement via Facebook et questionnaires web », *Netcom*, 31-3/4 | 2017, 309-334.

Disponible en ligne : <http://journals.openedition.org/netcom/2734>

<p>Ce texte de travail est mis à disposition des internautes afin de faire connaître les travaux de LUCAS Jean-François. Les droits d'auteur de ce document demeurent aux auteurs des textes et/ou aux détenteurs des droits. Ce texte est destiné à une utilisation personnelle et non-marchande. Il ne peut être en aucun cas distribué et utilisé à des fins lucratives.</p>
---

**LES RETOURS D'UNE EXPLORATION METHODOLOGIQUE CROISANT  
DONNEES TWITTER, RECRUTEMENT VIA FACEBOOK ET  
QUESTIONNAIRES WEB :**

**LA MOBILITÉ DES JEUNES DANS ET PAR LES RÉSEAUX SOCIAUX**

*Résumé – La présence de certaines catégories de population sur internet et l'activité importante qu'ils génèrent suggèrent aux chercheurs en sciences sociales de réfléchir aux perspectives méthodologiques qui permettent de capter l'information ainsi à disposition. Cet article vise à rendre compte de deux expériences ayant en commun le fait d'utiliser les réseaux sociaux, respectivement Twitter et Facebook, dans le but de renseigner les pratiques et les représentations sociales des adolescents en matière de mobilité. Chacune de ces deux méthodes, en tant que telles et articulées au sein d'un cheminement méthodologique original, présente des avantages vis-à-vis des méthodes classiques. Mais des limites persistent et les enjeux méthodologiques liés à la mise en œuvre de ces méthodes restent de taille.*

*Mots-clés – adolescents, mobilité, réseaux sociaux, recrutement par internet, topic detection*

*Abstract – The presence of some groups of population on the Internet and the important activity they generate suggest that social scientists think about the methodological perspectives to capture such information. This article aims to report two experiences which both use social networks, respectively Twitter and Facebook, to give some insights on practices and social representations linked with mobility. These methods, on their own and articulated in an original methodological path, have some advantages compared with more classical methods. But some important limits still exist and methodological issues are major.*

*Key-words - teenagers, mobility, social networks, web-based recruitment, topic detection*

Les auteurs de cet article tiennent à remercier Toyota Motors pour le financement de ce projet. A noter qu'en aucune manière le financeur de la recherche n'est intervenu dans les résultats obtenus. Par ailleurs, le cheminement méthodologique original de cette recherche, que nous présentons dans cet article, a été bien accepté et soutenu par Toyota Motors.

Nous souhaitons également adresser nos remerciements à Arnaud Le Lann de la société Euroleads ([www.euroleads.fr](http://www.euroleads.fr)) et Sylvain Métille de la société HDC Avocats ([www.hdclegal.ch](http://www.hdclegal.ch)) pour leurs conseils sur les questions légales.

## **INTRODUCTION**

Les adolescents utilisent massivement les réseaux sociaux numériques grâce auxquels ils produisent et partagent un volume de données variées (textes, images, sons, etc.) à une vitesse sans précédent ; ces trois propriétés, volume, variété, vitesse (les « 3V ») sont caractéristiques de ce que l'on appelle communément les « Big Data ». Pour les chercheurs qui s'intéressent aux pratiques et aux représentations sociales des adolescents, ces réseaux sociaux peuvent donc apparaître comme un nouvel eldorado de la donnée. Néanmoins, au-delà de l'effet de séduction induit par les possibilités du Big Data, l'analyste doit être en mesure d'évaluer la pertinence des données qu'il souhaite et qu'il peut récupérer, au regard de la problématique traitée. En outre, le développement et la généralisation de l'usage des réseaux sociaux numériques (nous utiliserons les termes « réseaux sociaux » dans la suite du texte) par une partie de la population invitent à saisir les opportunités qui peuvent être associées à cette pratique massive, par exemple en recrutant les enquêtés par ce biais. Pour autant, les questions d'ordre méthodologique restent nombreuses et beaucoup n'ont pas encore de réponses satisfaisantes.

Cet article propose de revenir sur le montage méthodologique d'une étude qui a utilisé les réseaux sociaux pour investiguer les pratiques et les représentations des adolescents en matière de mobilité, puis pour diffuser un questionnaire à leur attention. Dans une première partie, nous apportons des éléments théoriques afin de préciser le questionnement et la population cible de notre recherche dans le but de justifier le cheminement méthodologique choisi. Dans un second et un troisième temps, nous présentons et discutons comment nous avons extrait et analysé des millions d'expressions relatives à la mobilité grâce au site de micro-blogging Twitter, puis comment nous avons procédé pour recruter des milliers de personnes sur Facebook pour une enquête web. La première de ces deux activités visait, sous une forme exploratoire, à fournir des informations complémentaires pour la construction du questionnaire web. Nous revenons dans la discussion finale sur les manières d'articuler ces méthodes et les enjeux et les limites associés en termes disciplinaires et professionnels.

### **1. RETOUR SUR LA GENESE DU CHEMINEMENT METHODOLOGIQUE**

L'effort méthodologique, quel qu'il soit, est au service d'un questionnement et en lien avec une population cible. Cette première partie a ainsi comme objectif de mettre en lien la problématique et les spécificités de la population des adolescents d'une part et la méthodologie déployée d'autre part.

#### **1.1. La problématique de la recherche**

La mobilité est appréhendée dans cet article, et dans la recherche qui en est à l'origine, comme l'articulation de trois dimensions complémentaires que sont le champ des possibles (les conditions du déplacement), la motilité (une aptitude des personnes au mouvement) et les déplacements effectifs dans l'espace (Kaufmann, 2008). Nous nous concentrons sur l'expression locale de la mobilité, déployée à l'échelle du quotidien ou de l'hebdomadaire.

L'objectif de notre recherche était de questionner l'évolution de la place de la voiture dans les représentations sociales des adolescents. En effet, le rapport à la voiture change depuis le milieu des années 2000, dans la plupart des pays occidentaux, dans le sens d'une remise en cause de sa place hégémonique dans le quotidien et d'une baisse relative de son association symbolique avec la liberté et l'autonomie (Kaufmann *et al.*, 2010). Nous cherchions, dans notre étude, à identifier les représentations sociales de la voiture pour les jeunes et à comprendre comment ils les expriment. Pour les étudiants comme pour les plus âgés, la voiture semble doucement descendre de son piédestal (Delbosc et Curry, 2013 ; Vincent *et al.*, 2015), mais les discours des adolescents sur leurs futures pratiques restent très favorables à la voiture particulière (Kopnina, 2011 ; Sigurdardottir *et al.*, 2013).

Dans ce contexte, Toyota Motors a souhaité s'appuyer sur une équipe de recherche européenne pour évaluer le rapport des jeunes (des adolescents plus spécifiquement) à la voiture.

Nous entendons par représentations sociales « *une forme de connaissance, socialement élaborée et partagée, ayant une visée pratique et concourant à la construction d'une réalité commune à un ensemble social* » (Jodelet, 2003). Les représentations sociales construisent la réalité sociale : « *toute réalité est représentée, c'est-à-dire appropriée par l'individu ou le groupe, reconstruite dans son système cognitif, intégrée dans son système de valeurs dépendant de son histoire et du contexte social et idéologique qui l'environne* » (Abric, 1994:12). D'une certaine manière, on peut dire que les représentations sociales d'une « chose » sont la cristallisation d'imaginaire(s) à un temps donné pour un certain groupe. En ce sens, et comme le souligne Moscovici (1961), les représentations sociales sont toujours représentations de quelque chose et de quelqu'un. L'approche des représentations sociales en lien avec la mobilité est d'autant plus pertinente que l'automobile, l'avion ou encore le train « *constituent l'environnement quotidien des outils ou machines qui ont impulsé des révolutions techniques, mais dont la réception et l'usage ne peuvent être ramenés à la seule commodité pratique. Ils ont aussi été investis d'imaginaires puissants et complexes qui expliquent leur appropriation, leur valorisation et parfois leur idolâtrie* » (Wunenburger, 2016:16). Par conséquent, les fonctions de ces outils et machines (principalement nous transporter) sont toujours liées à des imaginaires, des représentations sociales, des fictions, et constituent de concert la double identité d'un objet technique (Musso *et al.*, 2014).

## **1.2. Les spécificités de la population étudiée**

Dans cette recherche, la population cible est celle des adolescents âgés de 14 à 17 ans dans cinq pays européens que sont l'Allemagne, l'Espagne, la France, la Grande-Bretagne et l'Italie.

Le premier élément sur lequel il nous semble nécessaire de nous attarder concerne l'âge de ces adolescents et le rapport qu'ils entretiennent avec leurs parents. Cette question de l'autonomie relative de la population étudiée a des conséquences importantes sur le plan

méthodologique. Ainsi, dans plusieurs pays occidentaux, il n'est pas autorisé de contacter directement des adolescents pour leur faire remplir un questionnaire au sein duquel des informations personnelles sont demandées. Une autorisation parentale préalable est alors requise. Cependant, l'activité des adolescents sur les réseaux sociaux étant censée être contrôlée par les parents, l'autorisation parentale devient tacite dans le cas où le chercheur contacte les adolescents par ce biais. Sur le plan méthodologique, nous avons donc été amenés à nous orienter vers un questionnaire web avec recrutement via les réseaux sociaux. D'un point de vue plus théorique, il est important de rappeler que les adolescents ne font pas des choix libres de modes de transport et doivent souvent intégrer la décision ou l'orientation parentale dans les équations décisionnelles qui président aux pratiques. Cela joue donc un rôle sur leurs comportements de mobilité, mais peut également venir influencer leurs représentations des modes de transport et plus généralement de la mobilité. Nous nous situons finalement dans un contexte tout à fait nouveau depuis une dizaine d'années pour la tranche d'âge que nous considérons, à savoir que les adolescents ont un rapport récent et intense au numérique, ce qui peut modifier substantiellement leurs pratiques de mobilité et les représentations sociales drainées par chacun des modes de transport. En quelques années seulement, les smartphones se sont installés dans le quotidien des adolescents au point de devenir un outil quasiment incontournable de leur sociabilité (Amey et Salerno, 2015). Comme le montre le tableau 1, les taux d'équipement atteignent près de 90% pour les filles françaises de 16 et 17 ans. Toujours pour la France, Crouette *et al.* (2016) établissent même à 93% la part des 12-17 ans ayant un smartphone. Si l'on s'intéresse au taux de participation des adolescents aux réseaux sociaux, les dernières données disponibles sur Eurostat<sup>1</sup> indiquent que les 16-19 ans actifs sur un réseau social en 2016 sont 73% en France, 77% en Italie, 88% en Allemagne et 89% en Espagne et au Royaume-Uni.

Plusieurs travaux mettent en évidence l'utilisation du smartphone pour occuper ou valoriser le temps de déplacement, en particulier en transports publics (Patriarche et Huynen, 2014 ; Vincent-Geslin *et al.*, 2014). Par ailleurs, et toujours dans les transports publics, le smartphone peut également être un outil de maintien des relations à distance et de constitution d'une bulle de familiarité qui protège de l'inconnu (Vincent-Geslin *et al.*, 2014).

L'accès à internet apparaît donc comme un des enjeux principaux lié à la possession d'un smartphone. Cela nous mène à questionner un second point dans le développement de la sociabilité des adolescents, à savoir leurs usages des réseaux sociaux. Dans les faits, ils peuvent y accéder aussi bien via un équipement fixe ou portable (ordinateur, tablette, etc.), partagé ou non au sein du foyer, que via leur smartphone. Pour autant, leur activité sur les réseaux sociaux se fait très largement sur le smartphone ; elle est devenue un moyen privilégié de rester en contact avec les amis tout en échappant à la surveillance et au contrôle de leurs parents (Chaulet, 2009).

---

<sup>1</sup> <http://ec.europa.eu/eurostat/web/digital-economy-and-society/>

<b>Taux d'équipement du smartphone en France</b>		
<b>Âges</b>	<b>Filles</b>	<b>Garçons</b>
14-15 ans	79%	67%
16-17 ans	88%	74%
<b>Taux de participation à des réseaux sociaux en France</b>		
<b>Âges</b>	<b>Filles</b>	<b>Garçons</b>
14-15 ans	90%	85%
16-17 ans	92%	82%

Tableau 1 : Place du smartphone et des réseaux sociaux dans la population adolescente. Source originale : CREDOC, Enquêtes « Conditions de vie et aspirations », 2014 et 2015.

En France en 2016, 79% des 12-17 ans et 94% des 18-24 ans étaient membres d'un réseau social (Crouette *et al.*, 2016). En termes de pratiques, 92% des 12-17 ans y partageaient en 2015 photos et vidéos et 96% d'entre eux s'en servaient pour entretenir des liens avec des proches (Brice *et al.*, 2015), ce qui rejoint les conclusions d'autres travaux qui déclarent que l'utilisation des réseaux sociaux par les adolescents permet plus d'entretenir des relations qui existent au quotidien plutôt que d'en créer des nouvelles (Moreau *et al.*, 2012 ; Betbout, 2015). Ainsi, « *les réseaux sociaux (...) ne s'apparentent nullement à une communauté virtuelle, mais plutôt des micro-communautés profondément ancrées dans les relations physiques de face-à-face* » (Fluckiger, 2010). Par ailleurs, les réseaux sociaux sont des plateformes privilégiées pour l'expression des jeunes et le partage d'information. Ils relèvent ainsi d'une articulation particulière entre un espace de publication et le réseau conversationnel de l'internaute publiant (Cardon et Prieur, 2016).

La problématique de la recherche sur les représentations sociales de la voiture et la spécificité de la population adolescente nous mènent à la question de recherche autour de laquelle cet article est construit : Qu'apportent les réseaux sociaux à l'étude et à la compréhension des pratiques et des représentations de la mobilité, notamment chez les jeunes ?

Pour tenter de répondre à cette question, nous rendrons compte d'une double expérience testée dans le cadre de notre recherche, à savoir l'analyse de posts Twitter et la mise en œuvre d'une enquête web avec recrutement sur Facebook. Dans le premier cas, nous avons mis en place une démarche inductive de récupération et d'analyse d'entités textuelles afin d'identifier des pratiques et représentations associées à la mobilité des jeunes, exprimées dans des millions de discours sur Twitter. Dans le second cas, il s'agit d'une démarche plus classique avec la réalisation d'un questionnaire diffusé via Facebook, afin d'identifier des pratiques, des valeurs, des croyances ou encore des connaissances associées à l'objet de notre recherche. Nous les avons choisis sur la base de la revue de littérature que nous venons de présenter.

## 2. ENQUETER VIA LES RESEAUX SOCIAUX : EXPLORER ET REPERER DES EXPRESSIONS GRACE A TWITTER

### 2.1. Twitter et la *topic detection*

Twitter est une solution gratuite de micro-blogging créée en 2006. Dix ans plus tard, elle comptabilisait près de 313 millions d'utilisateurs actifs chaque mois<sup>2</sup>. Twitter permet de partager de courts messages de 140 caractères maximum, que l'on appelle des « *tweets* », par messagerie instantanée ou par SMS, de manière publique ou privée. Un utilisateur consulte les *tweets* des personnes auxquelles il est abonné via une liste qui agrège tous les messages qu'ils publient, sa *timeline*, et qui est actualisée en temps réel. À son tour, il peut avoir des « *followers* », c'est-à-dire des personnes abonnées aux messages qu'il publie. En plus du texte, un *tweet* peut comporter une image, une vidéo, des URL et des « *hashtags* ». Un hashtag est un mot précédé du caractère dièse (« # » ; *hash* en anglais) qui permet d'indexer des mots-clés ou des sujets sur Twitter. Cette fonctionnalité permet aux utilisateurs de suivre facilement des sujets qui les intéressent (Boullier et Lohard, 2012:62) et aux chercheurs d'étudier les phénomènes de propagation et de viralité d'une information.

De nombreux projets de recherche en informatique et en linguistique utilisent Twitter pour faire du *topic detection* (en français, « détection de sujets » ou « détection de thématiques »), car son contenu est principalement public, petit en taille, diversifié et facilement récupérable grâce à son API<sup>3</sup> gratuite ; depuis 2001 Twitter impose néanmoins de nombreuses limitations quant à la quantité de données récupérables, ce qui contraint les chercheurs à ne pas maîtriser tous les aspects de la constitution de leur corpus, et incite les analystes et les entreprises spécialisées à se tourner vers des solutions payantes pour obtenir plus de données (Felt, 2016:2). La *topic detection* consiste à identifier les thèmes les plus partagés au sein d'une large collection non structurée de documents (Blei, 2012:77). Dans le cas de Twitter, il s'agit d'identifier des sujets à partir de l'ensemble du contenu textuel des tweets, et pas seulement à partir des *hashtags*.

Ainsi, nous avons utilisé Twitter pour analyser les expressions et les thématiques les plus partagées par les jeunes à propos de leurs représentations et de leurs pratiques de la mobilité. Cette première étape a été pensée comme un travail exploratoire, fondée sur des méthodes prospectives et de créativité afin de générer des éléments pour la construction d'un questionnaire web. A noter que si les tweets recueillis s'inscrivent dans un ensemble de représentations sociales qui apparaît de manière implicite, il reste que l'exploration de tweets n'est pas une méthode spécifiquement adaptée pour la mise en évidence des représentations

---

<sup>2</sup> Aude Fredouelle, « Nombre d'utilisateurs de Twitter dans le monde », *Le journal du Net*, mis en ligne le 29/07/16, consulté le 09/02/17, <URL : <http://www.journaldunet.com/ebusiness/le-net/1159246-nombre-d-utilisateurs-de-twitter-dans-le-monde>>.

<sup>3</sup> Une *Application Programming Interface* (API) est une petite « interface de programmation » qui permet de communiquer entre différentes applications, le plus souvent pour échanger des données.



sociales. En l'occurrence, ce travail d'exploration est mis en œuvre pour venir nourrir le questionnaire web qui permettra plus directement le recueil et l'analyse des représentations sociales des jeunes vis-à-vis de la voiture et de la mobilité en général.

## **2.2. Récupération et préparation des données pour la *topic detection***

La *topic detection* à partir de données issues de Twitter compte sans aucun doute autant de manière de faire qu'elle ne compte de travaux de recherche sur ce sujet, notamment parce que les chercheurs mettent souvent en place des processus *ad hoc*. Dans notre cas, la première étape de notre processus de *topic detection* a consisté à définir des requêtes (des mots-clefs ou des combinaisons de mots-clefs) dans le but d'identifier des *tweets* relatifs à la mobilité chez les adolescents.

Premièrement, nous avons défini cinq thématiques qui permettent de couvrir, en grande partie, l'ensemble des problématiques liées à la mobilité chez les jeunes, exprimées en français sur Twitter : la « voiture », le « véhicule deux roues », le « permis de conduire », les « transports en public » et la « mobilité partagée ». Ces cinq thématiques, dont la plupart sont relatives à des modes de déplacement, doivent être appréhendées comme des « portes d'entrée » pour questionner plus spécifiquement les mobilités chez les jeunes sur Twitter. À partir d'elles, nous avons défini un nombre important et diversifié de requêtes, dans le but de cibler spécifiquement des tweets reflétant par exemple des projets, des pratiques ou encore des valeurs associées aux mobilités des jeunes pour ces cinq catégories. Le principal écueil de cette méthode est d'arriver à identifier un maximum de contenu « pertinent » tout en minimisant le « bruit », c'est-à-dire les messages qui contiennent un mot-clef que nous avons inclus dans une requête, mais dont l'usage sert à une toute autre signification que celle qui nous intéresse. Ainsi, grâce à des opérations de racinisation (*stemming* en anglais) et des opérateurs booléens, la définition de requêtes permet par exemple de récupérer des *tweets* avec le mot « train », en distinguant ceux qui ont été utilisés pour évoquer le moyen de transport de ceux qui ont été utilisés pour exprimer une activité en cours par le recours à l'expression « en train de ».

Une fois la définition des requêtes réalisée, nous avons procédé au *harvesting* (« moissonnage » en français), c'est-à-dire à la récupération des données. Malheureusement, les diverses API de Twitter ne permettent pas d'associer les requêtes au contenu textuel d'un *tweet* uniquement, c'est toute la structure d'un *tweet* qui est « interrogée », c'est-à-dire son contenu texte ainsi que ses multiples métadonnées<sup>4</sup>. Par exemple, si l'on désire récupérer les tweets contenant l'expression « permis de conduire » et qu'un compte utilisateur porte ce nom, ce sont tous les tweets de ce compte qui seront récupérés (dans la période de temps établie), même si cette expression n'apparaît dans aucun contenu publié par ce compte. Grâce

---

<sup>4</sup> Voir par exemple la « Map of a Twitter status object » produite par Raffi Krikorian, qui montre en une image la diversité des informations contenues dans un tweet. Disponible en ligne, lien consulté le 09/02/17, <URL : <http://www.slaw.ca/wp-content/uploads/2011/11/map-of-a-tweet-copy.pdf>>.

à GNIP<sup>5</sup>, une API de Twitter, nous avons collecté 2 891 573 *tweets* publics entre le 12 avril et le 13 mai 2016.

Ensuite, nous avons mis en place une phase de pré-traitement (ou *pre-processing*) qui comprenait différentes étapes pour nettoyer et préparer les données collectées en vue de leur analyse. Pour isoler les données pertinentes, les optimiser afin de les stocker et les traiter dans une base de données, nous avons eu recours à différents algorithmes afin de détecter la langue du *tweet* (*language detection*) pour supprimer les messages qui ne sont pas en français, d'extraire le contenu textuel des métadonnées (*sentence extraction*), ou encore de supprimer les mots les plus communs (*stop-word removal*) tels que « le », « la », « du », « des », etc. Enfin, nous avons procédé à une étape de normalisation de la base de données (*data normalization*), qui consiste à attribuer des règles pour formaliser et structurer la base afin de respecter la cohérence des données, de créer, supprimer ou modifier des relations entre elles et de minimiser toute redondance d'informations.

### 2.3. Méthodes d'analyse et aperçu des résultats

Plusieurs méthodes d'analyses statistiques permettent de faire du *topic detection*. La plus utilisée est certainement la méthode LDA (Latent Dirichlet Allocation), « qui suppose une attribution aléatoire des mots à un thème ou à un sujet latent, puis utilise un processus génératif de classification afin d'analyser la probabilité qu'un document contienne une information sur un sujet compte tenu de la répartition des mots qui y sont » (Bail, 2014:466). La méthode du TF-IDF (pour l'anglais « *Term Frequency-Inverse Document Frequency* ») est une méthode statistique fréquentiste fondée sur l'analyse de similarités textuelles permettant de pondérer l'importance – ou le « poids » – d'un terme contenu dans un document, relativement à un ou des ensembles de documents différents (corpus partiel ou entier). Le *term frequency* (TF) rend compte de la fréquence des mots présents dans chaque document, pondérée en fonction de la longueur de ces mêmes documents. Dans le cas de Twitter, le calcul de la fréquence inverse (IDF) est plus intéressant, car cet indicateur mesure la rareté ou la banalité d'un terme par rapport à un ensemble de documents. Nous avons utilisé la méthode TF-IDF pour cette étude, car elle est plus adaptée à des grands corpus que ne l'est la méthode bayésienne LDA.

Une première étape de classification des données a été réalisée grâce à un algorithme de correspondance de termes (*keyword matching*) afin d'obtenir deux catégories : l'une comprenant des données classifiées dans l'une des cinq thématiques définies au début de la mise en place du processus, l'autre agrégeant le reste des données, dites non-classifiées. Plus exactement, le contenu textuel du *tweet* est classifié dans l'une des cinq catégories dès lors qu'un terme ou une expression de celui-ci correspond (« *match* ») avec une requête relative

---

<sup>5</sup> GNIP est une API payante, qui offre plus de possibilités que des versions gratuites en terme de paramétrage et de quantité de données récupérables.

à cette catégorie. Comme nous l'avons expliqué précédemment, les données non classifiées correspondent par exemple à des données dont le contenu n'est pas pertinent au regard de notre problématique, mais qui a pourtant été récupéré, car une métadonnée (et non le texte du tweet) comprenait peut-être une référence à une requête.

La deuxième étape a consisté à appliquer la méthode TF-IDF. Nous avons ainsi obtenu des listes de mots pour chacune des catégories. Par exemple, les termes « taxi », « uber », « vtc », « covoiturage », « blablacar » figurent parmi les termes les plus spécifiques du domaine « mobilité partagée », quand « train », « bus », « métro » et « chauffeur » sont les plus utilisés dans la catégorie « transport en commun ». Apparemment évidents, ces premiers résultats reflètent avant tout une partie de ce « qui s'est dit » sur Twitter entre le 12 avril et le 13 mai 2016, à propos de la mobilité, et de la manière dont nous l'avons définie.

Pour aller plus loin, nous avons réalisé une troisième phase d'analyse à partir des 100 mots les plus « spécifiques » (indicateur IDF) pour chacune des cinq catégories. Pour chaque terme, nous avons consulté des dizaines voire des centaines de *tweets* qui les incluaient. Cela nous a permis de définir des *clusters* (ou catégories) transversaux aux cinq catégories préalablement établies. Par exemple, nous avons considéré que de nombreux *tweets* reflétaient l'idée selon laquelle la mobilité définit un acte quotidien de déplacements spatio-temporels ponctués de micro-événements. De prime abord, ce constat peut sembler logique puisque Twitter est une application de micro-blogging qui semble favoriser ce mode d'expression, c'est-à-dire spontané, lié à « l'événement », qui plus est dans un contexte de mobilité. Néanmoins, ce n'est pas parce que ce dispositif sociotechnique rend possible ce type de pratique que tous les utilisateurs l'adoptent. En ce sens, les résultats obtenus lors de cette étude nous laissent présager que les adolescents sont plus enclins à partager des « événements » qui subviennent lorsqu'ils se déplacent, que d'autres populations. Voici quelques exemples de *tweets* illustrant cette proposition :

- « Je suis assis sur le quai dans le metro et la je vois qui en face de moi benoit du mag des anges »<sup>6</sup>.
- « Allé direction Paris avec la petite sœur pour la deuxième fois qu'elle monte dans le train et première fois qu'elle prend le métro! ».
- « Y a eu un accident sur la route de mon bus ce matin cetait chaud ».
- « j'ai pas eu le métro j'vais encore être en retard ».

Par ailleurs, de nombreux *tweets* illustrent l'idée selon laquelle la mobilité peut être appréhendée comme l'expérience d'une tension permanente entre l'individuel et le collectif :

- « pq y'a toujours des gens chelous qui s'assoient à côté de moi dans le tram? »
- « Dite moi qui est celui qui a décidé qu'on pouvait accepté des gens bourrés dans le bus ?!!! Merde ! ».

---

<sup>6</sup> Les tweets sont retranscrits tels qu'ils ont été écrits par les utilisateurs.

- « Les gens dans le bus, le gel douche ça existe !! ».

D'autres *tweets* dévoilent que la musique et la connectivité sont des ressorts principaux pour les adolescents qui permettent de s'isoler des nuisances extérieures et de se créer une « bulle » :

- « Je vais prendre le métro et j'ai oublié mes écouteurs chez moi. J'ai envie de mourir. »
- « Vous aussi vous commencer à faire semblant d'envoyer des sms et à bien replacer vos écouteurs quand une femme enceinte monte dans le bus ? »

D'autres *clusters* que nous avons définis reflètent la tension entre la liberté désirée de l'adolescent et sa subordination à ses parents, les plaintes des jeunes faces aux dysfonctionnements du matériel ou des équipements qui permettent la mobilité, ou encore l'importance que représentent les cours de conduite dans la vie d'un adolescent qui prépare le code ou le permis de conduire.

Ce travail exploratoire a pris la forme d'un rapport (Lucas et Boullier, 2016). Les résultats qui y sont présentés ont permis d'aider à la construction d'un questionnaire quantitatif interrogeant les représentations et les pratiques de mobilité des adolescents. Nous proposons maintenant de passer à la deuxième utilisation qui a été faite des réseaux sociaux dans notre recherche, cette fois dans la perspective du recrutement des adolescents susceptibles de remplir ce questionnaire.

### **3. ENQUETER VIA LES RESEAUX SOCIAUX : RECRUTER DES ADOLESCENTS GRACE A FACEBOOK**

#### **3.1. Une campagne de publicité et une page sur Facebook pour recruter des adolescents**

La seconde phase d'analyse de notre travail d'investigation a consisté en un recueil de données quantitatives avec pour ambition la représentativité de la population étudiante ciblée, à savoir les 14-17 ans dans cinq pays. Pour recueillir les données dont nous avons besoin en matière de pratiques effectives de mobilité et de représentations sociales, nous avons fait le choix d'une enquête web. Un questionnaire a été rédigé puis traduit pour que nous puissions le passer auprès des adolescents français, anglais, espagnols, allemands et italiens. Remplir ce questionnaire prenait, selon les estimations issues de la phase de test, quinze minutes environ. Cette durée peut déjà être considérée comme très élevée compte tenu du support web et de la population (Galesic et Bosnjak, 2009).

Pour atteindre les adolescents, nous avons décidé d'utiliser le réseau social Facebook, bien qu'il soit souvent fait mention du fait qu'il est moins utilisé désormais que Snapchat ou Instagram par les adolescents. Néanmoins, nous avons opté pour cette solution, car ses

fonctionnalités sont les plus adaptées pour ce type de recherche (création d'une page dédiée au projet, diffusion de contenus de différents formats, etc.) et parce que son interface publicitaire permet de sélectionner l'âge et la position géographique des utilisateurs pour un meilleur ciblage.

Le recours à la publicité sur internet pour recruter les répondants d'une enquête n'est pas nouveau. Plusieurs exemples de ce type sont documentés dans la littérature, avec une inscription disciplinaire plutôt reliée à la médecine (Fenner *et al.*, 2012 ; Bauermeister *et al.*, 2012 ; Morgan *et al.*, 2013 ; Ramo et Prochaska, 2012). Pour autant, ces expériences restent trop peu nombreuses pour permettre un recul suffisant sur l'efficacité financière de ce type de recrutement et la (non-)représentativité des échantillons obtenus vis-à-vis de la population ciblée (Schonlau *et al.*, 2009). Les expériences documentées sont généralement issues d'un terrain unique. Aussi, nous n'avons pas trouvé d'éléments sur la manière dont ce type de recrutement pouvait être dépendant du pays dans lequel il est mis en place. C'est donc sur ces éléments que nous proposons de revenir maintenant.

Deux différentes voies ont été testées, exactement comme l'a fait Batterham (2014) autour de questions relatives à la santé mentale. Nous avons tout d'abord décidé de lancer une campagne publicitaire sur Facebook en lien avec un site internet externe sur lequel l'enquête était hébergée<sup>7</sup>. Cette phase relevait d'un essai réalisé uniquement sur la France en mai 2016. Ce test s'est révélé extrêmement décevant, bien plus encore que ce que les résultats obtenus par Batterham (*ibid.*) ne le laissaient supposer. Pour un peu plus de 1000 francs suisses (CHF) dépensés, soit près de 1000 euros, et une campagne de dix-huit jours, nous n'avons récupéré que cinq questionnaires remplis. Cette expérience nous a incités à passer à une seconde méthode de recrutement, très proche finalement de la première, mais autrement plus efficace sur le plan financier.

La seconde méthode consistait à financer une campagne publicitaire pour une page Facebook dédiée à l'enquête, plutôt que vers un site indépendant et extérieur au réseau social. Pour développer l'attractivité de la page Facebook, nous publions des articles en lien avec les transports et la mobilité et renvoyant vers des articles de journaux, des vidéos publicitaires humoristiques, ou des reportages TV (une actualité par semaine en moyenne était publiée). Cette technique, comme la précédente, a d'abord été testée en France, pendant trois semaines, avec un financement de la campagne publicitaire de 50 CHF par jour. Au bout de cette période, et pour un montant équivalent à ce qui avait été dépensé pour le test de la première méthode, nous avons récupéré plus de 200 enquêtes remplies en totalité. Nous avons donc opté pour cette seconde méthode, poursuivi le terrain français et lancé les terrains dans les quatre autres pays. Nous avons alors ouvert quatre autres pages Facebook, afin d'en avoir une dans chaque langue. Chaque page était modérée par une personne dont la langue maternelle était celle du pays enquêté. Nous avons donc fonctionné avec une équipe de cinq

---

<sup>7</sup> Il s'agit du site internet du logiciel avec lequel nous avons codé l'enquête, à savoir Unipark.

modérateurs. Cela permettait une grande réactivité pour répondre aux commentaires et gérer la préparation des publications. Le recueil des questionnaires s'est étalé jusqu'à la fin du mois de septembre 2016 et a duré de 6 semaines en Italie et en Espagne jusqu'à 12 semaines en France. Nous cessons le financement de la campagne publicitaire pour chaque page lorsque nous atteignons 3000 enquêtes partiellement remplies dans chaque pays<sup>8</sup>.

### 3.2. Une forte variabilité des résultats selon les pays

	France	Allemagne	Espagne	Italie	Royaume-Uni
Nombre de "J'aime" enregistrés sur la fanpage Facebook	7994	6095	2542	5551	3479
Nombre de clics vers l'enquête à partir de la fanpage Facebook	8844	3642	5531	5654	3008
Nombre d'enquêtes débutées	3019	3001	3034	3305	2466
Nombre d'enquêtes terminées	1988	1656	1704	1789	1056
Coût par enquête débutée	1,26 CHF	1,99 CHF	0,80 CHF	0,63 CHF	2,54 CHF
Coût par enquête terminée	1,91 CHF	3,61 CHF	1,43 CHF	1,17 CHF	5,94 CHF
Part des enquêtés participant au tirage au sort	67.8%	55.6%	61.4%	47.8%	45.7%

Tableau 2 : l'enquête web en chiffres dans les cinq pays

Le tableau 2 montre la variabilité des résultats de la méthode de recrutement selon les pays. Ainsi, le taux de remplissage en totalité (part des adolescents qui ont rempli totalement le questionnaire sur l'ensemble de ceux qui l'ont débuté) est de 55% environ en Allemagne, en Espagne et en Italie, mais atteint 66% en France et chute à 43% au Royaume-Uni. Il est difficile d'expliquer ces deux écarts, en particulier parce que nous posons les questions de caractérisation socio-démographique en fin de questionnaire, ce qui empêche de comparer les profils socio-démographiques des deux populations (ceux qui ont fini le questionnaire et ceux qui ont abandonné avant la fin). La présence de plusieurs questions ouvertes et la longueur du questionnaire font sans doute partie des raisons qui expliquent ces abandons (Galesic et Bosnjak, 2009). Par ailleurs, 87% des Anglais, 89% des Espagnols, 90% des Français, 92% des Allemands et 96% des Italiens ont répondu à l'enquête depuis leur smartphone. La possibilité pour un enquêté de répondre à l'ensemble du questionnaire va

---

<sup>8</sup> Ce seuil constituait notre engagement contractuel. Ce choix du nombre d'enquêtes partiellement remplies plutôt que du nombre d'enquêtes totalement remplies s'explique par la forte incertitude liée au remplissage d'un questionnaire web de 15 minutes par des adolescents. Le seuil n'a pas été atteint au Royaume-Uni compte-tenu des délais importants et des coûts élevés du recrutement dans ce pays.

donc dépendre beaucoup du moment choisi pour commencer l'enquête, du temps à sa disposition et du confort de sa situation...

Les différences entre pays sont plus marquées encore si l'on observe le nombre de personnes ayant « aimé » la fanpage Facebook ou celles ayant cliqué vers l'enquête depuis cette fanpage.

Ainsi, il a fallu que près de 9000 personnes cliquent en France depuis la page Facebook vers l'enquête pour atteindre les 3000 enquêtes débutées, alors que 5500 environ ont suffi en Espagne et en Italie et 3500 environ en Allemagne. Au Royaume-Uni, 3000 personnes sont arrivées sur la page de l'enquête et 2500 se sont lancés dans le remplissage, ce qui est un résultat très élevé.

Sur les questions d'efficacité en termes de coût de la méthode de recrutement, les différences entre pays sont encore une fois très fortes. Contrairement au Royaume-Uni, la méthode s'est avérée très efficace en Italie (1.17 CHF en Italie par enquête totalement remplie contre 5.94 CHF au Royaume-Uni) ; les trois autres pays étant dans des situations intermédiaires à ces deux extrêmes. L'explication de ces différences n'est pas aisée et le fait d'animer la page Facebook ne semble pas avoir eu un impact direct sur le succès de la page et de l'enquête. Par ailleurs, la difficulté rencontrée au Royaume-Uni pourrait être expliquée par le grand nombre de publicités qui circulent en anglais comparativement à celles qui circulent dans les autres langues, rendant potentiellement moins efficace chacune d'elle. Nous n'avons malheureusement pas trouvé de sources permettant d'étayer cette hypothèse.

Deux autres éléments d'ordre méthodologique méritent d'être précisés. Le premier visait un effet boule de neige dans le recrutement des adolescents en proposant aux répondants, à la fin du questionnaire, d'inviter leurs amis à participer à cette enquête. Le partage de cette information sur Twitter ou sur Facebook n'a été que très faiblement réalisé ; seulement 10 questionnaires totalement remplis l'ont été grâce à cette méthode. Autre choix méthodologique, nous avons mis en place un système d'*incentive*, qui mentionnait la réalisation d'un tirage au sort parmi les répondants au questionnaire afin de gagner cinq bons d'achat Apple d'une valeur de 100 CHF chacun. Cet artifice a également connu un succès très relatif puisque 34% des Français ont refusé de laisser leur adresse mail pour participer au tirage au sort, et ils sont 40% en Espagne, 47% en Allemagne, 54% en Italie et 57% au Royaume-Uni à avoir fait ce choix.

### **3.3. Profils et représentativité des répondants**

Dans notre échantillon, certains profils sociodémographiques sont surreprésentés, si l'on s'en tient aux données de recensement disponibles dans les cinq pays. Ainsi, notre échantillon est

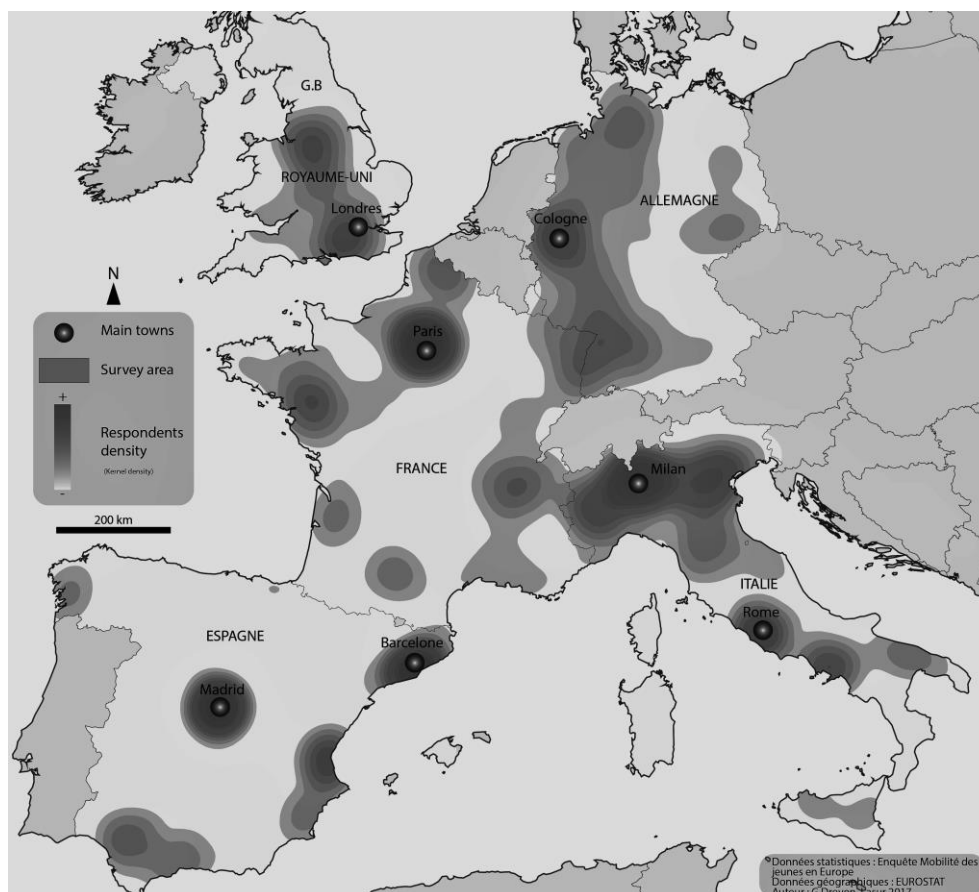
plutôt féminin, sauf en Allemagne et au Royaume-Uni où la tendance est plus équilibrée. Les plus âgés (16-17 ans) ont également plus répondu que les plus jeunes (14-15 ans). Ces deux résultats doivent s'interpréter comme une tendance liée à la pratique des réseaux sociaux dépendamment du genre et de l'âge. Si l'on s'en tient aux informations présentées dans le tableau 1 sur la France, on avait bel et bien une surreprésentation des filles et des adolescents les plus âgés parmi les utilisateurs du smartphone et des réseaux sociaux.

Le paramétrage des publicités permettait de n'afficher les publicités que sur les profils Facebook dont l'utilisateur est âgé entre 14 et 17 ans. Pour autant, l'âge indiqué par les internautes sur leur compte Facebook dépend de celui que l'individu a choisi lors de la création de son compte. Ce n'est donc pas nécessairement une donnée sûre. Les très jeunes peuvent par exemple être amenés à déclarer un âge supérieur au leur, afin d'ouvrir un compte Facebook avant d'avoir l'âge légal qui est de 13 ans. Cette pratique n'est pas anodine si l'on s'en tient aux informations récoltées au cours de la recherche EUKidsonline, qui montrent que 38% des 9-12 ans ont un compte sur un réseau social en Europe, et que Facebook était en 2010, le plus répandu parmi ces jeunes (Livingstone *et al.*, 2011). Pour limiter les possibles biais liés à cette information, nous demandions aux enquêtés leur âge à la fin du questionnaire. 2% à 4% des répondants selon les pays avaient alors moins de 14 ans ou plus de 17 ans et ont été retirés de notre échantillon pour l'analyse.

Nous observons également une légère surreprésentation des répondants qui vivent en milieu urbain vis-à-vis des résidents des zones rurales que la carte 1 permet de mettre en évidence.

Finalement, le recrutement des adolescents grâce au réseau social Facebook, notamment par la création de pages dédiées à l'enquête et à leurs promotions grâce à de la publication de contenus et des publicités ciblées sur la population visée, apparaît être une méthode relativement performante sur le plan de l'efficacité financière. Des biais d'échantillonnage existent et se doivent d'être corrigés par la construction d'un coefficient de pondération, en particulier sur les plans de l'âge, du genre et de la localisation résidentielle des répondants. Mais il n'y a pas de raisons que ces biais soient plus problématiques que ceux qui existent pour les autres méthodes. Finalement cette méthode s'avère très pertinente pour toucher des populations spécifiques comme l'est celle des adolescents (Fenner *et al.*, 2012), d'autant plus que malgré la longueur relative du questionnaire et la présence de plusieurs questions ouvertes, l'enquête a été remplie avec un grand sérieux. Sur les 8193 enquêtes remplies reçues, seules deux ont dû être supprimées à cause de la très faible qualité des réponses données.





Carte 1 : densité des répondants au questionnaire diffusé sur Facebook (réalisation : Guillaume Drevon, LaSUR).<sup>9</sup>

#### 4. DISCUSSION : DES LIMITES DE L'USAGE DES RESEAUX SOCIAUX A LA MISE EN ŒUVRE DE CHEMINEMENTS METHODOLOGIQUES ORIGINAUX

Dans cet article, nous avons fait le choix de décrire successivement la méthodologie mise en place lors de deux phases d'une recherche sur les pratiques et les représentations de la mobilité chez les adolescents. Nous voudrions revenir sur leur articulation et sur les enjeux et les limites théorique et méthodologique qu'elle implique.

---

<sup>9</sup> Cette carte a été construite sur la base d'une interpolation avec la méthode des noyaux de Kernel.

#### 4.1. Les enjeux de représentativité en lien avec l'usage des réseaux sociaux dans une démarche de recherche.

Deux utilisations très différentes des réseaux sociaux ont été présentées dans les parties 2 et 3 de cet article. Dans le premier cas, grâce à Twitter, nous avons analysé des « expressions » partagées et discutées par les adolescents à propos de leurs expériences et représentations de la mobilité. Dans le second cas, nous avons utilisé Facebook pour contacter les enquêtés, en nous appuyant sur plusieurs informations données par les internautes (en particulier l'âge et le pays de résidence). Nous proposons de discuter maintenant des enjeux de représentativité liés à ces deux méthodes de recueil de l'information via les réseaux sociaux.

Les expressions recueillies sur Twitter ne peuvent être considérées comme équivalentes à « l'opinion » constituée par les sondages, notamment parce que nous ne connaissons rien des propriétés socio-démographiques des émetteurs (Boullier et Lohard, 2012). Ainsi, à l'image des réserves qui ont été émises à propos de la qualité des informations liées aux profils des répondants au questionnaire diffusé sur Facebook dans le paragraphe précédent, ni le nom ni la photo qui apparaissent sur le profil d'un utilisateur de Twitter ne doivent être considérés comme des indicateurs attestant de son identité. Dès lors, comme le notent Boullier et Lohard, il faut faire le « deuil » de la représentativité et de l'exhaustivité avec ce type de données<sup>10</sup>, car ce qui est étudié « *ne fait plus référence à des émetteurs bien identifiés dont les chercheurs posséderaient les caractéristiques socio-démographiques classiques et considérées comme définitoires de la personne (en tant que sujet statistique tout au moins). Ici, les énoncés sont nivelés et c'est leur agrégation qui fait sens (...)* Les méthodes pour retrouver ces repères socio-démographiques dans les métadonnées et indices divers exploitables sur les sites ou plates-formes demandent encore de grandes avancées (...) Il faut donc faire le deuil et pour longtemps des repères socio-démographiques sur le web actuel et pour cela changer de paradigme d' "explicitation" dans les sciences humaines et sociales, tellement fondées sur une théorie trop implicite de l'individu comme individu et non comme sociation » (ibid, 189-190).

Par ailleurs, les données issues de Twitter ont été récoltées à partir d'une sélection de mots-clés, sans lien avec les caractéristiques sociodémographiques des utilisateurs. Autrement dit, il n'est pas possible de savoir si les messages que nous avons collectés ont été postés par des adolescents. Néanmoins, pour tenter de pallier à cette limite, nous avons créé un filtre lors de la phase d'analyse des messages. Pour cela, grâce aux données de l'INSEE, nous avons d'abord créé une liste des 200 prénoms les plus attribués aux personnes nées en France entre 1999 et 2002 (en faisant le choix de 100 prénoms féminins et de 100 autres masculins). Deuxièmement, nous avons développé un algorithme afin d'identifier tous les noms d'utilisateurs que nous avons dans notre base de données (1 610 650 utilisateurs pour 2 891

---

<sup>10</sup> Précisions que les auteurs font alors référence à un corpus web pour de l'analyse de sentiment (*sentiment analysis*).

573 tweets) qui se rapprochaient le plus des « racines » des 200 prénoms isolés précédemment. Grâce à cette technique, nous avons isolé des milliers d'utilisateurs. Pour aller plus loin dans cet objectif de sélection des profils des plus jeunes, nous avons consulté, manuellement, des centaines de profils (tous publics rappelons-le) afin de regarder la photo de profil et lire des messages qui étaient postés. Bien évidemment, cette méthode ne permet pas de certifier que l'utilisateur d'un compte a entre 14 et 17 ans, mais elle nous a néanmoins semblé relativement pertinente pour différencier, au moins, des utilisateurs qui semblaient être des adolescents d'autres utilisateurs plus âgés. En somme, cette méthode empirique nous a permis d'affiner certains de nos résultats, tout en restant prudents quant à la possibilité de relier un « profil » à un *tweet*.

Une fois ces limites identifiées (et acceptées !), ces matériaux (ici des *tweets*) peuvent être considérés pour ce qu'ils sont, c'est-à-dire des autopublications numériques caractéristiques de la conversation et du bavardage sur le web, qui ne sont pas produites en « situation de laboratoire » (Boyadjian, 2016), mais « sur le vif » et sans aucun biais déclaratif.

Pour ce qui est de notre travail de recrutement des enquêtés sur Facebook, le débat relatif à la représentativité est un peu différent. Sur ce réseau social, il est possible de diffuser des contenus en fonction de l'âge et de la localisation des personnes. Même si l'utilisateur peut avoir menti sur son âge ou peut avoir eu connaissance du questionnaire par un autre biais que la publicité qui a été diffusée vers les profils Facebook des 14-17 ans, la marge d'erreur sur cet indicateur est relativement faible et peut être corrigée en questionnant une seconde fois les enquêtés sur ces variables dans le questionnaire (rappelons que 2% à 4% des répondants seulement ont déclaré avoir moins de 14 ans ou plus de 17 ans à la fin du questionnaire). Ainsi, les informations sociodémographiques liées au profil des utilisateurs Facebook n'ont qu'un rôle partiel de ciblage vers les personnes susceptibles d'être concernées par l'enquête. Les informations sociodémographiques utilisées ensuite dans l'analyse des résultats sont issues de l'enquête elle-même. Mais peut-on assurer la représentativité de l'échantillon finalement obtenu ? La première limite est celle de la participation à Facebook pour les adolescents de cette tranche d'âge. Si la part des jeunes ayant un compte Facebook semble relativement élevée (elle serait de 79% pour les 15-18 ans en France selon une récente étude de Génération Numérique<sup>11</sup>), ils n'y sont pas toujours actifs. Ainsi, comme évoqué dans la partie 1, 75% (pourcentage pour la France) à 90% des adolescents de 16 à 19 ans participeraient activement à un réseau social selon le pays (qui peut ou peut ne pas être Facebook). Il est clair en ce sens que l'échantillon finalement obtenu souffre d'un biais directement lié à cette participation aux réseaux sociaux. Finalement, et grâce à une repondération des répondants sur la base de l'âge et du genre, nous avons limité en partie la déformation de l'échantillon lié à la non-réponse. Pour autant, et dans la mesure où les profils

---

<sup>11</sup> <http://asso-generationnumerique.fr/11-18ans-reseaux-sociaux/>

sociodémographiques des usagers de Facebook sont mal connus, des biais peuvent persister selon la position sociale ou encore selon la localisation résidentielle.

#### **4.2. Articuler les méthodes au sein de cheminements originaux**

La méthode d'exploration de contenus issus de Twitter que nous avons présentée ne prétend pas répondre, seule, aux questions de recherche qui animent notre travail, parce qu'elle ne peut prétendre à l'exhaustivité et ne donne qu'une image « furtive » de ce qui est partagé sur ce réseau social. Néanmoins, cette méthode possède des vertus exploratoires fortes à développer.

Premièrement, elle permet de recueillir des discours qui sont émis sans passer par le filtre de l'enquête scientifique (questionnaire, entretiens, etc.), bien qu'ils soient formatés selon des modalités imposées par le dispositif technique (140 caractères par exemple). Deuxièmement, les résultats fournissent aux chercheurs une sorte de cartographie de « ce qui se dit », dans un périmètre donné (ici Twitter, pour trente jours), à propos d'une ou de plusieurs thématiques.

Sur la base de ces éléments, nous avons articulé les deux méthodes qui ont été présentées dans cet article. Ainsi, les résultats obtenus lors de cette première phase, exploratoire, ont incité les responsables de l'enquête web à intégrer une nouvelle section dans leur questionnaire, afin d'interroger plus spécifiquement le rapport des jeunes aux technologies de l'information et de la communication lorsqu'ils se déplacent, comme des outils pour mieux occuper le temps de transport, mais qui permettent également de limiter le sentiment d'insécurité. De fait, en soulignant certaines thématiques particulièrement abordées sur Twitter par les adolescents à propos de la mobilité, diverses questions ont pu être affinées, et d'autres ajoutées. Ainsi, par la vitalité et la diversité qui caractérisent leur production, les millions de *tweets* se sont révélés être signifiants pour l'exploration, le repérage et l'identification de pratiques et de représentations sociales de la mobilité chez les adolescents, pour les chercheurs qui ont réalisé le questionnaire web. Les informations recueillies à travers les *tweets* récupérés ont en outre permis de nourrir la grille d'entretien à l'occasion d'une phase qualitative construite autour d'entretiens approfondis avec des adolescents et réalisés après le questionnaire web.

En écho aux méthodes de « photo-élicitation » (Rose, 2003), qui renvoie à l'utilisation de photographies pour appuyer un questionnement lors d'entretiens approfondis, il pourrait être envisagé de construire des entretiens sous forme de « *tweet elicitation* ». De la même manière, on pourrait imaginer de la « *comment elicitation* » à partir des contenus Facebook ou encore de la « *photo elicitation* » avec Instagram.

### 4.3. Enjeux et limites de la collaboration entre des chercheurs de disciplines différentes

Notre méthode est combinatoire à différents niveaux. Premièrement, nous avons combiné « la capacité humaine à comprendre et à interpréter – ce que les ordinateurs ne peuvent pas encore complètement faire – avec la capacité des ordinateurs à analyser des jeux massifs de données en utilisant des algorithmes que nous créons » (Manovich, 2012). Deuxièmement, nous avons croisé des méthodes d'enquête traditionnelles en sciences sociales avec des méthodes de collecte et de traitement de grands ensembles de données qui sont spécifiques au support numérique (Rogers, 2013). Enfin, cette étude résulte de la combinaison de différents savoirs et de différentes compétences, fruit de la collaboration entre des chercheurs en sciences humaines et sociales avec des chercheurs en *computer sciences*, qui ont conceptualisé et codé les algorithmes et le processus de *topic detection*. Comme le souligne Christopher A. Bail, la collaboration entre des chercheurs en sciences humaines et des *computer scientists* est potentiellement fructueuse d'un point de vue scientifique, car, si les *computer scientists* développent des outils puissants pour l'analyse automatique des Big Data, ils manquent souvent de cadres théoriques pour extraire le sens des données analysées, quand les chercheurs en sciences sociales (les « *cultural sociologists* » dans son cas) produisent des cadres théoriques sophistiqués, mais n'ont pas les compétences requises pour les explorer autrement qu'à un niveau micro, c'est-à-dire qualitatif (Bail, 2014 :465).

Dans notre cas, la collaboration avec les *data scientists* de l'EPFL, qui ont modélisé et développé le processus de *topic detection* à partir des données Twitter, nous a permis de connaître et contrôler la plupart des choix conditionnant chaque étape du processus, notamment la phase de *pre-processing*. Par exemple, ce qui apparaissait comme du contenu « non pertinent » pour certains, semblait pertinent pour d'autres. Ce constat, somme toute banal, s'avère pourtant fondamental dès lors qu'il devient possible de changer les conditions de prise en compte de tel ou tel contenu, quand ce type de décision est encapsulé dans des algorithmes que des chercheurs en sciences humaines utilisent souvent comme des kits clef en main, avec l'impossibilité de connaître ces choix, et encore moins de les modifier. Par exemple, de nombreux chercheurs en sciences humaines utilisent les réseaux sociaux, et Twitter notamment, pour faire de la *topic detection*. Le problème, malheureusement trop peu évoqué, est que les compétences techniques et les moyens financiers dont ils disposent les contraignent souvent à utiliser des dispositifs opaques ou peu paramétrables (comme les API gratuites) pour lesquels ils ne connaissent rien des modèles qui sont à l'origine des algorithmes constitutifs de ces boîtes noires. De fait, dans les travaux en sciences humaines, par choix ou non, les nombreuses étapes, processus et procédures qui permettent de récupérer les données, de les nettoyer, de les stocker ou encore de les classer, sont rarement renseignées : « tout ce passe comme si entre les données et leurs effets, une ligne directe, immédiate et automatique conduisait des *data* aux résultats » (Bastard *et al.*, 2013). Or, car les données ne sont pas brutes et qu'elles ne parlent jamais d'elles-mêmes (Bowker, 2005:184 ; Gitelman, 2013), tout travail qui requiert ce type d'opérations implique un travail d' « explicitation »

(Boullier, 2016) afin de déplier les choix et les actions des « travailleurs de la donnée » (Bastard *et al.*, 2013), pour comprendre le passage des « données » aux « obtenues » (Latour, 1999).

Pour autant, la collaboration entre des chercheurs en sciences humaines et sociales et des chercheurs en *computer sciences* implique de négocier au mieux avec les différentes « positions de savoir » (Darré, 1999 ; Morrissette et Desgagné, 2009) des différents acteurs, mais aussi de s'accorder sur des éléments de langage et une vision convergente (Star et Griesemer, 1989), ce qui nécessite un « travail d'articulation » (Strauss, 1988), c'est-à-dire des opérations de coordination afin de faire converger des points de vue différents, et des investissements en temps parfois coûteux. Or, la finalité de ces investissements diffère justement selon les acteurs, car les chercheurs en sciences humaines et sociales vont par exemple chercher à comprendre le sens des données récupérées, quand les chercheurs en *computer sciences* chercheront d'abord à optimiser les algorithmes (gain temporel, robustesse et fiabilité des systèmes, etc.), sans questionner les ontologies à partir desquelles ils conditionnent la récupération des données.

## CONCLUSION

Pour être en mesure de rendre compte des pratiques et des représentations sociales de la mobilité des adolescents en Europe, nous avons investigué cette thématique dans et par les réseaux sociaux. Le recueil et l'analyse d'expressions issues de Twitter ont été pensés dans le cheminement méthodologique de notre recherche comme un travail exploratoire visant à mettre en évidence des thématiques, des questions ou encore des témoignages du vécu permettant d'enrichir et de préciser le questionnaire quantitatif qui suivait. Une fois le questionnaire construit, nous avons utilisé Facebook pour recruter les adolescents, et ainsi toucher une population que l'on sait difficilement accessible dans les enquêtes.

Contrairement à une phase d'exploration plus « classique » pour une problématique donnée, qui peut être réalisée par une courte phase d'observation, quelques entretiens exploratoires ou encore l'identification d'un nombre limité de *tweets*, notre démarche présente une manière originale d'appréhender une thématique à partir de millions de données. En somme, nous avons utilisé des millions de tweets pour faire de l'élicitation pour le chercheur. Néanmoins, nous avons détaillé les étapes de construction du processus de récupération et d'analyse de ces matériaux afin de montrer qu'ils ne sont pas simplement « disponibles » en ligne, que leur récupération implique de nombreuses « médiations » qui les « traduisent » (Akrich *et al.*, 2006) et qui participent au façonnage des résultats. Nous avons également insisté sur le fait qu'il est impossible d'affirmer pouvoir relier ces données à des « individus ».

Pour ce qui est du recrutement réalisé via Facebook, nous considérons cette méthode comme efficace sur le plan économique (le prix par enquête rempli est très bas), rapide et pertinente pour construire des échantillons de grande taille, et relativement sûre, mais sujet à précaution

sur le plan de la représentativité des échantillons obtenus. Elle nous a permis de capter l'attention des jeunes en les contactant via les canaux qu'ils utilisent.

Le recours aux réseaux sociaux pour pouvoir adresser nos questions de recherche a finalement permis de mettre en évidence plusieurs forces et faiblesses propres à ces méthodes. De fait, l'explicitation (Boullier, 2016) de l'ensemble des choix réalisés apparaît comme une condition indispensable à une utilisation éclairée, féconde et transparente des réseaux sociaux dans la recherche en sciences sociales. L'articulation de ces méthodes se révèle également prometteuse.

Au-delà de l'objet de ce présent chapitre, quelques résultats peuvent être mentionnés. Ainsi, nous avons pu mettre en évidence la place encore prépondérante de la voiture dans l'imaginaire des adolescents et la très grande importance de l'insécurité dans les pratiques et représentations des adolescentes en matière de mobilité. Ces éléments feront l'objet de futures publications.

## **BIBLIOGRAPHIE**

ABRIC, J.-C. (ed.) (1994), *Pratiques sociales et représentations*, Paris, PUF, 251p.

AKRICH, M., CALLON, M., LATOUR, B. (ed) (2006), *Sociologie de la traduction : textes fondateurs*, Paris, Mines Paris, les Presses, 304p.

AMEY, P., SALERO, S. (2015), Les adolescents sur Internet : expériences relationnelles et espace d'initiation, *Revue française des sciences de l'information et de la communication* [En ligne], 6 | 2015, mis en ligne le 23 janvier 2015, consulté le 08 février 2017, <URL : <http://rfsic.revues.org/1283>>, <DOI:10.4000/rfsic.1283>.

BAIL, C.A. (2014), The cultural environment : measuring culture with big data, *Theory and Society*, vol.43 (3), 465-482.

BASTARD, I., CARDON, D., FOUETILLOU, G., PRIEUR, C., RAUX, S., (2013), Travail et travailleurs de la donnée, *Internetactu*, [en ligne], publié le 13 décembre 2013, consulté le 9 février 2017, <URL : <http://www.internetactu.net/2013/12/13/travail-et-travailleurs-de-la-donnee>>.

BATTERHAM, P.J. (2014), Recruitment of mental health survey participants using Internet advertising: content, characteristics and cost effectiveness, *International Journal of Methods in Psychiatric Research*, vol.23 (2), 184–191, <DOI:10.1002/mpr.1421>.

BAUERMEISTER, J.A., ZIMMERMAN, M.A., JOHNS, M.M., GLOWACKI, P., STODDARD, S., VOLZ, E. (2012), Innovative Recruitment Using Online Networks:

Lessons Learned From an Online Study of Alcohol and Other Drug Use Utilizing a Web-Based, Respondent-Driven Sampling (webRDS) Strategy, *Journal of Studies on Alcohol and Drugs*, vol.73 (5), 834–838, <DOI:10.15288/jsad.2012.73.834>.

BETBOUT, A. (2015), Les Réseaux Sociaux Numériques: Nouvelle forme de convivialité sociale. Presented at the Les réseaux sociaux numériques. Un espace de liberté à prendre ou une utopie mort-née ?, Nov. 2015, Strasbourg, France, <hal-01291886>.

BLEI, D.M. (2012), Probabilistic Topic Models, *Communications of the ACM*, vol.55 (4), 77-84.

BOULLIER, D. (2016), *Sociologie du numérique*, Paris, Armand Colin, 352p.

BOULLIER, D., LOHARD, A. (2012), Opinion mining et Sentiment analysis, Marseille, Opendition press, Sciences Po-Médialab, 234p.

BOWKER, G.C. (2005), *Memory Practices in the Sciences*, Cambridge, Massachusetts, The MIT Press, 274p.

BOYADJIAN, J. (2016), *Analyser les opinions politiques sur Internet, Enjeux théoriques et défis méthodologiques*, Paris, Dalloz, Nouvelle Bibliothèque de Thèses, 382p.

BRICE, L., CROUTTE, P., JAUNEAU-COTTET, P., LAUTIE, S. (2015), Baromètre du numérique, Rapport du CREDOC, 170p.

CARDON, D., PRIEUR, C. (2016), Comment la conversation a façonné le Web, in : MARTIN, O., DAGIRAL, E. (ed.), L'ordinaire d'internet - Le web dans nos pratiques et relations sociales, Paris, Armand Colin, 288p.

CHAULET, J. (2009), Les usages adolescents des tic, entre autonomie et dépendance. *Empan*, vol.76 (4), 57-66, <DOI:10.3917/empa.076.0057>.

CROUTTE, P., LAUTIE, S., HOIBIAN, S. (2016), Le baromètre du numérique 2016. Rapport du CREDOC, 244 p.

DELBOSC, A., CURRIE, G. (2013), Causes of Youth Licensing Decline: A Synthesis of Evidence, *Transport Reviews*, vol.33 (3), 217-290.

FELT, M. (2016), Social media and the social sciences: How researchers employ Big Data analytics, *Big Data & Society*, vol.3 (1), 1-15.

FENNER, Y., GARLAND, S.M., MOORE, E.E., JAYASINGHE, Y., FLETCHER, A., TABRIZI, S.N., GUNASEKARAN, B., WARK, J.D. (2012), Web-Based Recruiting for Health Research Using a Social Networking Site: An Exploratory Study, *Journal of Medical Internet Research*, vol.14 (1), <DOI:10.2196/jmir.1978>.



- FLUCKIGER, C. (2010), Blogs et réseaux sociaux : outils de la construction identitaire adolescente ?, *Diversité : ville école intégration*, CNDP, 38-43, <hal-01373990>.
- GALESIC, M., BOSNJAK, M. (2009), Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey. *Public Opinion Quarterly*, vol. 73 (2), 349–360, <DOI:10.1093/poq/nfp031>.
- GITELMAN, L. (ed.) (2013), *Row data is an oxymoron*, Cambridge, Massachusetts, The MIT Press, 192p.
- JODELET, D. (2003), Représentations sociales : un domaine en expansion, in : JODELET, D. (ed.), *Les représentations sociales*, Paris, PUF, 7<sup>ème</sup> édition, 45-78.
- KAUFMANN, V. (2008), *Les paradoxes de la mobilité, bouger, s'enraciner*, PPUR (Presses polytechniques et universitaires romandes), Lausanne, 115 p.
- KAUFMANN, V., TABAKA, K., LOUVET, N., GUIDEZ, J.-M. (2010), Et si les Français n'avaient plus seulement une voiture dans la tête ? : Evolution de l'image des modes de transport. CERTU, Lyon.
- KOPNINA, H. (2011), Kids and cars: Environmental attitudes in children, *Transport Policy*, vol.18 (4), 573–578, <DOI:10.1016/j.tranpol.2011.01.013>.
- LATOURE, B. (1999), *Pandora's Hope, Essays on the Reality of Science Studies*, Harvard University Press, 336p.
- LIVINGSTONE, S., ÓLAFSSON, K., STAKSRUD, E. (2011), Social networking, age and privacy, 13p., <URL : [http://www.lse.ac.uk/media@lse/research/EUKidsOnline/EU%20Kids%20II%20\(2009-11\)/EUKidsOnlineIIReports/ShortSNS.pdf](http://www.lse.ac.uk/media@lse/research/EUKidsOnline/EU%20Kids%20II%20(2009-11)/EUKidsOnlineIIReports/ShortSNS.pdf) > (accessed 1.31.17).
- LUCAS, J.-F., BOULLIER, D. (2016), *Mobility: topic detection in social networks*, rapport de recherche du Social Media Lab de l'EPFL.
- MANOVICH, L. (2012), Trending: The Promises and the Challenges of Big Social Data, in : GOLD, M.K. (ed), *Debates in the Digital Humanities*, Minneapolis, The University of Minnesota Press, 460-475.
- MOREAU, A., ROUSTIT, O., CHAUCHARD, E., CHABROL, H. (2012), L'usage de Facebook et les enjeux de l'adolescence : une étude qualitative, *Neuropsychiatrie de l'Enfance et de l'Adolescence*, vol.60 (6), 429-434, <DOI:10.1016/j.neurenf.2012.05.530>.
- MORGAN, A.J., JORM, A.F., MACKINNON, A.J. (2013), Internet-Based Recruitment to a Depression Prevention Intervention: Lessons From the Mood Memos Study, *Journal of Medical Internet Research*, vol.15 (2), <DOI:10.2196/jmir.2262>.

MORRISSETTE, J., DESGAGNE, S. (2009), Le jeu des positions de savoir en recherche collaborative: une analyse, *Recherches qualitatives*, vol.28 (2), 118-144. Disponible en ligne <URL : [http://www.recherche-qualitative.qc.ca/documents/files/revue/edition\\_reguliere/numero28\(2\)/morrissette\(28\)2.pdf](http://www.recherche-qualitative.qc.ca/documents/files/revue/edition_reguliere/numero28(2)/morrissette(28)2.pdf)>

MOSCOVICI, S. (1961), *La psychanalyse, son image, son public, étude sur la représentation sociale de la psychanalyse*, Paris, PUF, 2<sup>ème</sup> éd, 651p.

MUSSO, P., COIFFIER S., LUCAS J.-F. (2014), *Innover avec et par les imaginaires*, Paris, Manucius, Modélisations des imaginaires, 50p.

PATRIARCHE, G., HUYNEN, P. (2014), Les usages des temps de déplacement par les Bruxellois: Une analyse des activités mobiles pratiquées vers le lieu de travail ou d'études, in : MACHARIS, C., DOBRUSZKES, F., HUBERT, M. (ed.), *Mobilité et logistique à Bruxelles*, Bruxelles, VUBPRESS, 77-107.

RAMO, D.E., PROCHASKA, J.J. (2012), Broad Reach and Targeted Recruitment Using Facebook for an Online Survey of Young Adult Substance Use, *Journal of Medical Internet Research*, vol.14(1), <DOI:10.2196/jmir.1878>.

ROGERS, R. (2013), *Digital methods*, Cambridge, The MIT Press, 280p.

ROSE, G., (2003), *Visual Methodologies: An Introduction to the Interpretation of Visual Materials*, London, Sage Publications Ltd., 287 p.

SCHONLAU, M., SOEST, A. van, KAPTEYN, A., COUPER, M. (2009), Selection Bias in Web Surveys and the Use of Propensity Scores, *Sociological Methods & Research*, vol.37 (3), 291-318, <DOI:10.1177/0049124108327128>.

SIGURDARDOTTIR, S.B., KAPLAN, S., MØLLER, M., TEASDALE, T.W. (2013), Understanding adolescents' intentions to commute by car or bicycle as adults, *Transportation Research Part D: Transport and Environment*, vol.24, 1-9, <DOI:10.1016/j.trd.2013.04.008>.

STAR S.L., GRIESEMER J. (1989), Institutionnal ecology, 'translations', and boundary objects: Amateurs and professionals on Berkeley's museum of vertebrate zoologie, *Social Studies of Science*, vol.19 (3), 387-420.

STRAUSS, A. (1988), The articulation of project work: an organizational process, *The Sociological Quarterly*, vol.29 (2), 163-178.

VINCENT-GESLIN, S., RAVALET, E., KAUFMANN, Vincent (2014), L'appropriation des temps et des espaces de la grande mobilité à l'ère numérique, *Géo-regards*, vol.7, 17-35.

WUNENBURGER, J.-J. (2016), Imagologie de la technosphère, in : MUSSO, P. (ed.), Imaginaire, industrie et innovation, Actes du colloque de Cerisy, Paris, Manucius, Modélisations des imaginaires, 16-24.