Des Migrants et des Mots

Une analyse numérique des débats médiatiques sur les migrations et l'environnement

Tommaso Venturini, François Gemenne, Marta Severo

> *Vous savez, il y a un débat ici au sujet des réfugiés. Laissez-moi vous dire quelle est ma position et celle des celle des personnes autour de cette table : les personnes dont nous parlons ne sont pas des réfugiés. Ce sont des Américains, et ils méritent l'aide, l'amour et la compassion de nos compatriotes*[1].
>
> - Président George W. Bush, à propos des populations déplacées par l'ouragan Katrina, 6 septembre 2005.

Résumé

Si les dégradations de l'environnement apparaissent aujourd'hui comme un facteur majeur de migrations, le débat concernant la définition du lien entre environnement et migration reste profondément ouvert. Un très grand nombre de termes occupent aujourd'hui l'espace public pour désigner ceux qui doivent se déplacer à cause de dégradations de leur environnement, générant ainsi une controverse sur l'usage des différents termes. Cet article essaie d'éclaircir cette controverse en étudiant les usages de ces termes dans le débat public sur Internet. Pour ce faire, nous avons employé une nouvelle méthode numérique qui nous a permis, à travers l'interrogation du moteur de recherche Google.com, de collecter les pages web où la discussion autour des différentes définitions de ces migrations était la plus visible. Grâce à une analyse des expressions contenues dans ces pages, nous avons obtenu des cartes sémantiques qui ont nous permis de voir quels termes étaient associés les uns aux autres : en particulier, à quels acteurs, lieux ou concepts les différents termes étaient les plus connectés. Ces cartes révèlent la réalité polymorphe de ces migrations, mais aussi le vide catégoriel qui les entoure.

Abstract

Though environmental degradations appear today as a major driver of migration flows, the debate about the definition of the relationship between environment and migration remains vivid. In public debates, many terms exist to qualify those who have to migrate because of environmental disruptions: this has lead to controversies surrounding the use of these different terms. This article attempts to document this controversy through a study of the different uses of these terms in public debates online. In order to do so, we have used a new digital method that has allowed us, thanks to a search on Google.com, to collect webpages where the discussion on the different definitions of these migrations was most visible. Building on an analysis of the expressions contained in these webpages, we have generated semantic maps that show which terms are associated with each other: in particular, which actors, places and concepts were connected to the most terms. These maps reflect the polymorphic reality of these migrations, but also the categorical void that surrounds them.

---

[1] « You know, there's a debate here about refugees. Let me tell you my attitude and the attitude of people around this table: The people we're talking about are not refugees. They are Americans, and they need the help and love and compassion of our fellow citizens. » (traduction des auteurs)

Mots-clés
Migrations, Environnement, Terminologie, Internet, Catégories.

Keywords
Migration, Environment, Terminology, Internet, Categories.

Biographies
Tommaso Venturini est chercheur et professeur associé à Sciences Po où il coordonne les activités de recherche du médialab. Ses recherches se concentrent sur les méthodes numériques et quali-quantitatives pour les sciences sociales. Il enseignes cours et coordonne plusieurs projets sur la cartographie des controverses.

François Gemenne est chercheur en science politique (Iddri – Sciences Po / CEDEM – ULg). Ses recherches sont principalement consacrées aux migrations et aux déplacements de populations liés aux changements de l'environnement, ainsi qu'aux politiques d'adaptation au changement climatique. Il enseigne également ces questions à Sciences Po, à l'Université de Paris 13 et à l'Université Libre de Bruxelles.

Marta Severo Marta Severo est maître de conférences en Communication auprès du laboratoire Geriico de l'Université de Lille 3. Elle travaille sur les thèmes de la communication et de la gestion du patrimoine culturel et notamment sur la contribution des technologies numériques. En 2009, elle a publié l'œuvre *Heritage Networks. Managing Network Cultural Heritage with the Web*.

1. L'intérêt d'une controverse

L'ouragan Katrina, qui ravagea les côtes du Golfe du Mexique en août 2005, provoqua l'évacuation de plus d'un million de personnes. Certains trouvèrent refuge chez des proches, d'autres dans des centres d'accueil mis en place par les autorités ou des organisations caritatives. Très rapidement, les médias allaient jouer un rôle central dans la caractérisation des victimes de la catastrophe. Immédiatement après la catastrophe, le terme qui était alors le

plus fréquemment utilisé pour désigner les évacués était celui de 'réfugiés de Katrina'[2]. Le terme, néanmoins, fut immédiatement rejeté par ceux qu'il prétendait désigner. Ceux-ci insistèrent pour être appelés 'évacués', ou 'survivants', mais surtout pas 'réfugiés'. Pour beaucoup, être appelé(e) 'réfugié' revenait à être considéré comme un citoyen de seconde zone, un étranger dans son propre pays[3]. Refuser cette étiquette était aussi une manière de rappeler à leur propre Etat ses obligations envers ses citoyens. L'ancien maire de La Nouvelle-Orléans, Marc Morial, proposa l'emploi du terme 'citoyens réfugiés', sans guère de succès. Après que les victimes se furent indignées contre l'emploi du terme de 'réfugiés', la plupart des médias cessèrent de l'utiliser et le remplacèrent par d'autres termes, notamment celui d''évacués'. Mais d'autres médias, et notamment CNN, le New York Times et l'Associated Press, persistèrent à l'utiliser. Le 7 septembre 2005, un des présentateurs-vedettes de CNN, Lou Dobbs, justifiait ainsi à l'antenne l'emploi du terme :

> *Au fait, vous avez entendu dans cette émission plusieurs personnes, dont le Révérend Jesse Jackson, nous intimer de ne pas utiliser le terme de 'réfugiés' pour décrire les citoyens de La Nouvelle-Orléans qui avaient fui leur maison. Jackson et d'autres, notamment le Président Bush, ont dit ou sous-entendu que ce terme était racialement connoté. Le dictionnaire Miriam Webster définit un réfugié comme quelqu'un sui fuit. (...) Le Président, Jackson et d'autres ont l'air de penser que les médias ont créé le terme de réfugié dans le seul but de décrire les victimes de l'ouragan Katrina. C'est n'importe quoi. Même une revue rapide de la couverture médiatique de catastrophes comme l'ouragan Andrew, des inondations dans le Midwest en 1993 ou des feux de forêts dans l'Ouest fera apparaître l'usage du termes 'réfugié' par les médias. Je suis fier de vous dire que cette chaîne a résisté aux appels de ceux qui lui disaient quels mots employer. A même rejeté, en fait, la suggestion des Nations Unies que nous utilisions, plutôt que 'réfugiés', l'expression 'déplacés internes'. J'adore celle-là. Nous continuerons à utiliser le terme dans cette émission quand nous pensons que c'est le plus approprié*[4].

---

[2] Sommers, S. R., E. P. Apfelbaum, K. N. Dukes, N. Toosi, et E. F. Wang. "Race and Media Coverage of Hurricane Katrina: Analysis, Implications, and Future Research Questions." *Analyses of Social Issues and Public Policy* 6:39-55, 2006, p.40.

[3] Gemenne, F. "What's in a name: social vulnerabilities and the refugee controversy in the wake of hurricane Katrina." In Jäger, J. et Afifi, T. (Ed.) *Environment, Forced Migration and Social Vulnerability*, Berlin, Springer, 2010.

[4] « You've heard on this broadcast, by the way, several people, including Reverend Jesse Jackson and others admonish us not to use the term refugee when describing the New Orleans citizens who have had to flee their

Outre la colère des victimes, d'autres raisons furent avancées pour rejeter l'emploi du terme 'réfugiés' : d'aucuns estimaient que le terme retirait aux victimes leur dignité[5], tandis que d'autres, tels les Révérends Jackson et Sharpton, voyaient dans le terme une preuve évidente de discrimination raciale. Le linguiste Geoffrey Nunberg, dans une étude postée sur son site personnel, montre ainsi que le terme 'réfugié' était deux fois plus employé que le terme 'évacué' quand il était employé en conjonction avec les termes 'noir' ou 'pauvre'[6].

Pourquoi faire état ici de cette controverse ? D'abord parce qu'elle illustre la catastrophe sociale qu'a été Katrina, et les profondes inégalités sociales que l'ouragan a mises au jour[7]. Ensuite parce qu'elle révèle un vide catégoriel. Aucun terme, aujourd'hui, ne fait l'unanimité pour désigner ceux qui sont déplacés en raison de dégradations de leur environnement. Et s'il n'existe aucun terme consensuel, c'est avant tout parce que la réalité des migrations liées à l'environnement est polymorphe : certain(e)s sont forcé(e)s de quitter leur domicile à la suite de catastrophes naturelles brutales, tandis que d'autres font le choix de partir, à la recherche d'un environnement plus clément. Certain(e)s partent très loin, d'autres ne peuvent se déplacer qu'à quelques kilomètres. Certain(e)s peuvent revenir chez eux après quelques semaines ou quelques années, tandis que d'autres ne reviennent jamais. La relation entre changements environnementaux et mouvements migratoires est à ce point complexe et multiforme qu'il est sans doute impossible de trouver un seul terme qui en fasse la synthèse, et qui soit à même de refléter la diversité des vécus des migrants[8].

Pourtant les termes que l'on emploie pour désigner cette réalité comptent, et la controverse de

---

homes. Jackson and others including President Bush have said or implied that term is racially insensitive. In my opinion, straightforwardly, Reverend Jackson and President Bush are not entirely correct. The Miriam Webster Dictionary defines refugee as one who flees. (…) The president, Jackson and others apparently think that news organizations created the term refugee just to describe victims of Hurricane Katrina. Hardly. Even a cursory review of reporting of such disaster of Hurricane Andrew, the 1993 midwestern floods and wildfires through the West have all prompted the use of the term refugee by news organizations. I'm proud to tell you that this network has resisted others telling them how to use words. Rejecting, in fact, the United Nations suggestion that we use, instead of refugee, the expression 'internally displaced persons'. I love that one. We'll continue here to use the term on this broadcast where we think it is most descriptive » (traduction des auteurs). Dans « Lou Dobbs tonight », CNN, 7 septembre 2005.

[5] Masquelier, A. « Why Katrina's Victims Aren't Refugees: Musings on a "Dirty" Word », *American Anthropologist* 108 (4): 735-743, 2006.

[6] Nunberg, G. "When Words Break Down.", 2005. http://people.ischool.berkeley.edu/~nunberg/looting.html (consulté le 20 décembre 2012).

[7] Huret, R. *Katrina, 2005. L'ouragan, l'Etat et les pauvres aux Etats-Unis*. Paris, Editions de l'EHESS, 2010.

[8] Piguet E., Pécoud A., et de Guchteneire P. « Migration and Climate Change: An Overview », *Refugee Survey Quarterly* 30(3): 1-23, 2011.

Katrina est là pour nous le rappeler. Depuis les travaux de Becker[9] et Goffman[10] sur la théorie de l'étiquetage, les sciences sociales savent bien que les termes employés impliquent une catégorisation, et cette catégorisation forge l'identité de ceux qu'elle désigne. Alors que de nombreux travaux existent sur les conséquences de la catégorisation des réfugiés sur leur identité personnelle et les réponses politiques qui sont déployées[11], peu de travaux similaires existent pour la relation environnement-migration, malgré la grande variété de vocables utilisés pour désigner un même phénomène. Pourtant, les mots comptent : à chaque terme est associée une valeur symbolique, qui peut à la fois affecter la vulnérabilité de ceux qu'il désigne, mais également déterminer les réponses politiques qui seront mises en œuvre pour encadrer ces migrations. Ainsi, plusieurs études ont montré combien ces questions identitaires étaient déterminantes dans la capacité d'adaptation de ceux qui étaient touchés par des changements environnementaux[12]. Mais rares sont les travaux de réflexion et d'analyse sur les termes employés pour désigner ceux qui migraient pour des raisons environnementales[13]. Seuls les termes de 'réfugié climatique' ou de 'réfugié environnemental' se sont attirés les foudres des juristes, ceux-ci relevant généralement que le terme de 'réfugié' était strictement défini par la Convention de Genève de 1951, dont ne relevaient pas ceux qui migraient pour des raisons environnementales[14].

Dans la communauté académique, les débats sur les termes à employer sont vifs, et recoupent largement les débats qui touchent à la définition des migrations liées à l'environnement, et sur la protection à accorder aux migrants. Ainsi, il est possible d'identifier plusieurs tendances dans la littérature. En premier lieu, certains auteurs et organisations militent pour une reconnaissance officielle du statut de « réfugié environnemental » ou de « réfugié climatique » dans le droit international, en soulignant le vide juridique qui entoure leur protection. Logiquement, ces auteurs emploieront volontiers le terme de « réfugié », précisément pour insister sur l'absence de critères environnementaux dans la définition

---

[9] Becker, H. *Outsiders : Studies in the Sociology of Deviance*. New York, Free Press, 1963.

[10] Goffman, E. *Stigma: Notes on the Management of Spoiled Identity*. Englewood Cliffs, N.J., Prentice-Hall, 1963.

[11] Zetter, R. "Labelling Refugees: Forming and Transforming a Bureaucratic Identity." *Journal of Refugee Studies* 4:39-62, 1991.

[12] Farbotko, C. "Tuvalu and climate change: Constructions of environmental displacement in the Sydney Morning Herald." *Geografiska Annaler* 87 B:279-293, 2005.

[13] Dun, O. et F. Gemenne, « Defining 'environmental migration », *Forced Migration Review* 31: 10-11, 2008

[14] McAdam, J. "Swimming against the Tide: Why a Climate Change Displacement Treaty is Not *the* Answer." *International Journal of Refugee Law* 23:2-27, 2011 ; Cournil, C. "Vers une reconnaissance des "réfugiés écologiques"? Quelle(s) protection(s), quel(s) statut(s)?" *Revue du Droit Public et de la Science Politique* 4:1035-1066, 2006.

officielle du concept de « réfugié » de la Convention de Genève de 1951[15]. La plupart de ces auteurs souhaitent dès lors soit que la Convention de Genève soit révisée pour inclure des critères environnementaux, soit qu'une nouvelle convention soit adoptée pour garantir à ces personnes une protection dans le droit international.

A l'inverse, d'autres auteurs estiment que l'emploi du terme de « réfugiés » est particulièrement préjudiciable, et qu'une protection spécifique dans le droit international ne serait pas appropriée, voire contre-productive. Logiquement, ces auteurs condamnent l'emploi du terme de « réfugiés », et lui préféreront plus volontiers le terme de « déplacés »[16]. Certains auteurs contestent également l'idée qu'il soit possible d'identifier une catégorie spécifique de migrants dont la migration soit liée à l'environnement, et insistent plutôt sur la multi-causalité des migrations, et l'impossibilité d'isoler les facteurs environnementaux parmi d'autres facteurs de migration. Ces auteurs éviteront dès lors l'emploi d'un terme spécifique pour désigner ces migrants[17].

La controverse dépasse largement la sphère académique, puisqu'elle touche à la protection à accorder aux migrants, et donc aussi aux mandats des différentes agences qui pourraient fournir cette protection. A cet égard, il est particulièrement intéressant d'observer le positionnement du Haut Commissariat des Nations Unies pour les Réfugiés (HCR) et de l'Organisation Internationale pour les Migrations (OIM), qui sont les deux principales organisations en charge de ces migrants sur le terrain, sans que cette mission fasse explicitement partie de leur mandat. Soucieux de ne pas introduire de confusion avec son mandat de protection des réfugiés, le HCR préférera ainsi généralement parler de 'personnes déplacées pour des raisons environnementales'[18], soulignant ainsi la caractère forcé de leur

---

[15] Westing, A. H. « Environmental refugees: A growing category of displaced persons. » *Environmental Conservation* 19 (3): 201-207, 1992 ; McCue, G. S. « Environmental Refugees: Applying International Environmental Law to Involuntary Migration. » *Georgetown International Environmental Law Review* 6 (1): 151-190, 1993 ; Myers, N. « Environmental refugees: a growing phenomenon of the 21st century. » *Philosophical Transactions of the Royal Society B* 357 (1420): 609-613, 2002 ; Conisbee, M. et A. Simms. *Environmental Refugees. The case for Recognition*. London, New Economics Foundation, 2003 ; Williams, A. « Turning the Tide: Recognizing Climate Change Refugees in International Law », *Law & Policy* 30 (4): 502-529, 2008.

[16] Kälin, W. "Conceptualising Climate-Induced Displacement." In McAdam, J. (Ed.) *Climate Change and Displacement. Multidisciplinary Perspectives,* Oxford, Hart, pp. 81-104, 2010 ; McAdam, J. "Swimming against the Tide: Why a Climate Change Displacement Treaty is Not *the* Answer." *International Journal of Refugee Law* 23:2-27, 2011.

[17] Kibreab, G. « Environmental Causes and Impact of Refugee Movements: A Critique of the Current Debat » *Disasters* 21 (1): 20-38, 1997 ; Jonsson, G. *The environmental factor in migration dynamics – a review of African case studies*. Oxford, International Migration Institute, 21, 2010.

[18] Gorlick, B. "Environmentally Displaced Persons: A UNHCR Perspective." In *'Environmental Refugees: the*

migration – et donc la légitimité du HCR à se saisir de la question – tout en veillant à les distinguer clairement des réfugiés. A l'inverse, l'OIM emploiera plus volontiers le terme de 'migrants environnementaux'[19], dont elle proposera une définition – manière de souligner que l'organisation se trouve bien dans le cadre de son mandat général de protection des migrants.

La controverse sur les termes à employer n'est donc pas simplement affaire de vocabulaire : elle touche à la conceptualisation même des migrations liées à l'environnement, et à la protection de ces migrants, ainsi à la politique qui leur sera appliquée quant à leurs droits fondamentaux ou au franchissement de frontières. L'emploi d'un terme plutôt qu'un autre révélera non seulement une manière différente d'appréhender et de conceptualiser ces mouvements de populations, mais également une indication quant à la manière la plus adéquate d'assurer leur protection. A cet égard, une fracture importante se situe autour de l'emploi du terme de « réfugié », car cette fracture traduit aussi, généralement, un positionnement quant à la nécessité de créer dans le droit international un statut de « réfugié environnemental » ou de « réfugié climatique ».

Cet article est l'une des premières tentatives visant à analyser la production de discours sur la relation entre environnement et migrations au travers des termes employés pour désigner les migrants. Jusqu'ici, cette discussion était restée confinée aux débats sur la recherche d'une définition consensuelle, sans que la question de l'usage de ces termes ne soit véritablement abordée. Or, au cours des dernières années, les liens entre changements environnementaux et migrations ont fait l'objet de mobilisations importantes, et ont logiquement généré de nombreux discours médiatiques. Les liens entre changement climatique et migrations sont de plus en plus débattus, non seulement dans les contextes institutionnels de la recherche et des organisations internationales, comme on vient de le voir, mais également dans les discours des citoyens ordinaires. Ce sont ces discours qui constituent l'objet de cet article, et qui sont analysés à l'aide d'une méthode innovante fondée sur une cartographie d'internet[20].

Forgotten Migrants', Discussion Event. New York, 2007.

[19] Laczko, F. et C. Aghazarm. *Migration, Environment and Climate Change: Assessing the Evidence*. Geneva, International Organization for Migration, 2009.

[20] Cet article constitue un premier test des méthodes de recherche d'un plus large projet dédié à la cartographie des controverses sur le changement climatique. Financé par le 7ème programme cadre de l'Union Européenne, le projet EMAPS (Grant Agreement numéro 288964) vise à analyser et à représenter le débat autour des stratégies d'adaptation au réchauffement global (voir http://www.emapsproject.com/ pour plus d'informations).

Se déployant sous l'impulsion d'un nombre croissant d'acteurs et sous l'élan d'injonctions fortes et souvent opposées, la question du lien entre migrations et environnement se présente volontiers comme un 'moteur à controverses'. D'un côté, le phénomène des migrations prend de plus en plus d'ampleur, accru par le fossé grandissant entre pays riches et pays pauvres, et facilité par un meilleur accès aux moyens de transport international. Ce phénomène suscite des tensions toujours plus vives, qui opposent le désir et le droit à la mobilité de millions de personnes aux exigences de contrôle des pays d'origine et de destination. Par ailleurs, depuis quelques années, le changement climatique et ses conséquences ont acquis une beaucoup plus grande visibilité dans le débat international. Avec le changement climatique, ce n'est pas une technologie ou une institution particulière qui est remise en discussion, mais l'organisation même de notre vie collective, des moindres aspects de la vie quotidienne jusqu'aux relations internationales.

Le débat académique et institutionnel, loin d'avoir tranché la question de la définition et de la terminologie, emploie une très grande variété de vocables différents pour caractériser la relation entre environnement et migrations, tandis que les organisations de la société civile utilisent volontiers des images et notions à fort impact émotionnel pour nourrir l'imaginaire du public.

Les contours encore flous de la relation entre environnement et migrations doivent-ils nous conduire à ignorer le débat public sur ce sujet, sous prétexte qu'il s'agirait d'un débat stérile et abstrait, du fait de l'impossibilité d'isoler les causes environnementales des migrations?

Au contraire, précisément à cause de sa nature de 'moteur à controverses', la question du lien entre migrations et environnement nous semble être un sujet extrêmement intéressant. Comme toutes les controverses bien vivantes, ce débat permet d'expliciter et donc d'analyser les lignes de fracture traversant notre vie collective[21].

Ce débat complexe se cristallise en partie autour de la question de l'étiquette utilisée pour les personnes déplacées par des changements environnementaux. La question est loin d'être purement formelle, puisqu'elle emporte des conséquences dans le droit international, dans les réponses politiques à mettre en œuvre, mais aussi, plus fondamentalement, dans la manière dont les migrants se perçoivent et dont nous les percevons. En caricaturant à peine, si on les

[21] Venturini, T. « Diving in Magma: How to Explore Controversies with Actor-Network Theory » *Public Understanding of Science*, 20:4, 2010a ; Venturini, T. « Building on Faults: How to Represent Controversies with Digital Méthodes », *Public Understanding of Science,* 2012a, à paraître.

appelle ‘réfugiés’, cela suppose que leur migration est forcée. Si on les appelle ‘migrants’, c’est qu’au contraire leur migration est choisie. Si on les appelle ‘refugiés’, on reconnaît implicitement le besoin d’assistance de ces populations. Si on les appelle ‘migrants’, on a volontiers tendance à les considérer eux-mêmes comme un problème. Si on les appelle ‘climatiques’, on attribue directement la cause de leur migration au réchauffement global. Si on les appelle ‘environnementaux’, la cause de leur déplacement apparaît tout de suite plus vague et incertaine. Et, bien sûr, les acteurs n’hésitent pas à mélanger substantifs et adjectifs, pour obtenir l’exacte nuance qui leur convient.

Dans la discussion sur les étiquettes se trouve au fond condensée toute la controverse sur les migrations liées à l’environnement, et avec elle beaucoup des débats sur les conséquences de la crise écologique.

3. Méthode

Dans cet article, nous ne nous intéressons pas directement au débat académique et institutionnel sur le sujet, mais uniquement à ses manifestations médiatiques. Et même, plus précisément, au vocabulaire qui est employé dans ces manifestations médiatiques. Comment ces débats ont-ils investi les sphères de l’opinion publique ? Question difficile, car l’opinion publique est par définition fuyante et hétérogène[22]. Elle ne s’exprime pas dans une arène prédéfinie, elle ne suit aucune cohérence formelle, elle est constituée par un flux de voix diverses, impossible à prévoir et constamment changeant. L’opinion publique est bien sûr composée par les chercheurs, les institutions, les médias, les organisations, mais elle accueille aussi une multitude d’acteurs inattendus (activistes, bloggeurs, savants amateurs, citoyens…) qu’on ne peut ignorer. Comment étudier une telle cacophonie, et comment l’appréhender ?

Internet nous aide dans ce travail car il offre un espace assez neutre pour que toutes ces voix s’y expriment, mais suffisamment stable, techniquement, pour qu’on puisse toutes les enregistrer. Mais le Web est vaste et mal indexé : comment y retrouver les discours qu’on y cherche ? Demandez à n’importe quel expert en *information mining* comment s’y prendre pour trouver l’aiguille dans la botte de foin : il vous répondra qu’il faut surtout avoir la bonne *query*, la bonne requête, la combinaison magique qui permet d’identifier les informations qu’on cherche et de filtrer le bruit qui les entoure. Or, la discussion sur le vocabulaire des migrations liées à l’environnement nous offre exactement cela : une bonne requête ou – encore mieux – une série de bonnes requêtes capables d’identifier les différents discours sur

[22] Lippmann, W. *The Phantom Public*. New York, The Macmillan Company, 1927.

ces migrations.

Les cartes que nous allons discuter dans les prochaines pages n'ont pas été produites de manière automatique. Trop souvent, lorsqu'on parle de *méthodes numériques* pour les sciences sociales[23] ou de *computational social sciences*[24], on imagine des algorithmes fantastiques qui permettraient d'appuyer sur un bouton et d'obtenir automatiquement un résultat. La réalité est tout autre. Les méthodes numériques demandent une intervention continue du chercheur, et sont jalonnées par des centaines de choix subjectifs qui ne peuvent s'appuyer que sur l'expertise de l'analyste[25].

Dans les paragraphes suivants, nous procéderons à une description très détaillée des passages qui nous ont permis d'arriver aux cartes que nous allons analyser. Avant de nous lancer dans les détails méthodologiques, il est toutefois important de donner au lecteur une idée d'ensemble de l'approche employée dans cet article. Au delà de ses technicités, notre démarche est relativement linéaire. Il s'agit de s'appuyer sur Google (aujourd'hui le plus influent des moteurs de recherche) pour collecter un ensemble de pages web utilisant les diverses étiquette associées aux migrations climatiques. Sur la base de ces *corpus*, nous avons ensuite analysé quelles expressions lexicales apparaissaient le plus fréquemment dans les pages associées à l'une ou à l'autre étiquette. Nous avons enfin visualisé les relations entre étiquettes et expressions par un graphe spatialisé afin de rapprocher les mots souvent utilisés ensemble. L'objective de la méthode est simple : il s'agit d'identifier des univers de discours liés aux différentes façons de désigner les personnes se déplaçant pour des raisons environnementales.
Le lecteur nous pardonnera d'aborder ici des détails techniques et quelque peu fastidieux : il n'y a pas d'autre moyen d'exposer la validité de notre démarche.

Le point de départ est l'interrogation du moteur de recherche Google.com (dans la version la

---

[23] Rogers, R. *Information Politics on the Web*. Cambridge, Mass, MIT Press, 2004 et Rogers, R. *The End of the Virtual: Digital Methods*. Amsterdam, Amsterdam University Press, 2009.

[24] Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., Christakis, N., et al. « Computational social science » *Science*, 323(5915), 721-3, 2009.

[25] Venturini, T. « Great Expectations : méthodes quali-quantitative et analyse des réseaux sociaux ». In Fourmentraux J.P. (ed.) *L'Ere Post-Média*, Paris, Hermann, 2012b.

plus 'standard' possible[26]) avec les requêtes présentées dans le tableau suivant :

Requête	Nombre de pages moissonnées
"Climate refugee"	84 pages
"Environmental migrant"	76 pages
"Climate migrant"	69 pages
"Environmental Refugee"	72 pages
"Eco-migrant"	80 pages
"Ecological migrant"	64 pages
"Ecological refugee"	65 pages
"Climate-induced migrant"	12 pages
"Climate-induced refugee"	16 pages
"Environmentally-induced migrant"	62 pages
"Environmental displacee"	86 pages
"Environmentally displaced person"	48 pages
"Eco-refugee"	79 pages
"Ecologically displaced person"	90 pages

Tableau 1. Résultats considérés pour chaque requête

Pour chacune des requêtes nous avons visité les 100 premiers résultats (ou moins, si la requête en produisait moins) et, quand il s'agissait de pages web, nous avons moissonné leur texte[27]. Le corpus ainsi constitué est composé du texte de 903 pages. La constitution du corpus a été réalisée le 9 février 2012. Il est important d'indiquer la date, car notre analyse reflète la situation du débat à cette date. Contrairement à ce qu'on croit souvent, le web a très peu de mémoire. Il est donc très difficile de réaliser des analyses diachroniques du débat sur le Web à moins de disposer d'un corpus de pages web archivé dans le temps[28].

Nous avons ensuite employé le moteur d'extraction sémantique Alchemy (http://www.alchemyapi.com/) pour extraire les *ngrams* et les *entités nommées* présents dans chacun des 903 textes de notre corpus. Par *ngram*, on entend une expression de *n* mots apparaissant plusieurs fois dans le même texte (par exemple *anthropogenic climate change*). Par *entités nommées*, on entend un *ngram* qui est reconnu par Alchemy comme faisant

[26] Il est connu que le moteur de recherche Google.com 'personnalise' les résultats des recherches qui lui adressés en se fondant sur l'historique de recherche et de navigation, le profil Google de l'utilisateur, le lieu d'où provient la requête, la langue du navigateur et d'autres variables qui ne sont pas complètement connues. Par conséquent, deux utilisateurs posant à Google la même requête peuvent très bien obtenir des réponses différentes. Nous avons essayé de fournir à Google le moins d'informations possibles sur notre identité afin d'obtenir les résultats les moins personnalisés possible. Mais en toute occurrence, étant donné que la personnalisation concerne surtout l'ordre des résultats et que nous avons visité les 100 premiers résultats pour chaque requête, nous sommes confiants quant au fait que la personnalisation de Google ne devrait pas avoir trop biaisé les résultats obtenus.

[27] Pour des raisons techniques, nous n'avons pas pu récupérer le texte correspondant aux résultats qui n'étaient pas des pages html (vidéos, animations flash, documents PDF, présentations PowerPoint, documents DOC…). Ceci explique pourquoi nous n'avons pu récupérer la totalité de 100 premiers résultats pour aucune des requêtes.

[28] Le médialab de Sciences Po travaille à la réalisation d'un outil capable de permettre cet archivage, mais à ce jour cet type d'analyse n'est pas encore possible.

référence à une personne (par exemple Richard Black), à une organisation (par exemple l'OIM) ou à un lieu géographique (par exemple Tuvalu). Par la suite, nous désignerons *ngrams* et *entités nommées* par le terme générique d' 'expressions'.

Une précaution méthodologique s'impose. Tout à l'heure, nous avons affirmé que l'objectif de notre analyse était de dégager les 'univers des discours' autour des différentes étiquettes des migrations liées à l'environnement. Le choix de nous concentrer sur les expressions *extraites des pages web* comporte cependant la réserve que l'analyse que nous avons menée n'est pas une analyse de discours au sens strict du terme. Une analyse de ce type aurait demandé une lecture méticuleuse, difficile à réaliser pour un *corpus* aussi vaste que celui que nous avons collecté. Afin de garder une large couverture des débats en ligne, nous avons donc dû nous contenter d'une analyse plus légère et nous limiter à compter l'occurrence d'expressions lexicales dans les pages de notre *corpus*. Nous sommes bien conscients que la façon d'utiliser une certaine expression, le contexte rhétorique et pragmatique sont souvent plus importants que le contenu de l'expression elle-même.

Pour contrebalancer cette difficulté, deux remèdes ont été mis en place. En premier lieu, l'interprétation des résultats de notre analyse, loin de s'appuyer exclusivement sur les cartes que nous avons produites, se fonde également sur une connaissance experte du débat. Deuxièmement, nous avons essayé de récolter et d'analyser un très grand nombre d'expressions afin de reconstruire en partie le contexte d'usage.

Le traitement par Alchemy des 903 textes du corpus a restitué plus de 50 000 expressions[29] (52.542 exactement) : un nombre d'expressions beaucoup trop grand pour pouvoir les analyser et les interpréter. Un travail d'analyse préliminaire a donc été nécessaire pour réduire les expressions à une quantité plus raisonnable. Pour cela, nous avons d'abord appliqué un premier tri automatique en éliminant toutes les expressions les moins fréquentes. En particulier, nous avons éliminé toutes les expressions apparaissant dans moins de trois des textes du corpus et toutes les expressions moins fréquentes. Le tri automatique a permis d'exclure 90,5% des expressions trouvées par Alchemy, en réduisant le nombre d'expressions à 4 995. Nous avons ensuite trié manuellement les expressions restantes, en éliminant les expressions trop génériques et ambiguës (par exemple *major issue*, ou *significant consequences*) et en fusionnant les synonymes (par exemple *United Kingdom* et *UK*). Cela

[29] Dans le paramétrage de l'analyse de *Alchemy,* nous avons choisi de n'extraire que les *n-grams* de langue anglaise. Malgré l'évident intérêt que cela aurait eu, une analyse multilingue aurait posé l'énorme problème de l'équivalence des expressions dans les différentes langues. On sait bien, par exemple, que 'développement durable' et 'sustainable development' ne désignent pas tout à fait la même chose. C'est pourquoi il nous a semblé plus prudent de nous limiter à une seule langue : nous avons choisi l'anglais comme langue du débat international.

nous a permis d'obtenir un *corpus* de 609 expressions. Nous avons ensuite manuellement corrigé la classification des entités nommées de Alchemy, en identifiant tous les acteurs (individuels ou collectifs) et les lieux (continents, pays, régions, villes…) présents dans notre liste finale d'expressions.

Nous avons ensuite exporté un réseau non-pondéré de documents et d'expressions (voir Fig. 1). Les nœuds de ce réseau sont les 903 textes et les 609 expressions que nous avons retenus dans notre *corpus* final. Le réseau est *biparti,* car ses arcs ne connectent jamais deux nœuds du même type (deux textes ou deux expressions), mais toujours des nœuds de type différent : chaque expression est connectée à tous les textes où elle apparaît, et chaque texte est connecté à toutes les expressions qu'il contient.

Toutes les phases décrites jusqu'ici ont été réalisées à l'aide de ANTA, un logiciel développé par le médialab de Sciences Po[30].Toutes les phases suivantes et impliquant l'analyse du réseau des documents et entités ont été réalisées à l'aide de GEPHI (gephi.org), un logiciel d'analyse de réseaux co-développé par le médialab de Sciences Po[31].

Le choix d'une analyse et d'une représentation par les réseaux mérite d'être brièvement discuté, d'autant plus que les donnés analysées ne sont pas de nature intrinsèquement relationnelle. Pour réaliser nos cartes, nous avons en effet choisi d'ignorer le liens explicites présents sur le pages web constituant notre *corpus*, pour nous concentrer sur les liens indirects de co-occurrence d'expressions. Cette information aurait pu très bien se prêter à une analyse matricielle : les même données pourraient être présentées dans un tableau ayant en colonnes les documents (ou les requêtes) et en lignes les expressions sélectionnées. Ce tableau aurait pu être analysé avec des méthodes statistiques et visualisé par des diagrammes classiques (histogrammes, camemberts, etc.). Si nous avons décidé d'utiliser une analyse de réseaux, c'est parce que ce type d'analyse nous semble particulièrement adapté aux objectifs d'exploration de cet article. S'agissant de la première tentative d'analyse à large échelle du débat en ligne sur les migrations environnementales, notre analyse ne peut être qu'exploratoire[32]. Dans cet article, nous n'avons pas des hypothèses précises à tester. Notre objectif est plutôt de défricher les grandes tendances du débat en ligne, en préparant le terrain pour des études plus approfondies. Grâce à leur capacité à faciliter à la fois l'observation des

[30] Voir Venturini, T., Guido, D. « Once Upon a Text: An ANT Tale in Text Analysis », *Sociologica,* à paraître, 2012.

[31] Les deux outils sont diffusés avec une licence *open-source* et leur code est disponible ici : http://github.com/medialab.

[32] Tukey, J. W. *Exploratory Data Analysis*. Reading, Addison-Wesley, 1977.

données individuelles (chaque nœud dans le réseau reste visible) et des caractéristiques globales (densité, *clustering*, centralité…), les réseaux nous semblent l'outil idéal pour ce travail d'exploration[33], en particulier s'il sont spatialisés avec un algorithme force-vecteur.

Figure 1. Le réseau biparti des expressions (en gris foncé) et des documents (en gris clair).

Le réseau de la Figure 1, ainsi que les réseaux suivants, sont spatialisés à l'aide d'un algorithme force-vecteur (en l'occurrence ForceAtlas2 LinLog mode[34]). Cette étape est très importante : elle consiste à positionner les nœuds du graphe en utilisant un algorithme qui simule un système de forces physiques. Dans ce système, les nœuds exercent une force de répulsion les uns sur les autres, tandis que les arcs fonctionnent comme des élastiques, gardant proches les nœuds qu'ils connectent. Une fois que l'algorithme trouve sa position d'équilibre, la position relative des nœuds devient significative. Si, par exemple, deux documents se retrouvent proches dans le graphe, c'est parce qu'ils partagent plusieurs expressions. Ou encore, si une expression est liée à deux documents seulement, elle va occuper une position intermédiaire entre les deux.

Le réseau présenté en Figure 1 représente les 'données brutes' de notre analyse, mais il est beaucoup trop complexe pour pouvoir être interprété. Afin de rendre le réseau lisible, nous lui

[33] Venturini, T. (2012b) « Great Expectations : méthodes quali-quantitative et analyse des réseaux sociaux ». In Fourmentraux J.P. (Ed.) *L'Ere Post-Média*, Paris, Hermann, 2012b.

[34] Jacomy, M., Heymann, S., Venturini, T., & Bastian, M. ForceAtlas2, a graph layout algorithm for handy network visualization, 2011 : http://www.medialab.sciences-po.fr/fr/publications-fr/, consulté le 20 décembre 2012.

avons fait subir plusieurs transformations graphiques et mathématiques. En premier lieu, nous avons décidé d'éliminer les documents liés aux requêtes *Climate-induced migrant* (12 documents) et *Climate-induced refugee* (16 documents). Ayant généré beaucoup moins de documents que les autres, leur analyse risquait en effet d'être biaisée par l'insuffisance de l'échantillon de textes considérés. Nous avons ensuite groupé les 875 documents dans 12 groupes correspondant aux requêtes qui avaient permis de les trouver. Une conséquence importante de cette action de regroupement est la pondération des liens du graphe[35]. Le résultat est visible en Figure 2.

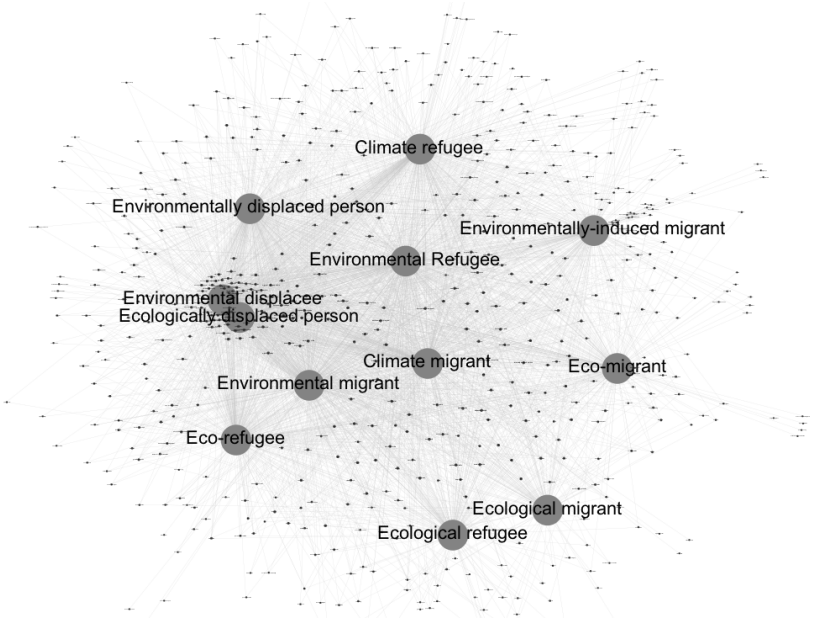

Figure 2. Le réseau des expressions et de requêtes

Tout en étant beaucoup plus lisible que la figure précédente (il est maintenant possible de lire les noms des requêtes, et donc de voir quelles requêtes sont proches parce qu'elles partagent le même contexte lexical), la Figure 2 reste encore difficile à interpréter. En particulier, le nombre d'expressions est encore trop grand pour qu'on puisse les afficher dans le graphe.

Afin de réduire le nombre d'expressions sur la carte, nous avons suivi deux stratégies. En premier lieu, nous avons enlevé toutes les expressions connectées à une seule des requêtes ou à l'ensemble des douze requêtes. Ces entités, bien sûr, ne sont pas moins importantes que les autres, mais pour elles une visualisation par liste (comme celle présentée dans le Tableau 2) peut remplacer une visualisation par graphe.

[35] Le pois d'un arc connectant une expression et une requête est égal au nombre de résultats de la requête contenant cette expression.

Label	N-docs	Requête
asia	126	Toutes
pakistan	67	Toutes
social networks	23	Toutes
environmentalists	27	Toutes
european commission	292	Toutes
unhcr	227	Toutes
security	79	Toutes
water	171	Toutes
australia	249	Toutes
bangladesh	296	Toutes
urgency	99	Toutes
urban areas	37	Toutes
biofuels	63	Toutes
desertification	154	Toutes
human rights	106	Toutes
emergent crisis	251	Toutes
geneva refugee convention	96	Toutes
international framework	119	Toutes
natural disaster	225	Toutes
dry land	113	Toutes
indigenous people	81	Toutes
unep	482	Toutes
sea level rise	371	Toutes
legal	276	Toutes
environmental factors	104	Toutes
environmental degradation	291	Toutes
climate change	522	Toutes
pacific ocean	48	Toutes
new scientist	70	Toutes
india	296	Toutes
south asia	41	Toutes

Label	N-docs	Requête
climate refugee hoax	5	Climate refugee
human cost	4	Climate refugee
political migrants	3	Environmental migrant
recent earthquake	3	Environmental migrant
nansen passport	4	Environmental migrant
migration control	3	Climate migrant
climate shifts	2	Climate migrant
climate debate	2	Climate migrant
climate stresses	5	Climate migrant
early eco migrant	4	Eco-migrant
human smugglers	23	Eco-migrant
northern hemisphere barren	3	Eco-migrant
undocumented immigrants	3	Eco-migrant
logical correlation	3	Eco-migrant
public housing	3	Eco-migrant
labour rights	2	Ecological migrant
new houses	8	Ecological migrant
environmental rights	3	Ecological refugee
our common future	4	Ecological refugee
southern dust bowl	3	Ecological refugee
cara nine	10	Ecological refugee
philosophical traditions	3	Ecological refugee
marc sheppard	22	Environmentally-induced migrant
shakir husain	14	Environmentally-induced migrant
whacky climate refugees	22	Environmentally-induced migrant
terrifying words	20	Environmentally-induced migrant
nobel laureate	20	Environmentally-induced migrant
environmental horror	14	Environmentally-induced migrant
barbara boxer	14	Environmentally-induced migrant
impossible doomsday scenarios	14	Environmentally-induced migrant
excessive statements	14	Environmentally-induced migrant
robust findings	12	Environmentally-induced migrant
rajendra k. pachauri	12	Environmentally-induced migrant
eua	10	Environmentally-induced migrant
pat carlson	12	Environmentally-induced migrant
corrupt governments	12	Environmentally-induced migrant
environmental groups	4	Environmentally-induced migrant
robust science	10	Environmentally-induced migrant
lisa jackson	12	Environmentally-induced migrant
new country	8	Environmentally-induced migrant
operative phrase	4	Environmentally-induced migrant
apocalyptic view	4	Environmentally-induced migrant
intense heat waves	12	Eco-refugee
rural vietnam	9	Eco-refugee
territorial rights	12	Ecological refugee

Tableau 2. Expressions connectées à l'ensemble des 12 requêtes ou à une seule d'entre elles.

Nous avons ainsi enlevé 76 expressions, mais les 533 expressions restantes sont encore trop nombreuses pour pouvoir être lues dans le graphe. Il convient donc de scinder le graphe en trois sous-graphes, en distinguant les 76 expressions faisant référence à des noms d'acteurs individuels ou collectifs (Fig. 3) et les 100 toponymes géographiques (Fig. 4). Dans les deux

images suivantes, la taille des nœuds-expressions est proportionnelle au nombre de documents (pages web) dans lesquels ils apparaissent. Quant à la couleur des nœuds, elle dépend du nombre de requêtes auxquelles ils sont connectés : plus ce nombre est grand, plus la couleur du nœud est foncée.

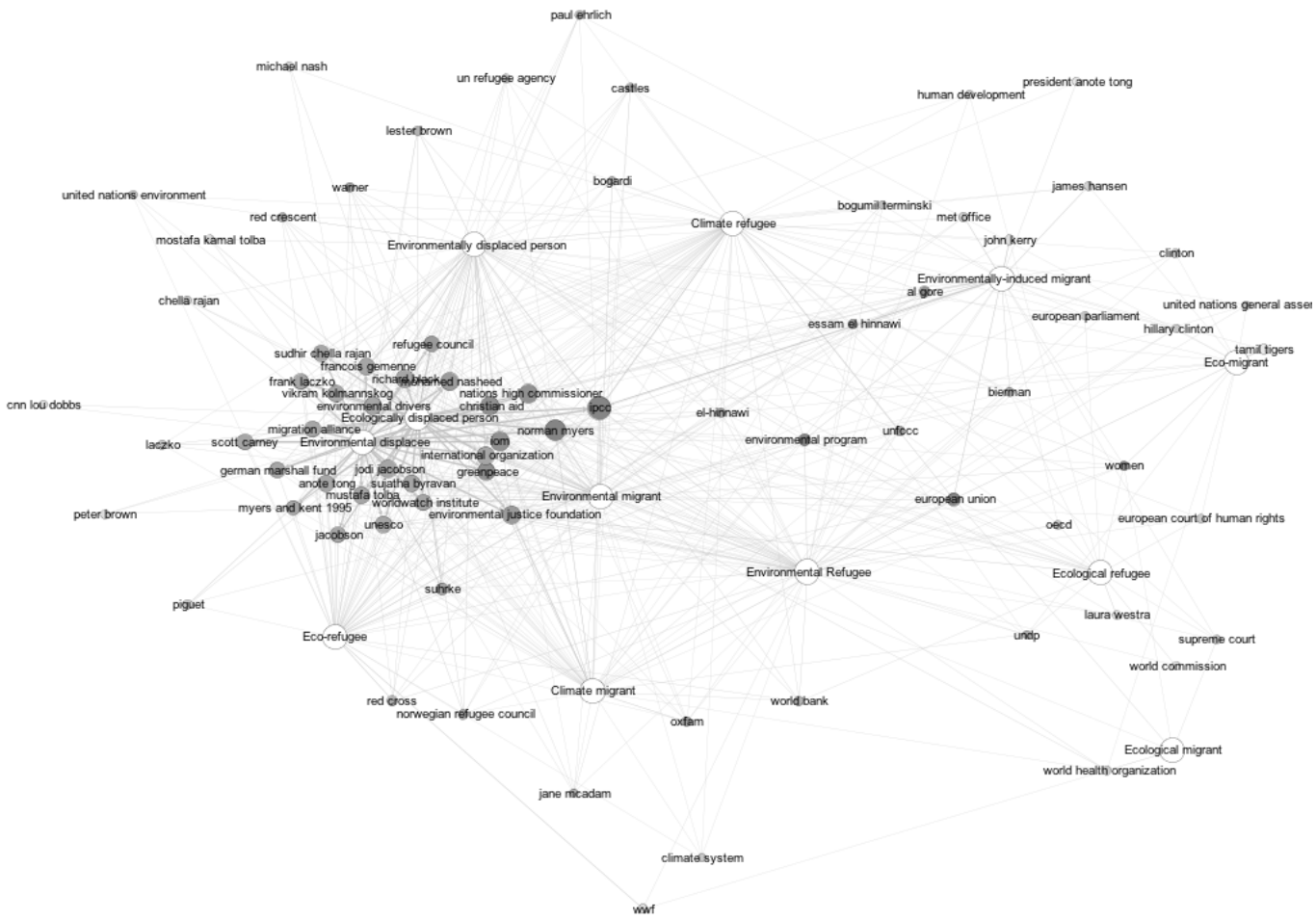

Figure 3. Le réseau des noms d'acteurs (gris) et des requêtes (blanc)

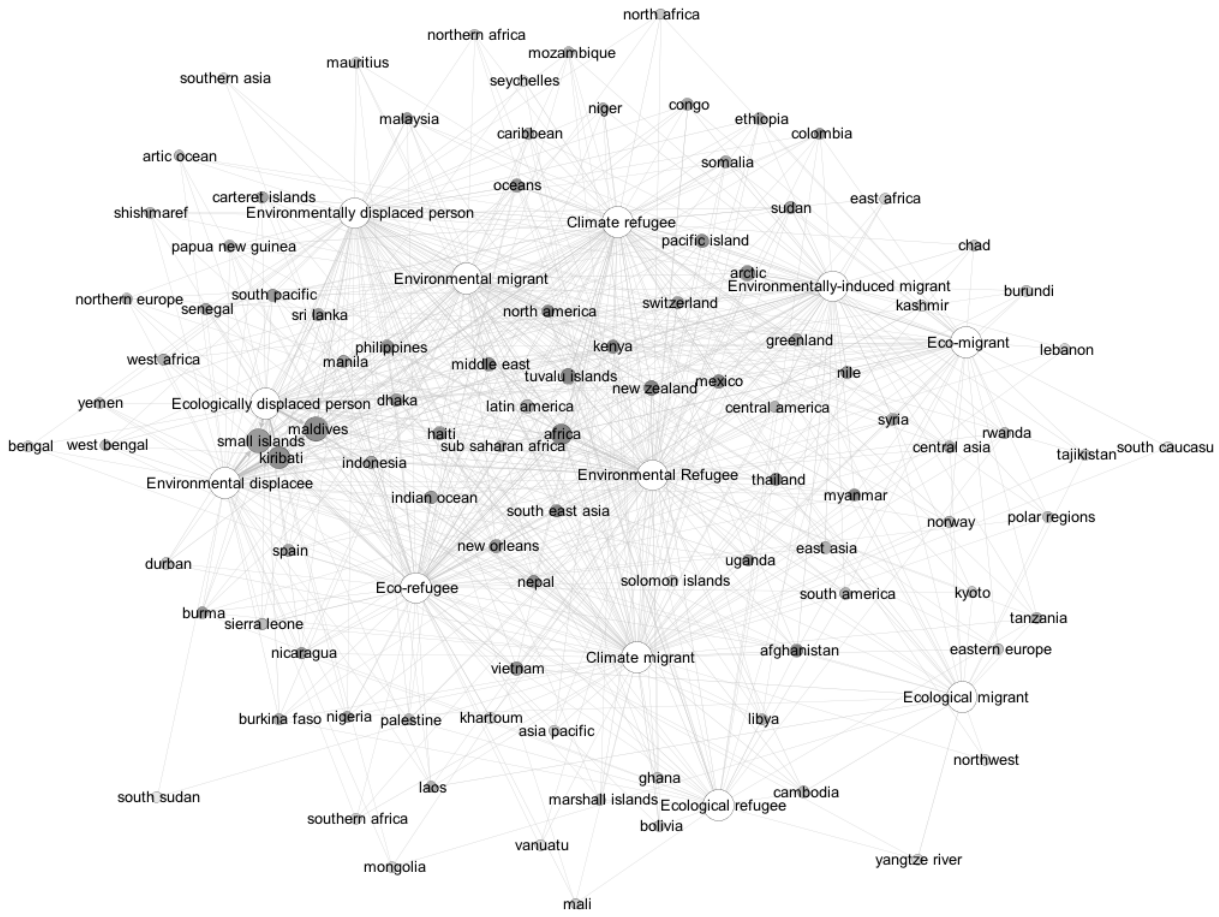

Figure 4. Le réseau des toponymes (gris) et des requêtes (blanc)

Une fois enlevés les noms d'acteurs et les toponymes, il est possible de dessiner un réseau – enfin lisible – des 357 expressions restantes (Fig. 5). Il est intéressant de remarquer que, dans les Figures 3, 4 et 5, la position relative des nœuds représentant les requêtes ne change pas, ce qui en prouve la robustesse.