

Université de Montréal

Qui est à blâmer pour la pandémie de la COVID-19 ? Analyse des perceptions de la responsabilité pendant la crise et évaluation de l'Allocation de Dirichlet latente dans l'étude de questions ouvertes.

Par  
Marianne Chevalier

Département de sociologie  
Faculté des arts et des sciences

Mémoire présenté en vue de l'obtention du grade de Maîtrise ès sciences  
(M. Sc.) en sociologie

Août 2023

© Marianne Chevalier, 2023

Qui est à blâmer pour la pandémie de la COVID-19 ? Analyse des perceptions de la responsabilité pendant la crise et évaluation de l'Allocation de Dirichlet latente dans l'étude de questions ouvertes.

Par

Marianne Chevalier

Le présent mémoire a été évalué par un jury composé des personnes suivantes :

Anne Calvès, Ph. D.

Président-rapporteur

Éric Lacourse, Ph. D.

Directeur de recherche

Roxane de la Sablonnière, Ph. D.

Co-directrice de recherche

Claire Durand, Ph. D.

Membre du jury

## Résumé

La crise de la COVID-19 a provoqué des bouleversements majeurs dans la vie des populations du monde entier et a suscité des réactions sociales importantes. La propagation du virus contagieux de la COVID-19 a été rapidement suivie d'une « épidémie » d'explications et de discours tentant de donner un sens à la crise. Lorsqu'un événement dévastateur se produit, les gens se demandent ce qui se passe et ce que cela signifie. Le premier but de cette recherche est de suivre l'évolution de la dynamique du blâme et de la désignation de boucs émissaires au fur et à mesure que la pandémie de COVID-19 se déroule. Le deuxième but de cette recherche est d'évaluer l'intérêt d'utiliser l'Allocation de Dirichlet latente (ADL), un modèle de mélange/classe latente génératif bayésien, dans l'analyse de questions ouvertes. Les données ont été recueillies auprès d'un échantillon représentatif de 3617 Canadiens selon un devis de recherche longitudinal intensif (avec 12 temps de mesure). Neuf thématiques ont été identifiées, dont six sont récurrentes à différents temps de mesure. Les résultats indiquent que, durant les premiers mois de la pandémie, les Canadiens blâment majoritairement les collectivités distantes, telles que la Chine et les marchés aux animaux vivants (*wet markets*). Au fil du temps, ils blâment de plus en plus les collectivités locales, tels que les individus qui ne respectent pas les mesures sanitaires. Cette recherche met en évidence le rôle de la proximité géographique et de l'évaluation du risque dans la manière dont le public perçoit la pandémie.

Mots clés : Attribution du blâme, Bouc émissaire, Pandémie, COVID-19, *Text Mining*, Allocation de Dirichlet latente

## Abstract

The COVID-19 crisis has caused major disruptions in the lives of populations around the globe and provoked important social responses. The spread of the contagious COVID-19 virus was quickly followed by an outbreak of explanations and discourses trying to make sense of the crisis. When devastating events occur, people ask themselves what happened, why the event happened and what it means. The first goal of this paper is to track the changing dynamics of blame attribution and scapegoating as the COVID-19 pandemic unfolds. The second goal of this paper is to evaluate the relevance of LDA (Latent Dirichlet Allocation), a Bayesian generative mixture/latent class model, to analyze open-ended survey responses. Data was collected from a representative sample of 3,617 Canadians following an intensive longitudinal research design (with 12 waves). Nine topics were identified, six of which were recurring. Canadians mostly blame distant collectives in the early months of the pandemic, especially China and wet markets. Over time, they increasingly blame local collectives, such as individuals who do not comply with sanitary measures. This study highlights the role of geographic proximity and perceived risk in shaping public perceptions of the pandemic.

Keywords: Blame attribution, Scapegoating, Pandemic, COVID-19, *Text Mining*, Latent Dirichlet Allocation

# Table des matières

Résumé .....	3
Abstract.....	4
Liste des tableaux .....	7
Liste des graphiques.....	8
Liste des illustrations .....	9
Avant-propos .....	10
<b>1. Introduction.....</b>	<b>12</b>
<b>1.1 Introduction générale.....</b>	<b>12</b>
<b>1.2 Objectifs de recherche.....</b>	<b>13</b>
<b>2. Contexte .....</b>	<b>15</b>
<b>2.1 Théories sur le blâme et la désignation de boucs émissaires .....</b>	<b>15</b>
<b>2.2 Théories sociologiques sur le blâme et la désignation de boucs émissaires.....</b>	<b>21</b>
<b>2.3 Histoire du blâme durant des épidémies .....</b>	<b>23</b>
<b>2.4 Attribution du blâme pour la COVID-19.....</b>	<b>28</b>
<b>3. Méthodologie .....</b>	<b>31</b>
<b>3.1 Explication de la méthode de l'Allocation de Dirichlet latente.....</b>	<b>31</b>
<b>3.1.1 Modèle analytique du langage .....</b>	<b>31</b>
<b>3.1.2 Inférence bayésienne.....</b>	<b>34</b>
<b>3.1.3 L'Allocation de Dirichlet latente .....</b>	<b>35</b>
<b>3.2 Application de la méthode de l'Allocation de Dirichlet latente.....</b>	<b>44</b>
<b>3.2.1 Prétraitement des données .....</b>	<b>44</b>
<b>3.2.2 Modélisation avec l'ADL.....</b>	<b>48</b>
<b>3.3 Enquête sur les conséquences sociales de la crise de la COVID-19 ...</b>	<b>51</b>
<b>4. Résultats.....</b>	<b>54</b>
<b>4.1 Caractéristiques sociodémographiques des répondants .....</b>	<b>54</b>
<b>4.2 Quelques événements importants liés à la pandémie de la COVID-19 .....</b>	<b>57</b>

<b>4.3 Statistiques descriptives</b> .....	59
<b>4.3 Principaux mots mentionnés</b> .....	61
<b>4.4 Proportion des thématiques de l'ADL et sommaire des thématiques</b>	62
<b>5. Discussion</b> .....	70
<b>5.1 Considérations théoriques</b> .....	71
<b>5.1.1 Attribution du blâme durant la COVID-19</b> .....	71
<b>5.1.2 Évolution du blâme durant la pandémie</b> .....	81
<b>5.2 Considérations méthodologiques</b> .....	83
<b>6. Limites</b> .....	85
<b>7. Conclusion</b> .....	88
<b>Références</b> .....	91
<b>Annexe 1. Tableaux</b> .....	104
<b>Annexe 2. Graphiques</b> .....	111
<b>Annexe 3. Termes remplacés ou réduits en un mot</b> .....	114
<b>Annexe 4. Liste des mots vides supprimés</b> .....	117
<b>Annexe 5. Formulaire de consentement (français)</b> .....	118
<b>Annexe 6. Formulaire de consentement (anglais)</b> .....	122

## Liste des tableaux

<b>Tableau 1. Informations sur les temps de mesure des enquêtes .....</b>	<b>53</b>
<b>Tableau 2. Répartition des répondants selon le pourcentage de femmes et la moyenne d'âge (Question ouverte et échantillon total) .....</b>	<b>55</b>
<b>Tableau 3. Répartition des répondants de l'enquête COVID-19 selon le plus haut niveau d'éducation atteint et recensement de 2016.....</b>	<b>56</b>
<b>Tableau 4. Participation à l'enquête et taux d'attrition.....</b>	<b>57</b>
<b>Tableau 5. Quelques événements importants de la pandémie de la COVID-19 selon les temps de mesure .....</b>	<b>58</b>
<b>Tableau 6. Statistiques descriptives du nombre de mots.....</b>	<b>60</b>
<b>Tableau 7. Pourcentage des réponses ne comportant qu'un seul mot .....</b>	<b>60</b>
<b>Tableau 8. Mots seuls les plus fréquents et leurs proportions sur le nombre total de réponses .....</b>	<b>61</b>
<b>Tableau 9. Paramètres de l'ADL et leurs intervalles.....</b>	<b>104</b>
<b>Tableau 10. Paramètres des modèles sélectionnés et scores de stabilité ...</b>	<b>104</b>
<b>Tableau 11. Répartition des répondants selon les catégories d'âge.....</b>	<b>104</b>
<b>Tableau 12. Sommaire des thématiques au temps 1 .....</b>	<b>105</b>
<b>Tableau 13. Sommaire des thématiques au temps 3 .....</b>	<b>106</b>
<b>Tableau 14. Sommaire des thématiques au temps 5 .....</b>	<b>107</b>
<b>Tableau 15. Sommaire des thématiques au temps 8 .....</b>	<b>108</b>
<b>Tableau 16. Sommaire des thématiques au temps 10 .....</b>	<b>109</b>
<b>Tableau 17. Sommaire des thématiques au temps 12 .....</b>	<b>110</b>

## Liste des graphiques

<b>Graphique 1. Proportion des thématiques de l'ADL.....</b>	<b>63</b>
<b>Graphique 2. 10 principaux mots mentionnés au temps 1 .....</b>	<b>111</b>
<b>Graphique 3. 10 principaux mots mentionnés au temps 3 .....</b>	<b>111</b>
<b>Graphique 4. 10 principaux mots mentionnés au temps 5 .....</b>	<b>112</b>
<b>Graphique 5. 10 principaux mots mentionnés au temps 8 .....</b>	<b>112</b>
<b>Graphique 6. 10 principaux mots mentionnés au temps 10 .....</b>	<b>113</b>
<b>Graphique 7. 10 principaux mots mentionnés au temps 12 .....</b>	<b>113</b>



## **Liste des illustrations**

<b>Illustration 1. Représentation graphique du modèle .....</b>	<b>39</b>
<b>Illustration 2. Exemples de documents avant le prétraitement des données .....</b>	<b>46</b>
<b>Illustration 3. Exemples de documents après le prétraitement des données .....</b>	<b>46</b>
<b>Illustration 4. Représentation d'un DTM .....</b>	<b>47</b>
<b>Illustration 5. Résumé des étapes d'application de l'ADL .....</b>	<b>50</b>

## Avant-propos

Ce projet de mémoire s’inscrit dans un programme de recherche dirigé par la professeure Roxane de la Sablonnière et une équipe composée des professeurs Éric Lacourse, Jean-Marc Lina et Dietlind Stolle et du chercheur Mathieu Pelletier-Dumas. Ce programme vise à comprendre les impacts sociaux et psychologiques de la crise de la COVID-19 sur la population canadienne. Le projet de mémoire est basé sur un article soumis pour publication dans une revue spécialisée. Il a été rédigé sous forme d’article scientifique détaillé. La première auteure, Marianne Chevalier, a formulé la question de recherche, l’hypothèse et les objectifs, produit les analyses statistiques, les tableaux et les graphiques, et rédigé intégralement le texte. Le deuxième auteur, Éric Lacourse, a contribué significativement à cette étude par son expérience en modélisation statistique. La dernière auteure, Roxane de la Sablonnière, a apporté une contribution majeure à ce travail par son expertise sur les changements sociaux dramatiques. Ils ont supervisé l’ensemble des étapes de cette étude sur les plans théoriques et méthodologiques. La professeure Sylvie Ratté de l’École de technologie supérieure et son étudiant à la maîtrise en génie logiciel Simon-Olivier Harel ont collaboré à la réalisation de ce projet par leur expérience en traitement automatique du langage naturel. Tous les autres membres du groupe de recherche de l’enquête sur la COVID-19, c’est-à-dire Anna Dorfman, Mathieu Pelletier-Dumas, Jean-Marc Lina et Dietlind Stolle, ont apporté une contribution en produisant le devis de recherche, les questionnaires et en organisant la collecte de données. Leurs commentaires et suggestions ont également joué un rôle important dans l’élaboration de l’article. Le programme de recherche a été approuvé par le Comité d’éthique de la recherche en éducation et en psychologie (CEREP) de l’Université de Montréal et financé par les Instituts de Recherche en santé du Canada (IRSC)<sup>1</sup>. La présente étude a été soutenue par la Bourse

---

<sup>1</sup> Sous la subvention 170633

d'initiation à la recherche au 1<sup>er</sup> cycle de l'Institut de valorisation des données (IVADO)<sup>2</sup> et par une subvention de recherche du Centre pour l'étude de la citoyenneté démocratique (CÉCD).

L'article, dont le titre est *Who's To Blame for the COVID-19 pandemic? Perceptions of responsibility during the crisis using text mining and Latent Dirichlet Allocation* a été soumis à la revue *Health Sociology Review* en avril 2023. Il y a des similitudes et des différences entre le mémoire et l'article. Tout d'abord, pour ce qui est des similarités, les deux textes traitent de la même question de recherche, partagent une hypothèse commune et poursuivent des objectifs semblables. Ensuite, la méthodologie de l'article est similaire à celle mobilisée dans le mémoire, et les analyses ont mené à des conclusions et constats analogues. Par la suite, les différences se trouvent essentiellement dans l'approfondissement de certains aspects du texte et l'ajout de nouvelles analyses. La problématique, la discussion et les limites de la recherche sont abordées de manière plus détaillée dans le mémoire. Enfin, le mémoire reprend les analyses et résultats présentés dans l'article et inclut également des données supplémentaires liées au dernier temps de mesure de l'enquête (temps 12). En effet, l'étude présentée dans ce mémoire est longitudinale et comprend plusieurs vagues de collecte.

---

<sup>2</sup> Sous la subvention BSc-2021a-3893462506

# 1. Introduction

## 1.1 Introduction générale

La pandémie de la COVID-19 et les mesures sanitaires pour limiter sa propagation ont provoqué des bouleversements rapides et importants parmi les populations du monde entier. Il s'est écoulé près de deux mois entre la détection du premier cas du virus à Wuhan, en Chine, et la déclaration de l'Organisation mondiale de la Santé selon laquelle la COVID-19 a atteint le stade de pandémie (Labbé et coll., 2022). Un grand nombre de pays ont ensuite instauré des protocoles sanitaires et économiques pour réduire sa transmission. Les mesures sanitaires comprennent la distanciation sociale, le confinement, le port du masque, le lavage des mains dans les lieux publics ainsi que l'isolement des personnes atteintes de la maladie. Les mesures économiques incluent notamment la fermeture d'entreprises, de commerces et de services non essentiels. Près de trois ans plus tard, environ 670 millions de personnes ont été infectées et 6,8 millions d'individus en sont morts (Johns Hopkins University, 2023). Le virus de la COVID-19 a muté au fur et à mesure de sa propagation ; plusieurs variants sont apparus et sont devenus des souches dominantes dans le monde.

Parallèlement à la pandémie causée par le virus lui-même, des « pandémies » de recherches d'explications ou d'éclaircissements à propos de la COVID-19 ont émergé (Labbé, 2022). Lorsqu'il se produit un événement dévastateur, les individus se posent des questions et sont perplexes. Ils cherchent à comprendre ce qui s'est passé, les raisons qui ont conduit à cet événement et les conditions sous lesquelles il pourrait se reproduire. Les individus sont généralement très sensibles aux événements négatifs (Brandt et Rozin, 1997 ; Malle et coll., 2014). En effet, ces derniers mobilisent plus de « ressources attentionnelles » et sont plus représentés dans le langage (Malle et coll., 2014).

De plus, lorsqu'un tel événement est soudain et inattendu, la quantité d'informations disponibles est souvent rare et changeante (Attema, 2021). Ce manque d'information peut se traduire par l'absence de réponses satisfaisante, ce qui incite les gens à chercher des responsables (Strong, 1990; Mayor et coll., 2012 ; Dionne & Turkmen, 2020). Ce type de raisonnement peut donner lieu à une rhétorique de blâme (Mayor et coll., 2012 ; Dionne & Turkmen, 2020).

Le blâme et la désignation de boucs émissaires en temps de crise ont des conséquences sociales, psychologiques et politiques. Ces phénomènes renforcent les préjugés, la discrimination et la stigmatisation de groupes marginalisés (Douglas, 1992 ; Brandt et Rozin, 1997). Durant la crise de la COVID-19, les communautés chinoises ont été particulièrement ciblées comme responsables de l'origine de la maladie, entraînant leur exclusion ainsi que des actes discriminatoires et racistes à leur égard (de Rosa & Mannarini, 2020 ; Choli & Kuss, 2021 ; Bouguettaya, 2022). La stigmatisation affecte les conditions de vie et le bien-être des communautés touchées (Douglas, 1992 ; Bor et coll., 2022). En effet, elle peut avoir des conséquences sur l'estime de soi, sur les relations sociales, sur la santé, ainsi que sur les opportunités de réussite dans la vie, que ce soit en matière d'éducation, de profession ou de revenu (Bor et coll., 2022). Finalement, le blâme et la désignation de boucs émissaires peuvent être encouragés par certains politiciens, ce qui peut mener à un affaiblissement de la cohésion sociale et créer des divisions au sein de la société (Forester & McKibbin, 2020). Ceux-ci peuvent profiter du climat de tensions pour alimenter des discours populistes, renforcer leur pouvoir et promouvoir des politiques discriminatoires (Forester & McKibbin, 2020).

## **1.2 Objectifs de recherche**

Récemment, un intérêt croissant s'est manifesté afin de comprendre l'attribution du blâme et la désignation de boucs émissaires durant la pandémie

de la COVID-19 (Hardy et coll., 2021 ; Nguyen et coll., 2021). Bien que certaines recherches aient étudié l'attribution « dynamique » du blâme à travers une crise en constante transformation (Choli & Kuss, 2021 ; Labbé et coll., 2022), il existe un manque important de connaissances sur sa manifestation et son évolution au sein de la population. Cette recherche vise à présenter des perspectives d'analyse qui peuvent être applicables à un contexte populationnel. Le premier objectif de ce mémoire est donc de suivre l'évolution de la dynamique du blâme et de la désignation de boucs émissaires au cours des développements de la pandémie de la COVID-19 auprès de la population canadienne. L'hypothèse de recherche est que le blâme suit une tendance où il est initialement imputé à des communautés distantes (Mayor et coll., 2012 ; Roy et coll., 2020 ; Labbé et coll., 2022). Au fur et à mesure que le virus se propage, les attributions « éloignées » diminuent et les assignations à des entités plus proches augmentent (Mayor et coll., 2012 ; Roy et coll., 2020 ; Labbé et coll., 2022).

Jusqu'à présent, l'attribution du blâme durant la pandémie de la COVID-19 a été abordée par une revue de la littérature (Barreneche, 2020), un codage d'entretiens et de questions ouvertes (Hardy et coll., 2021) ainsi que des analyses thématiques conventionnelles (Choli & Kuss, 2021 ; Nguyen et coll., 2021). Cependant, aucune étude n'a examiné l'attribution du blâme en utilisant des approches automatisées qui permettent d'analyser rapidement de grandes quantités de données textuelles. Ces approches restent encore peu explorées en sciences sociales et la plupart des chercheurs privilégient des techniques plus familières, telles que le codage manuel (Lindstedt et coll., 2019). L'Allocation de Dirichlet latente (en anglais *Latent Dirichlet Allocation*), un modèle de mélange issu des approches computationnelles bayésiennes, permet d'accomplir cette tâche en générant les thèmes dissimulés dans du texte. La présente étude se concentrera sur l'exploration d'une question ouverte d'enquête, afin de mieux comprendre le processus d'attribution de la responsabilité durant la pandémie de

COVID-19. Le deuxième objectif de ce mémoire est d'évaluer l'intérêt d'utiliser l'Allocation de Dirichlet latente dans l'étude de questions ouvertes d'enquêtes.

Ce mémoire est organisé comme suit. Les aspects théoriques (théories sur le blâme et la désignation de boucs émissaires, histoire du blâme durant des épidémies et attribution du blâme pour la COVID-19) sont abordés dans la section « Contexte ». Dans la section « Méthodologie », une explication détaillée de la méthode et la population à l'étude sont présentées. Ensuite, les aspects théoriques et méthodologiques qui en découlent sont abordés dans les sections « Résultats » et « Discussion ». Les limites de la recherche sont présentées dans le volet « Limites ». Pour terminer, les conclusions et les possibilités de recherches futures sont abordées à la fin du mémoire.

## **2. Contexte**

### **2.1 Théories sur le blâme et la désignation de boucs émissaires**

Dans le domaine de la cognition, la théorie du cerveau bayésien est une des approches les plus courantes pour comprendre comment les individus perçoivent le monde qui les entoure (Bottemanne et coll., 2021 ; Attema, 2021). Cette théorie offre une explication de la façon par laquelle les individus obtiennent et traitent des informations qui leur parviennent. Selon cette conception, le cerveau utilise des connaissances et des croyances préexistantes pour interpréter les données qu'il reçoit et générer des projections (Kelly et coll., 2019). Quand il fait des erreurs de prédiction, il actualise ses connaissances lorsque de nouveaux renseignements lui parviennent (Longstaff et Yang, 2008 ; Bottemanne et coll., 2021 ; Attema, 2021). La combinaison des stocks de connaissances, de croyances et de nouvelles informations produit des

*typifications* (Schutz, 1967 [1932] ; Kelly et coll., 2019). Les typifications sont des processus où l'individu utilise ses connaissances générales pour construire des idées sur le monde, ce qui lui permet d'appréhender, d'interpréter et de simplifier la complexité de la vie sociale (Schutz, 1967 [1932]). Les individus tentent donc de donner du sens à un événement dévastateur, tel qu'une pandémie, en le rattachant à des représentations préexistantes découlant de la mémoire collective et sociale (de la Sablonnière, 2013 ; de Rosa & Mannarini, 2020 ; Dionne & Turkmen, 2020 ; Caron-Diotte et coll., 2022). La mémoire collective désigne l'ensemble des conceptions partagées par les membres d'un groupe, lesquelles contribuent à la définition de l'identité commune de ce groupe (Caron-Diotte et coll., 2022). Ces représentations proviennent de connaissances et de perceptions antérieures qui sont ensuite généralement mises à jour en fonction de nouvelles informations (Lynch & Bartlett, 2019). Toutefois, les stocks de connaissances et de croyances dont disposent les individus pour comprendre le monde peuvent parfois être ambigus, incohérents ou faux (Longstaff et Yang, 2008 ; Kelly et coll., 2019). Les nouveaux renseignements qu'un individu reçoit peuvent occasionnellement servir à renforcer des croyances fallacieuses plutôt qu'à les corriger (Douglas, 1992). Ces croyances peuvent mener à des réactions émotionnelles et cognitives telles que le blâme (Landau et coll., 2015).

Le blâme peut prendre deux formes : il peut être « cognitif », c'est-à-dire un jugement personnel et privé, ou « social », soit un acte public où il est exprimé (Malle et coll., 2014). Le blâme « cognitif » se manifeste en réponse à des stimuli extérieurs, à des émotions ou à un traitement particulier de l'information, alors que le blâme « social » se produit notamment en raison de violation de normes (Malle et coll., 2014). La violation de normes survient lorsqu'il y a infraction à une règle socialement établie (Malle et coll., 2014). Par exemple, durant la COVID-19, les travailleurs du secteur de la santé ont été accusés de propager la maladie dans les résidences pour personnes âgées (Bouguettaya, 2022). Il peut y avoir des différences dans l'attribution du blâme en fonction de la norme qui est



transgressée ou de la vision de la responsabilité (Bouguettaya, 2022). Une personne peut imputer une responsabilité individuelle en pointant du doigt les travailleurs infectés ou une responsabilité collective en désignant les résidences pour personnes âgées (Bouguettaya, 2022). Le processus d'attribution du blâme permet de donner un sens à une crise, c'est-à-dire de trouver « des explications ou des réponses satisfaisantes » à une situation qui ne peut être élucidée de manière conventionnelle ou qui ne provient pas de facteurs connus (Bucher, 1957 ; Nelkin & Gilman, 1988 ; Barreneche, 2020). En effet, dans des contextes où il est difficile de trouver des causes ou des coupables à une crise, il est plus simple de construire des acteurs imaginaires ou fictifs qui peuvent être tenus responsables de la situation (Barreneche, 2020).

La signification accordée à un événement dévastateur peut varier selon le groupe social et l'époque (Rateau et coll., 2021). Rateau et coll. (2021) énumèrent quatre catégories distinctes qui permettent de classer l'origine des événements perturbateurs, basées sur la taxonomie de Rouquette (2013) concernant les peurs collectives. Premièrement, l'événement peut être imputé aux actions d'individus ou de groupes (ou intervention humaine directe). Par exemple, les groupes engagés dans une guerre biologique ou le bioterrorisme peuvent être pointés du doigt comme responsables. Deuxièmement, il peut être attribué au développement technologique ou à une politique économique perçue négativement (ou intervention humaine indirecte). À titre d'exemple, les changements climatiques peuvent être blâmés pour une crise. Troisièmement, l'événement peut être interprété comme étant le résultat du châtement de Dieu en punition d'une faute (ou intervention non humaine). Dernièrement, l'événement peut être perçu comme étant une coïncidence (ou aucune intervention intentionnelle) (Rateau et coll., 2021). Cette typologie permet de comprendre la diversité des perceptions et réactions des individus face à des crises. Par exemple, une épidémie peut être à la fois considérée comme une punition divine, une

conséquence de pratiques sanitaires irresponsables, ou un enjeu de santé publique.

Le phénomène de l'attribution du blâme a été particulièrement étudié en psychologie et en psychologie sociale (Malle et coll., 2014 ; Landau et coll., 2015 ; Proulx & Inzlicht, 2012). D'après ces perspectives, les humains ressentent un besoin inhérent de percevoir le monde comme étant ordonné et structuré et d'avoir un sentiment de contrôle sur leur vie (Proulx & Inzlicht, 2012). Lorsque les individus sont confrontés à des circonstances qui diminuent leur perception de contrôle, ils emploient des stratégies compensatoires pour le rétablir (Landau et coll., 2015). Une de ces stratégies est de rechercher et de préférer des interprétations claires et simples de la réalité (Landau et coll., 2015). Selon le modèle de maintien du sens (*Meaning maintenance model*), tout événement menaçant qui porte atteinte au cadre de signification de l'individu, c'est-à-dire à l'impression que ses expériences sont compréhensibles et familières, entraîne une compensation psychologique (Proulx & Inzlicht, 2012). Elle peut prendre la forme d'actions immédiates et manifestes (ou proximales) qui visent à éviter la menace ou des modes de compensation indirects (ou distaux) qui servent à atténuer le sentiment d'anxiété (Landau et coll., 2015). Par exemple, l'individu peut recourir à des mécanismes de défense, tels que la désignation d'un bouc émissaire, pour pallier le sentiment d'impuissance (Malle et coll., 2014 ; Landau et coll., 2015). La désignation de boucs émissaires consiste à blâmer une personne pour un événement négatif qui n'a pas de cause déterminante (Jensen, 2007). Ceci ne résout pas le problème sous-jacent de la menace et peut occasionner des préjudices. Des communautés peuvent être stigmatisées et persécutées pour des crises qu'elles n'ont pas provoquées (Jensen, 2007 ; Barreneche, 2020). Des exemples concrets de communautés stigmatisées pour des crises sanitaires seront présentés dans la section portant sur l'histoire du blâme durant des épidémies.

En psychologie sociale et en sociologie, les recherches sur le blâme abordent les notions de catégorisation sociale et de perception de « l'autre » (Bucher, 1957 ; Petersen et Lupton, 2000 ; Jensen, 2007 ; Barreneche, 2020). Selon ces perspectives, les individus s'associent à des groupes afin de se sentir valorisés et spéciaux (Bouguettaya, 2022). Les groupes sont en compétition les uns avec les autres pour obtenir une identité sociale positive, c'est-à-dire pour être considérés comme étant à la fois différents et meilleurs que les autres (Bouguettaya, 2022). Cette tendance à la compétition affecte les façons par lesquelles les collectivités interagissent entre elles (Bouguettaya, 2022). La catégorisation sociale correspond à la propension à classer le monde en groupes selon certaines dimensions générales telles que le genre, l'âge, l'occupation ou l'ethnicité (Baron et coll., 1997). Toutefois, cette propension va au-delà de la simple catégorisation ; il s'agit de scinder les collectivités en deux principales entités : l'endogroupe et l'exogroupe (Petersen et Lupton, 2000). L'endogroupe est le groupe social auquel un individu appartient (« nous ») et l'exogroupe correspond à l'ensemble des individus qui n'y appartiennent pas (« eux ») (Baron et coll., 1997). Le premier est perçu de manière positive, alors que le deuxième est perçu de façon négative. Plus précisément, selon l'étude de Nguyen et coll. (2021), dont le but est de comprendre le processus d'attribution du blâme durant la pandémie de la COVID-19 auprès de la population néo-zélandaise, les membres de l'exogroupe sont considérés comme ayant davantage de caractéristiques indésirables et sont souvent détestés. Ils sont également perçus comme étant plus semblables que les membres de l'endogroupe (Baron et coll., 1997). Allport (1979) apporte une précision supplémentaire en introduisant la notion de groupe de référence. Celui-ci correspond au groupe auquel l'individu se rattache ou auquel il aspire à appartenir (Allport, 1979). Si les concepts d'endogroupe et d'exogroupe décrivent le simple fait d'appartenir à un groupe, le concept de groupe de référence indique si l'individu souhaite cette appartenance ou s'il cherche plutôt à se rallier à un autre groupe (Allport, 1979). Le blâme permet d'attribuer une forme de responsabilité et de culpabilité auprès

de l'exogroupe et de maintenir « l'identité positive » de l'endogroupe (Bouguettaya, 2022 ; Ittefaq et coll., 2022).

En politique, la théorie des jeux permet de comprendre les motivations derrière des discours ou des actions de politiciens (Wagner, 1986). Ils peuvent recourir au blâme et à la désignation de boucs émissaires comme des stratégies politiques afin de créer des divisions au sein de la société, détourner l'attention sur leurs propres échecs et mobiliser leur base électorale en exploitant des sentiments de peur ou de préjugés (de Rosa & Mannarini, 2020 ; Forester & McKibbon, 2020 ; Choli & Kuss, 2021). En attribuant le blâme à des individus ou des groupes spécifiques, les politiciens cherchent à se soustraire à toute responsabilité pour leurs propres actions ou inactions (Forester & McKibbon, 2020). Ils peuvent également tirer profit des réactions émotionnelles et des préoccupations de leurs électeurs, ainsi que des préjugés existant envers les groupes marginalisés afin de renforcer leur attrait politique (de Rosa & Mannarini, 2020). Dans un contexte de compétition électorale, l'attribution du blâme et la recherche de boucs émissaires peuvent paraître comme des tactiques politiques efficaces, puisque les individus tendent à accorder plus d'importance aux informations négatives qu'aux informations positives (Brandt et Rozin, 1997 ; Malle et coll., 2014). La capacité des individus et des organisations à recevoir des informations fiables risque ainsi d'être compromise, ce qui peut conduire à une diffusion plus rapide de rumeurs et de fausses nouvelles (Longstaff et Yang, 2008). La société devient donc moins disposée à agir de manière coordonnée pour contribuer à limiter la propagation de la maladie (Forester & McKibbon, 2020).

## **2.2 Théories sociologiques sur le blâme et la désignation de boucs émissaires**

En sociologie, le blâme et la désignation de boucs émissaires sont des processus complexes et structurés qui s'appuient sur des préoccupations sociales, politiques et idéologiques (Nelkin & Gilman, 1988). La désignation de boucs émissaires correspond à un construit provenant de croyances préexistantes et protégeant les relations de pouvoir, les normes et les frontières sociales (Nelkin & Gilman, 1988 ; Jensen, 2007 ; Barreneche, 2020 ; de Rosa & Mannarini, 2020 ; Dionne & Turkmen, 2020 ; Rateau et coll., 2021). Désigner un bouc émissaire consiste à rejeter le blâme, la colère ou l'anxiété sur des individus qui ne sont pas responsables d'un problème, mais qui constituent des cibles faciles à attaquer en raison de caractéristiques qu'elles possèdent (Jensen, 2007). Elles peuvent représenter une menace réelle ou potentielle aux yeux du groupe qui attribue le blâme. Les perspectives fonctionnalistes et conflictuelles sont couramment utilisées pour comprendre les comportements sociaux et le changement sociétal (Jensen, 2007). La première met l'accent sur le maintien de l'ordre et la stabilité globale de la société, alors que la deuxième focalise sur les dynamiques de pouvoir et les inégalités sociales.

Stinchcombe (1968) propose une explication fonctionnaliste des comportements sociaux qui permet de comprendre la notion de bouc émissaire dans une perspective sociologique. Il s'agit d'un raisonnement selon lequel « les conséquences d'un comportement sont des éléments essentiels des causes de ce comportement » (Stinchcombe, 1968, p.80). Plus précisément, une menace ou un risque pour la société compromet l'état d'équilibre d'un système social (ou homéostasie sociétale). La diminution de cet état d'équilibre provoque une réponse qui permet de le rétablir ou de le renforcer. Par exemple, l'arrivée d'une épidémie entraîne une baisse du sentiment de sécurité et une augmentation de la

peur. Au fur et à mesure que la sécurité décline et que la peur croît, la recherche d'explications augmente, ce qui peut mener à la désignation de boucs émissaires. Finalement, pour que les explications à la menace fonctionnent comme prévu, elles doivent rétablir la sécurité et diminuer, voire déplacer la peur si un bouc émissaire est identifié (Jensen, 2007). Par exemple, en attribuant la responsabilité de l'épidémie à un groupe, il est possible que les individus se sentent momentanément en sécurité, car ils pensent pouvoir se protéger en évitant tout contact avec celui-ci (Petersen et Lupton, 2000 ; Roy et coll., 2020). En se différenciant du groupe à risque ou « malsain », le groupe en bonne santé ou « sain » peut projeter ses craintes relatives à la mort et à la maladie sur ce dernier (Petersen et Lupton, 2000). La désignation de bouc émissaire a lieu « lorsque les individus sont incertains de la source de la menace ou ont besoin de construire une explication qui déplace ou externalise la faute » (Jensen, 2007, p.185). Elle fonctionne comme un mécanisme visant à préserver l'équilibre social en attribuant la responsabilité de la menace à des groupes spécifiques.

Selon la théorie du conflit social, qui relève de la perspective conflictuelle, le blâme et la désignation de boucs émissaires sont utilisés par les groupes en position de pouvoir afin de maintenir leur statut et de préserver leurs intérêts (Jensen, 2007). En effet, selon Bardosh et coll. (2022), lors d'une crise, les inégalités sociales et économiques existantes peuvent s'aggraver, ce qui engendre des tensions et des conflits au sein de la société. Le blâme et la désignation de boucs émissaires sont des stratégies qui permettent aux groupes en position de pouvoir, tels que les politiciens, les élites économiques et les médias de masse, de faire diversion sur les véritables problèmes. Par exemple, durant une crise sanitaire, certains politiciens peuvent pointer du doigt des communautés en tant que responsables de l'origine ou de la propagation d'une maladie comme stratégie visant à détourner l'attention sur la mauvaise gestion de l'épidémie. Ceux-ci disposent davantage de ressources et d'influence pour façonner les discours dominants sur une crise sociale (Jensen, 2007).

L'attribution du blâme et la désignation de boucs émissaires peuvent exacerber les tensions et les inégalités sociales existantes, ainsi que renforcer les préjugés ou les discriminations envers certains groupes de la société (Dionne & Turkmen, 2020 ; Roy et coll., 2020). Les groupes déjà marginalisés ou vulnérables, tels que les minorités ethniques, les migrants et les personnes économiquement défavorisées sont souvent les plus ciblés en raison de leur position sociale précaire liée à leur origine ethnique, leur nationalité ou leur religion (Douglas, 1992 ; Brandt et Rozin, 1997 ; Casciano, 2020 ; Banerjee et coll., 2020 ; Jedwab et coll., 2021 ; Bardosh et coll., 2022 ; Voigt, 2022).

### **2.3 Histoire du blâme durant des épidémies**

Le blâme a toujours représenté un moyen d'expliquer des maladies mystérieuses, inconnues et graves, ce qui permettait éventuellement de les contrôler (Brandt et Rozin, 1997 ; Nelkin & Gilman, 1988). Lorsque la science médicale n'est pas en mesure de fournir des explications définitives à des pathologies, les individus tentent de formuler leurs propres interprétations afin de réduire leur sentiment de vulnérabilité (Nelkin & Gilman, 1988 ; Roy et coll., 2020). L'attribution de la responsabilité à des exogroupes, tels que des classes sociales ou groupes ethniques, est un processus qui s'est produit de manière récurrente durant différentes crises sanitaires (Nelkin & Gilman, 1988 ; Brandt & Rozin, 1997 ; Mondragon et coll., 2017 ; Roy et coll., 2020 ; Choli & Kuss, 2021 ; Ittefaq et coll., 2022).

Les communautés juives ont été blâmées pour l'origine et la propagation de la peste noire qui a sévi au milieu du 14<sup>e</sup> siècle ; ils étaient accusés d'avoir empoisonné l'eau et de répandre la maladie (Bouguettaya, 2022). Les couches vulnérables et défavorisées de la société étaient également pointées du doigt ; beaucoup d'individus n'ont pas été soignés et ont été abandonnés à leur sort (Brandt et Rozin, 1997 ; Banerjee et coll., 2020). Dans certaines régions

européennes, comme Bâle et Strasbourg, beaucoup de juifs ont été brûlés en guise de punition pour la maladie (Banerjee et coll., 2020).

Durant l'épidémie de variole de Montréal, qui s'est déroulée entre 1885 et 1886, les Canadiens français ont été désignés comme coupables de répandre la maladie dans certains journaux anglophones (Bliss, 1993). Ils ont été accusés de ne pas se faire vacciner ou de faire vacciner leurs enfants. Un inspecteur sanitaire interviewé par le *Witness* mentionne ceci : « Si elles [les classes francophones] sont frappées par la maladie, c'est bien de leur faute » (Bliss, 1993, p.122). En réalité, si un grand nombre de francophones ont refusé la vaccination, ce n'était pas en raison d'une quelconque infériorité ethnique, mais en raison de la peur, de l'ignorance et du fatalisme (Bliss, 1993, p.250). En effet, durant cette période, le risque de contamination bactérienne, le manque de communication des autorités de la santé publique sur l'efficacité du vaccin et les discours des « experts » opposés à cette pratique ont pu contribuer à la faible vaccination de cette population. Certaines familles cachaient également les cas de variole par peur de perdre leur source de revenus. De plus, l'isolation en cas de maladie était difficile, puisqu'elles vivaient dans des quartiers insalubres à plusieurs ménages dans un appartement. Certains Canadiens-français croyaient également que les vaccins étaient une manœuvre des anglophones pour les empoisonner. D'autre part, les journaux francophones blâmaient la communauté anglophone de Montréal pour la propagation de l'épidémie en évoquant le dépôt de ses ordures et des déchets sanglants des animaux à proximité des quartiers francophones (Bliss, 1993).

Les épidémies ont également été associées à des exogroupes parmi les populations marginalisées et moins privilégiées provenant de pays plus pauvres (Mondragon et coll., 2017). Ceux-ci ont été représentés comme étant responsables de pratiques nocives ou de complots intentionnels et malveillants visant à propager la maladie (Mondragon et coll., 2017). Les immigrants chinois en Amérique du Nord ont souvent servi de boucs émissaires pour des éclosions



de variole durant le 19<sup>e</sup> siècle (Dionne & Turkmen, 2020). Pourtant, ils ne représentaient qu'une petite fraction des décès liés aux maladies infectieuses ; à San Francisco, dans les années 1870, les taux de morbidité et de mortalité étaient bien plus élevés parmi les immigrants irlandais que parmi les immigrants chinois (Dionne & Turkmen, 2020). Les Asiatiques ont été également victimes d'actes de discrimination au Canada lors d'une épidémie de variole en 1892, alors qu'il n'y avait eu que quatre décès liés à cette maladie (Dionne & Turkmen, 2020). Ils ont été pointés du doigt pour l'épidémie de peste qui est apparue à San Francisco entre 1900 et 1904, même si les découvertes liées à la théorie des germes ont rompu le lien direct entre les épidémies et le quartier chinois (Dionne & Turkmen, 2020 ; Ittefaq et coll., 2022). Ce quartier a notamment été qualifié de « laboratoire d'infection » et les Chinois ont été décrits comme étant « sans scrupules, menteurs et traîtres » (Dionne & Turkmen, 2020, p.216). De plus, cette épidémie a été désignée en tant que « maladie orientale, propre aux mangeurs de riz » (Ittefaq et coll., 2022, p.18).

Récemment, l'écllosion du SRAS (syndrome respiratoire aigu sévère) a été associée à la Chine et aux Chinois, et plus précisément à leurs pratiques « agro(culturelles) » (Abeysinghe, 2016 ; Dionne & Turkmen, 2020). Celle-ci a été qualifiée de « maladie mortelle qui menaçait les frontières nationales » par les médias occidentaux (Dionne & Turkmen, 2020, p.218). Les immigrants chinois de la ville de New York étaient fréquemment pointés du doigt et stigmatisés par les autres résidents (de Rosa & Mannarini, 2020 ; Choli & Kuss, 2021). Lorsque l'épidémie de SRAS a atteint son apogée aux États-Unis, des avertissements demandant d'éviter des quartiers habités par des Asiatiques ont été largement diffusés (Joffe, 2011). Durant les épisodes de SRAS et de grippe aviaire, les Asiatiques ont été perçus comme étant des communautés ayant des pratiques culturelles et alimentaires « malsaines » ou « insalubres » (Barreneche, 2020 ; Martikainen et Sakki, 2020).

Selon les études de Wagner-Egger et coll. (2011) et de Mayor et coll. (2012) portant sur les perceptions publiques de la pandémie de grippe A (H1N1) qui a duré de 2009 à 2010, le Mexique était le pays le plus cité à titre de lieu d'origine du virus. Si les collectivités pauvres ont été présentées en tant que victimes de la maladie, elles étaient également tenues responsables de leur propre malheur, notamment en raison de leurs habitudes soi-disant peu hygiéniques (Mayor et coll., 2012). L'épidémie d'Ebola de 2014 et 2015 était perçue comme étant fondamentalement africaine par les Britanniques et les Américains (Joffe, 2011 ; Abeysinghe, 2016). Aux États-Unis, des immigrants africains ont vécu de la discrimination durant cette période. Des enfants ont été victimes d'intimidation dans des écoles de New York, du New Jersey et du Texas (Dionne & Turkmen, 2020). Selon Abeysinghe (2016), certains médias ont exagéré le rôle de certains facteurs culturels, tels que la consommation de la viande de brousse et les pratiques liées aux rites funéraires et aux soins des malades, dans la propagation du virus. Ceci a pu contribuer à la stigmatisation des populations affectées par l'épidémie. Ces représentations médiatiques ont généralement négligé d'autres facteurs « modernes », tels que les mouvements de population, les voyages et la concentration de la population dans les zones urbaines (Abeysinghe, 2016).

Plusieurs épidémies passées ont été nommées en fonction de régions du monde : grippe espagnole (H1N1) de 1918 à 1921, grippe asiatique (H2N2) de 1956 à 1958, grippe de Hong Kong (H3N2) de 1968 à 1970 (Joffe, 2011), grippe porcine mexicaine (H1N1) de 2009 et épidémie d'Ebola de 2014-2015 (Hoppe, 2016). L'association entre une maladie et un groupe « extérieur » permet souvent de se distancier de l'infection et de ses dangers (Joffe, 2011). Toutefois, le fait de nommer une épidémie en se référant à un lieu précis peut provoquer des stéréotypes, des préjugés ou des peurs irrationnelles envers la population concernée (Hoppe, 2016). Associer une maladie infectieuse à des groupes marginalisés peut également entraîner une réticence de la part des dirigeants à

prendre des mesures significatives pour lutter contre celle-ci de manière à protéger leurs communautés (Dionne & Turkmen, 2020).

Le blâme est également dynamique et fluctue en fonction de l'évolution d'une épidémie, alors que le public tente de comprendre ce qui se passe en scrutant les actions humaines qui ont pu contribuer à sa propagation (Roy et coll., 2020). Les personnes marginalisées de l'exogroupe ne constituent pas les seules cibles à être pointées du doigt pour une maladie. En effet, les perceptions des individus changent notamment selon l'emplacement géographique de l'épidémie (Roy et coll., 2020 ; Mondragon et coll., 2017 ; Mayor et coll., 2012). Lorsque la menace prend une dimension mondiale et que des individus appartenant à l'endogroupe commencent à être infectés par la maladie, de nouveaux comportements de blâme émergent (Mondragon et coll., 2017). Les autorités peuvent être accusées de ne pas avoir suffisamment agi face à l'épidémie afin de faire avancer leurs propres agendas (Mondragon et coll., 2017). Les individus infectés ou plus vulnérables faisant partie de l'endogroupe peuvent également être blâmés pour la propagation de la maladie (Mondragon et coll., 2017).

D'après Mayor et coll. (2012), les représentations du public à propos d'une épidémie évoluent en fonction de son « état ». Les individus sont portés à associer une maladie distante à des pays éloignés et tendent à associer une maladie omniprésente à des groupes plus susceptibles de la contracter (Mayor et coll., 2012). De plus, selon Roy et coll. (2020), qui ont analysé les conversations sur les réseaux sociaux à propos des entités accusées d'avoir causé ou répandu le virus Ebola en 2014-2016, le blâme était d'abord dirigé vers les populations touchées en Afrique. Au moment où l'épidémie menaçait de se propager au-delà du continent africain, le blâme s'est déplacé vers les gouvernements nationaux et les immigrants (Roy et coll., 2020). Enfin, toujours concernant l'épidémie d'Ebola, l'étude d'Abeyasinghe (2016) révèle un changement dans les représentations médiatiques du virus au fil du temps, passant des enjeux liés à la

contagion en Afrique de l'Ouest à une préoccupation croissante quant à sa propagation dans le monde occidental. En somme, lorsque l'épidémie est considérée comme étant géographiquement éloignée, le blâme se concentre sur les collectivités distantes, telles que les populations lointaines et leurs modes de vie. Ensuite, quand le virus est perçu comme étant une menace imminente ou présente, il est transféré à des collectivités locales, telles que les figures d'autorité ou les individus qui ont contracté la maladie (Mondragon et coll., 2017).

## **2.4 Attribution du blâme pour la COVID-19**

Le blâme durant la crise de la COVID-19 est également dynamique. Au début de la pandémie, l'attention était portée vers la Chine, premier pays à avoir établi des mesures sanitaires pour limiter la transmission du virus (Barreneche, 2020 ; Choli & Kuss, 2021 ; Ittefaq et coll., 2022 ; Labbé et coll., 2022). Lorsque le virus s'est propagé dans d'autres pays et que des protocoles de santé publique ont été instaurés, l'attention s'est tournée vers la responsabilité de la population, et plus particulièrement les groupes d'individus qui ne suivent pas les mesures sanitaires (Barreneche, 2020 ; Labbé et coll., 2022). Selon Choli et Kuss (2021), qui se sont penchés sur le blâme pour la crise de la COVID-19 auprès des utilisateurs du réseau social Twitter, les individus ont d'abord fait référence à la restriction de l'information et à certaines théories du complot dans les premières phases de la pandémie. Ils ont ensuite désigné les dirigeants politiques et les médias (Choli & Kuss, 2021). Les auteurs affirment que ces résultats divergent des études précédentes d'après lesquelles le blâme se déplace des collectivités distantes vers les collectivités locales en raison de la nature « décentralisée » des réseaux sociaux (Choli & Kuss, 2021). De plus, selon Labbé et coll. (2022) qui ont étudié le blâme et la stigmatisation à travers des caricatures éditoriales sur la COVID-19, les étrangers, les voyageurs et les citoyens étaient d'abord pointés du doigt pour l'origine de la pandémie. Le blâme s'est ensuite tourné vers les individus qui ne suivent pas les mesures sanitaires et les personnes qui remettent

en question les discours scientifiques sur la maladie (Labbé et coll., 2022). Si diverses entités ont été désignées comme responsables de la COVID-19, tels que le tourisme, les marchés aux animaux vivants et les individus qui ne suivent pas les mesures sanitaires, la Chine est demeurée la plus critiquée depuis l'apparition du virus (de Rosa & Mannarini, 2020 ; Choli & Kuss, 2021 ; Nguyen et coll., 2021). Selon Shangguan et coll. (2020), ce pays n'a pas réussi à freiner la propagation rapide de la COVID-19, malgré l'expérience acquise en gestion de crise de santé publique lors de la pandémie du SRAS et la mise en place d'un Centre pour le contrôle et la prévention des maladies (*Center for Disease Control and Prevention* ou CDC).

Certaines communautés ethniques, plus particulièrement les personnes d'ascendance asiatique, sont devenues des boucs émissaires de la pandémie de la COVID-19 (de Rosa & Mannarini, 2020 ; Martikainen et Sakki, 2020 ; Choli et Kuss, 2021 ; Ittefaq et coll., 2022 ; Voigt, 2022). Ces communautés sont considérées comme étant « à risque » en raison de « pratiques malsaines » ou de « comportements immoraux » et ont été accusées de propager intentionnellement la maladie (Martikainen et Sakki, 2020). Selon Hardy et coll. (2021), dont l'étude vise à comprendre les perceptions et les impacts de la pandémie de la COVID-19 sur la population américaine, les perspectives essentialistes culturelles ont pu contribuer à la stigmatisation des Asiatiques. L'essentialisme culturel est la pratique qui consiste à considérer la culture comme une catégorie absolue et exclusive à partir d'interprétations restreintes et souvent erronées (Hardy et coll., 2021). Par exemple, associer la culture chinoise et la consommation de chauve-souris relève de l'essentialisme culturel (Hardy et coll., 2021). Le blâme à l'égard des personnes d'ascendance asiatique s'est accompagné d'une forte augmentation des incidences de racisme, de discrimination et de violence envers ces individus (Barreneche, 2020 ; Dionne & Turkmen, 2020 ; Choli et Kuss, 2021 ; Nguyen et coll., 2021 ; Hardy et coll., 2021 ; Lin et coll. ; 2021 ; Rateau et coll., 2021 ; Ittefaq et coll., 2022 ; Labbé et coll., 2022). Ces incidents ont pris la

forme de refus de service, de commerces asiatiques boycottés et d'agressions physiques et verbales sur les lieux publics et les lieux de travail.

Toutefois, d'après une étude de Ferrante et coll. (2022) menée auprès de Canadiens durant la crise de la COVID-19, ce portrait des préjugés envers les Asiatiques doit être nuancé. En effet, la majorité des Canadiens ont des niveaux faibles de préjugés et ces niveaux demeurent stables au fil du temps (Ferrante et coll., 2022). Il existe également différents groupes de trajectoires de préjugés en fonction de caractéristiques sociodémographiques (Ferrante et coll., 2022). Par exemple, les personnes plus jeunes semblent plus disposées à appartenir au groupe ayant des niveaux faibles et décroissants et les conservateurs sont plus susceptibles de se trouver dans le groupe présentant la trajectoire la plus élevée (Ferrante et coll., 2021). Cette tendance peut être observée en politique et dans les médias.

L'ancien président des États-Unis Donald Trump et d'autres politiciens de droite se sont engagés dans une rhétorique de blâme envers la Chine, en désignant la COVID-19 comme étant un « virus chinois » ou un « virus de Wuhan » (Barreneche, 2020 ; Dionne & Turkmen, 2020 ; Choli & Kuss, 2021 ; Ittefaq et coll., 2022). Les mots « virus chinois » ont également été répétés par certains médias traditionnels de droite (Ittefaq et coll., 2022). Selon White (2020), la dénomination de la maladie comme étant « le virus chinois » et les violences physiques et verbales envers les personnes d'ascendance asiatique proviennent tous les deux d'une perception historique d'après laquelle les épidémies sont associées aux mouvements migratoires des peuples asiatiques. L'analyse du langage permet de comprendre comment le blâme, la désignation de boucs émissaires et les préjugés se manifestent et comment ces formes d'expression évoluent à travers le temps (Strong, 1990).

En résumé, le blâme et la désignation de boucs émissaires sont des processus complexes et structurés qui s'appuient sur des préoccupations sociales, politiques et idéologiques. En temps de crise, tel que la pandémie de la COVID-19, ils représentent des « stratégies » visant à retrouver une perception de contrôle sur la situation. Bien que certaines recherches aient étudié l'attribution « dynamique » du blâme durant la pandémie, il existe un manque important de connaissances sur leur manifestation et leur évolution au sein de la population. Le premier objectif de ce mémoire est de suivre l'évolution de la dynamique du blâme et de la désignation de boucs émissaires au cours des développements de la pandémie de la COVID-19 auprès de la population canadienne. De plus, les approches automatisées qui permettent d'analyser rapidement de grandes quantités de données textuelles sont peu utilisées en sciences sociales. Le deuxième objectif de ce mémoire est d'évaluer l'intérêt d'utiliser l'Allocation de Dirichlet latente dans l'étude de questions ouvertes d'enquêtes.

### **3. Méthodologie**

#### **3.1 Explication de la méthode de l'Allocation de Dirichlet latente**

##### **3.1.1 Modèle analytique du langage**

Le langage naturel correspond au langage humain, ou aux façons par lesquelles les individus s'expriment naturellement (Lyons, 1991). Il s'oppose au langage artificiel ou construit, comme en programmation informatique (Lyons, 1991). L'étude du langage naturel permet de comprendre les manières par lesquelles les personnes organisent et analysent le monde, et traitent et interprètent différentes informations (Tausczik & Pennebaker, 2010). Les mots employés dans la vie quotidienne reflètent leurs préoccupations, leurs opinions,

leurs préférences, leurs croyances et leurs réflexions actuelles (Strong, 1990). Ils sont également le reflet de leur éducation et de leur origine sociale (Bernstein, 1971). De plus, La longueur et la complexité des groupes lexicaux utilisés peuvent témoigner de la profondeur de leurs pensées (Tausczik & Pennebaker, 2010). Lorsqu'ils sont employés pour caractériser un événement, les termes évoluent à travers le temps, ce qui permet d'observer de quelle manière le sens octroyé par les individus à cet événement peut se transformer (de Saussure, 1983). La signification qu'on accorde à une expression ne réside pas dans sa seule utilisation, mais dépend également du contexte (de Saussure, 1983 ; DiMaggio et coll., 2013). En effet, les mots et leurs significations sont intrinsèquement liés. Le langage est un système de termes interdépendants dont le sens est déterminé par leurs relations avec d'autres mots (de Saussure, 1983 ; DiMaggio et coll., 2013).

Le langage a un rôle central dans les recherches en sciences sociales. En effet, dans la plupart des domaines de la vie, les individus communiquent à travers l'écrit ou la parole (Roberts et coll., 2014). Les entretiens de recherche qualitatifs, les analyses de contenu et les questions ouvertes provenant d'enquêtes permettent de l'étudier. Toutefois, il est moins courant de collecter et d'analyser des données textuelles provenant de questionnaires d'enquêtes que des données chiffrées. Ceci peut s'expliquer par le fait qu'elles sont plus difficiles à étudier (Roberts et coll., 2014). De plus, peu de chercheurs ont recours aux méthodes d'analyse automatique du texte ; la plupart opteront pour des techniques plus familières, telles que le codage manuel (Lindstedt et coll., 2019). Si les logiciels comme Excel, NVivo, Atlas.TI ou QDA Miner sont communément utilisés pour effectuer de l'analyse qualitative, le classement des informations doit tout de même être effectué manuellement. Autrement dit, chaque idée pertinente dans le texte doit être rattachée à la main au thème ou à la catégorie qui lui correspond le mieux, même en utilisant ces logiciels. L'utilisation de procédés automatisés est indispensable pour comprendre les perceptions et les expériences d'un grand



nombre de personnes, en particulier sur des enjeux nouveaux et évolutifs tels que la crise de la COVID-19. En effet, ceux-ci permettent d'explorer davantage les attitudes et les opinions décrites par les individus, comparativement aux méthodes d'analyse qualitative traditionnelles (Finch et coll., 2018).

Le traitement automatique de la langue (TLN), ou *natural language processing* (NLP) en anglais, correspond à l'étude du langage à travers des techniques computationnelles (Liddy, 2001). Le but du TLN est d'obtenir un traitement du langage semblable à celui de l'humain pour différentes tâches et applications (Liddy, 2001). Un procédé qui rend possible l'étude du langage naturel à l'aide de techniques automatisées est la fouille de textes (*text mining*). Elle comprend un ensemble de techniques visant à extraire des renseignements pertinents provenant d'une masse d'informations semi-structurées selon des critères de similarité et de nouveauté (Cheng et coll., 2022). Le texte est transformé en données numériques organisées en mots (ou termes) et en documents à l'aide de méthodes statistiques (Cheng et coll., 2022).

Les algorithmes de modélisation thématique (ou *topic models*) sont des outils de fouille de textes utilisés pour découvrir des thèmes abstraits dans une collection de documents (Blei et coll., 2003 ; Nelson, 2020 ; Finch et coll., 2018). Ces modèles statistiques sont réalisés au moyen de l'apprentissage non supervisé (*unsupervised machine learning*), une méthode qui regroupe automatiquement des éléments textuels dans des catégories/classes (Nelson, 2020). L'apprentissage non supervisé permet d'inférer des thématiques à partir des données brutes, plutôt que de les présupposer ou de les définir à l'avance (Roberts et coll., 2014).

Les algorithmes de modélisation thématique (*topic models*) prennent en considération le caractère relationnel du sens des mots, c'est-à-dire que la signification varie selon le contexte d'utilisation (DiMaggio et coll., 2013). En

effet, la présence simultanée de deux ou de plusieurs termes est primordiale dans l'extraction de thèmes significatifs (DiMaggio et coll., 2013 ; Nelson, 2020). Les algorithmes de modélisation thématique réduisent un corpus de texte complexe en groupes de mots plus simples et plus aisément interprétables (Nelson, 2020). Ils comprennent des processus automatisés pour le codage de très grandes quantités de textes (Mohr & Bogdanov, 2013 ; Nelson, 2020 ; Westrupp et coll., 2022). De plus, ils permettent d'analyser des documents sous un autre angle et de découvrir de nouvelles idées ou des concepts qui n'auraient pas été décelés par un codage manuel, tout en restant ancré dans les données (Mohr & Bogdanov, 2013 ; Nelson, 2020). Ce sont donc des méthodes qui offrent la possibilité d'intégrer des données de taille presque infinie, sans que le chercheur ait à fournir un travail beaucoup plus important (Mohr & Bogdanov, 2013 ; Nelson, 2020).

### **3.1.2 Inférence bayésienne**

L'inférence bayésienne est une méthode statistique couramment utilisée pour analyser le langage humain (DiMaggio et coll., 2013 ; Kelly et coll., 2019). Elle est fréquemment employée dans le domaine du TLN pour augmenter la précision et la performance des outils de compréhension du langage naturel. Les modèles bayésiens diffèrent des modèles fréquentistes, qui représentent une mesure de la fréquence d'un événement. En effet, contrairement aux modèles fréquentistes qui n'intègrent que les données observées, ils intègrent les connaissances précédentes (ou *a priori*) et les nouvelles informations (ou *a posteriori*) dans les distributions de probabilités (Lynch & Bartlett, 2019). Les connaissances précédentes sont des données qui peuvent provenir d'études antérieures ou de considérations théoriques, et les nouvelles informations correspondent aux données observées lors d'études ou d'expériences spécifiques (Lynch & Bartlett, 2019). En statistiques fréquentistes, une hypothèse nulle est souvent élaborée et testée pour déterminer sa plausibilité par rapport aux données observées. En statistiques bayésiennes, les informations disponibles sont

combinées aux croyances préalables pour produire de nouvelles conceptions ou connaissances sur les hypothèses et paramètres (Western, 1999 ; Jackman, 2009). Les procédures bayésiennes sont donc conceptuellement plus intuitives et plus flexibles que les méthodes fréquentistes (Western, 1999 ; Jackman, 2009).

Les modèles bayésiens sont moins couramment utilisés dans les études sociologiques que dans d'autres domaines des sciences sociales. Cela peut s'expliquer par le manque de formation des sociologues à ces méthodes et à leurs applications (Lynch & Bartlett, 2019). Pourtant, ils incorporent des informations sur les connaissances ou les croyances initiales dans l'étude de phénomènes sociaux, ce qui peut mener à des résultats plus précis et fiables. Les méthodes bayésiennes conviennent également aux modèles hiérarchiques dans lesquels les observations sont structurées en plusieurs niveaux imbriqués, tels que les données spatiales ou longitudinales (Western, 1999). En effet, ces modèles tiennent compte de l'incertitude à chaque niveau de la hiérarchie, puisque la probabilité *a priori* est considérée comme une variable aléatoire dont la valeur peut varier (Jackman, 2009). L'augmentation de la puissance de calcul des ordinateurs et la possibilité d'utiliser les méthodes Monte-Carlo par chaînes de Markov (MCMC) (techniques d'inférence qui permettent de générer des échantillons aléatoires à l'aide de paramètres) ont rendu l'application des modèles statistiques bayésiens plus accessible et efficace (Jackman, 2009).

### **3.1.3 L'Allocation de Dirichlet latente**

#### *3.1.3.1 Brève présentation*

L'Allocation de Dirichlet latente (ADL), une approche statistique provenant de la fouille de textes et du traitement automatique des langues, est l'algorithme de modélisation thématique le plus fréquemment utilisé (Blei et coll., 2003 ; Tudoran, 2018 ; Westrupp et coll., 2022). Elle permet notamment de

répondre aux défis liés à la conceptualisation des phénomènes en sciences sociales. En effet, certains concepts étudiés par les chercheurs, tels que le statut social, l'aliénation ou l'anomie, sont souvent complexes à définir et ne peuvent être directement observés (Boudon, 1962; McCutcheon, 1987). L'ADL extrait automatiquement les thèmes latents (ou *topics*) dans des textes, ce qui facilite la compréhension des concepts sous-jacents qu'ils contiennent. À partir des mots observés, les documents (chaque unité dans une collection de textes) sont affectés à des thématiques (Blei, 2012 ; Cheng et coll., 2022 ;). Plus précisément, le modèle analyse les mots présents dans les textes et cherche à découvrir des relations statistiques entre ces mots pour déterminer les thèmes, en examinant les combinaisons (ou co-occurrences) de mots qui sont souvent observés ensemble (Westrupp et coll., 2022). L'ADL a été développée par les professeurs David Blei, Michael I. Jordan et Andrew Ng et présentée dans leur article « *Latent Dirichlet Allocation* », publié en 2003. Il a été conçu pour répondre aux défis liés à la détection et à la classification de thèmes, notamment en raison de la grande quantité de données non structurées et de la complexité de la langue naturelle (Blei et coll., 2003).

*Latent* signifie que les thématiques sont « non observées », et que l'algorithme a pour fonction de regrouper une quantité de mots en des catégories qui devront être analysées par le chercheur (Scarpino et coll., 2021). *Dirichlet* provient de la Loi de Dirichlet utilisée pour générer les distributions document/thématique et les distributions mots/thématique (Scarpino et coll., 2021). *Allocation* indique que l'ADL attribue des thématiques à une sélection de textes (Scarpino et coll., 2021). Les thématiques (ou *topics*) sont des distributions sur un vocabulaire donné (Blei, 2012 ; Tudoran, 2018). Cela signifie que certains mots ont une probabilité plus élevée d'apparaître dans certaines thématiques que dans d'autres. Par exemple, le thème « voyage » peut avoir une probabilité plus élevée de comporter les termes « international », « monde » ou « frontières ». Les

thématiques découvertes par l'ADL sont également communes à l'ensemble du corpus de textes analysé.

L'ADL repose sur les statistiques bayésiennes pour estimer les paramètres des modèles. Il appartient aux modèles de mélange, car il utilise différentes distributions de probabilité (Dirichlet et multinomiale) pour représenter les données (Blei et coll., 2003). La distribution de Dirichlet décrit la répartition des thématiques dans les documents et des mots dans les thématiques, alors que la distribution multinomiale représente la répartition des mots dans les documents (Pietsch & Lessman, 2018). L'ADL est également considérée comme un modèle de mélange infini, puisque la distribution de Dirichlet permet de générer un nombre illimité de combinaisons possibles de thématiques dans les documents (Blei, 2003). Enfin, elle fait partie des modèles à *mixed membership*, car les observations peuvent être associées à plusieurs groupes simultanément (DiMaggio et coll., 2013). En effet, l'idée derrière la modélisation de l'ADL est que chaque document peut comporter plusieurs thématiques et avoir des probabilités différentes d'appartenir à ces thématiques (Blei, 2012). Chaque document est traité comme un ensemble de thématiques et chaque thématique correspond à un ensemble de mots (Blei, 2012).

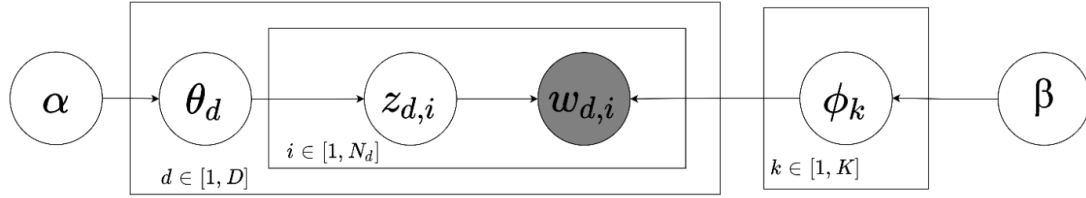
La méthode de l'ADL a été employée dans une variété de recherches en sciences sociales. Tudoran (2018) l'a utilisée pour étudier les perceptions de 4093 individus à l'égard des publicités en ligne et du blocage publicitaire. Westrupp et coll. (2022) ont examiné 918 publications sur le site Reddit pour comprendre quels sont les problèmes parentaux les plus fréquemment discutés sur cette plateforme. Alvarez-Melis et Saveski (2016) ont développé une nouvelle technique d'analyse des *tweets* à l'aide de l'ADL. Cheng et coll. (2022) l'ont utilisée pour obtenir un résumé de la littérature existante sur la COVID-19, le coronavirus lié au syndrome respiratoire du Moyen-Orient (MERS) et le SRAS, afin de soutenir le milieu médical. Ils ont étudié plus de 50 000 textes sur

la plateforme *CORD-19*, un ensemble de données ouvertes sur la COVID-19, le MERS et le SRAS. Enfin, Finch et coll. (2018) ont analysé des questions ouvertes avec l'ADL dans le but de comprendre les pensées et les réflexions de 620 professionnels en milieu scolaire concernant l'automutilation.

#### 3.1.4.2 Variables modélisées dans l'Allocation de Dirichlet latente

L'**illustration 1** est une reproduction graphique des variables modélisées dans l'ADL (Harel et coll., 2023). Les cercles représentent les variables aléatoires et la direction des flèches indique la dépendance qui existe entre elles (Harel et coll., 2023). Les cercles non ombragés correspondent aux variables aléatoires latentes alors que le cercle ombragé décrit la variable aléatoire observée (Harel et coll., 2023). Un rectangle est utilisé pour illustrer les arguments qui se répètent dans la modélisation, ce qui évite de les représenter individuellement (Blei, 2012). Chaque rectangle correspond à un objet modélisé par l'ADL ;  $d$  représente la distribution des thématiques par document,  $i$  désigne l'attribution d'un mot à une thématique et  $k$  décrit la distribution des mots par thématique (Harel et coll., 2023). Le caractère alpha ( $\alpha$ ) représente le paramètre Dirichlet pour la répartition des thèmes dans un document (Blei et coll., 2003). La proportion thématique par document est décrite par  $\theta$ , l'assignation d'une thématique pour un mot correspond à  $z$ , le mot observé équivaut à  $w$  et  $\phi$  représente la distribution des mots par thèmes (Harel et coll., 2023). Beta ( $\beta$ ) correspond au paramètre Dirichlet pour la répartition des mots dans une thématique (Blei et coll., 2003). Un paramètre supplémentaire ( $N$  ou le nombre d'itérations de Gibbs) est utilisé pour l'inférence statistique dans la modélisation de l'ADL. Les informations définies avant l'analyse sont appelées des hyperparamètres (Panichella, 2021). Par exemple, le nombre de thématiques à découvrir représente un hyperparamètre à spécifier avant de produire l'ADL. Les paramètres  $\alpha$  et  $\beta$  fixent des données initiales sur les probabilités des thématiques et des mots dans les documents avant d'être ajustés en fonction des données réelles.

### Illustration 1. Représentation graphique du modèle<sup>3</sup>



Les algorithmes derrière l'ADL sont dérivés des modèles bayésiens, où  $\phi$ ,  $\theta$  et  $z$  sont des variables latentes provenant d'un modèle probabiliste hiérarchique (DiMaggio et coll., 2013). L'objectif est de les définir à partir des variables observées, c'est-à-dire une collection observée de documents (DiMaggio et coll., 2013). Puisque la structure de l'ADL permet à la variable observée d'interagir avec les variables latentes, il est possible de déduire leur distribution *a posteriori* (Harel et coll., 2023). Cette interaction est représentée par un processus génératif qui produit les données observées.

Le processus génératif de l'ADL est le suivant<sup>4</sup> :

1. Pour chaque thématique  $k = \{1, \dots, K\}$ ,
  - Choisir une distribution sur les mots  $\phi_k \sim \text{Dir}(\beta)$
2. Pour chaque document  $d$ 
  - Choisir un vecteur de proportions de thématiques par document  $\theta_d \sim \text{Dir}(\alpha)$
  - Pour chaque mot  $i$  dans le document  $d$ 
    - Choisir une attribution de thématiques par mot  $z_{d,i} \sim \text{Mult}(\theta_d)$ ,  $z_{d,i} \in \{1, \dots, K\}$
    - Choisir un mot  $w_{d,i} \sim \text{Mult}(\phi_{z_{d,i}})$ ,  $w_{d,i} \in \{1, \dots, V\}$

<sup>3</sup> Harel et coll., 2023, p.6

<sup>4</sup> Harel et coll., 2023, p.7

Voici une explication simplifiée du processus génératif. Tout d'abord, l'ADL suppose que chaque document est composé d'un mélange de thématiques différentes (Müller & Guido, 2016). Par exemple, une collection de textes sur la pandémie de la COVID-19 peut contenir les thèmes « santé publique » et « économie ». L'algorithme infère une distribution de probabilité sur les mots du texte en se basant sur leurs co-occurrences (Pietsch & Lessmann, 2018). En reprenant l'exemple précédent, « santé publique » peut comporter les termes « vaccination, quarantaine, mesures de prévention », car ils sont souvent observés ensemble dans le texte. De même que le thème « économie » peut inclure les termes fréquemment observés « faillite, licenciement, récession ». Le mot « vaccination » a donc une probabilité plus élevée d'être observé avec des termes relevant de la thématique « santé publique » que de l'« économie ». Ensuite, l'algorithme détermine une distribution qui représente la proportion de chaque thématique dans chaque document (Müller & Guido, 2016). Un document sur la pandémie peut avoir les répartitions suivantes : 40 % sur la santé publique, 30 % sur la politique, 20 % sur l'économie, 10 % sur la technologie. Par la suite, l'algorithme associe une thématique à chaque mot (Pietsch & Lessmann, 2018). Le terme « vaccination » serait rattaché à « santé publique » et le terme « faillite » à l'« économie ». Enfin, l'algorithme détermine les mots associés à chaque thématique en fonction de leurs distributions dans la thématique (Pietsch & Lessmann, 2018).

Le but de ce processus génératif est d'estimer les deux matrices suivantes<sup>5</sup> :

1. La distribution des thématiques par document ( $\Theta$ )
2. La distribution des mots par thématique ( $\Phi$ )

Les matrices  $\Theta$  et  $\Phi$  permettent respectivement d'évaluer la pertinence de chaque document dans chaque thématique et d'identifier les termes les plus

---

<sup>5</sup> Harel et coll., 2023, p.7



couramment utilisés dans chaque thème. (Blei et coll., 2003 ; Pietsch & Lessmann, 2018). Ces deux matrices sont estimées afin de mieux comprendre les thèmes latents contenus dans les documents ainsi que les mots qui y sont associés.

#### 3.1.4.3 Avantages de l'Allocation de Dirichlet latente

En tant qu'algorithme de modélisation thématique, l'ADL comporte plusieurs avantages. Elle a la capacité de diminuer les biais du chercheur, puisque la méthode est entraînée sur les données (Blei, 2012 ; Cheng et coll., 2022 ; Roberts et coll., 2014 ; Westrupp et coll., 2022). En effet, cette méthode se base directement sur les propriétés statistiques des mots dans les textes pour former les thèmes, ce qui permet d'obtenir une analyse plus objective. De plus, elle est utilisée pour sa capacité d'organiser et de comprendre de très grands ensembles de textes non structurés, une tâche que les analyses qualitatives traditionnelles ne peuvent faire adéquatement (Cheng et coll., 2022 ; Westrupp et coll., 2022). Plus le nombre de textes augmente, plus le codage à la main est difficile et long. En revanche, les résultats de l'ADL sont plus fiables, précis et détaillés lorsque ce nombre s'élève.

L'ADL comporte certains avantages par rapport à d'autres méthodes de traitement automatique du langage, tels que l'analyse sémantique latente (ALS – en anglais *Latent Semantic Analysis* ou LSA) et l'analyse sémantique latente probabiliste (ALSP – en anglais *Probabilistic Latent Semantic Analysis* ou PLSA) (Blei et coll., 2003).

L'ALS, également appelée indexation sémantique latente (LSI, de l'anglais *Latent Semantic Indexation*, a été développée par Deerwester, Dumais, Furnas, Landauer et Harshman en 1990 en tant que méthode d'indexation et de recherche automatique. Cette approche est basée sur l'idée que les mots dotés de significations similaires ont tendance à apparaître dans des contextes semblables

(Williams & Betak, 2018). Elle simplifie les données en regroupant les mots similaires dans des concepts sémantiques (Williams & Betak, 2018). Cependant, l'ALS suppose une distribution gaussienne des termes dans les documents, alors que dans la réalité, les données textuelles peuvent présenter des distributions différentes (Hofmann, 1999). Cela peut entraîner des estimations moins précises et une perte de sensibilité dans la modélisation des relations entre les termes et les documents (Hofmann, 1999). Contrairement à l'ALS, l'ALSP et l'ADL se basent sur une approche probabiliste et permettent de modéliser des distributions de termes plus flexibles (Barde & Bainwad, 2017). Un autre inconvénient de l'ALS est son manque de prise en compte de la polysémie, c'est-à-dire le fait qu'un terme peut avoir plusieurs significations différentes (Hofmann, 1999). L'ALS traite les mots comme des entités uniques sans distinguer les différents sens possibles, ce qui peut entraîner une confusion dans la représentation des concepts sémantiques (Niraula et coll., 2013). Au lieu de représenter chaque terme par un point unique dans un espace sémantique, l'ALSP et l'ADL attribuent une distribution de probabilités à chaque mot pour chaque thème (Niraula et coll., 2013). Chaque mot peut être associé à plusieurs thèmes et chaque thème est lié à un groupe de mots qui ont des significations similaires.

L'ALSP, également appelée indexation sémantique latente probabiliste (PLSI, de l'anglais *Probabilistic Latent Semantic Indexing*), est une méthode statistique pour l'analyse de textes (Barde & Bainwad, 2017). Elle a été introduite par Hofmann (1999) afin d'améliorer la représentativité des thèmes latents dans les données textuelles. Plutôt que d'utiliser des matrices comme dans l'ALS, l'ALSP utilise une méthode probabiliste pour représenter les relations entre les mots et les sujets dans un corpus de textes (Barde & Bainwad, 2017). Elle se fonde sur un modèle de mélange (*mixture model*) dérivé des classes latentes (Liu et coll., 2011). Toutefois, le nombre de paramètres à modéliser dans l'ALSP augmente avec le nombre de documents à traiter (Blei et coll., 2003). Cela signifie que le modèle devient de plus en plus complexe à mesure que la taille du

corpus augmente, ce qui peut mener à des phénomènes de surapprentissage (*overfitting*) (Blei et coll., 2003). Contrairement à l'ADL, l'ALSP est incapable de générer de nouveaux documents basés sur les propriétés statistiques du corpus et n'effectue que de la classification d'information (Crain et coll., 2012). Son rôle se limite à classer les documents existants dans des thèmes déjà identifiés. L'ADL utilise un petit nombre de paramètres qui suivent la distribution de Dirichlet, ce qui réduit les risques de surajustement et permet une affectation plus précise des documents aux thématiques (Liu et coll., 2011 ; Crain et coll., 2012). En tant que modèle génératif, elle permet d'intégrer de nouvelles données reflétant les caractéristiques statistiques du corpus de référence. L'ADL est plus robuste et reproductible dans différentes conditions et sur différents ensembles de données (Blei et coll., 2003). Depuis son introduction en 2003, la méthode de l'ADL a été beaucoup plus largement utilisée que l'ALSP (Crain et coll., 2012).

Dans le contexte de la présente étude, l'ADL a été choisie pour sa capacité à analyser un grand nombre de réponses en relativement peu de temps et à découvrir des thématiques qui n'auraient pas été décelées par le chercheur. L'ADL peut également détecter des nuances et des relations dans les données qui peuvent ne pas être décelées par le codage manuel (Nelson, 2020). En effet, contrairement à cette méthode, qui emploie une approche déductive guidée par la théorie, l'ADL utilise une approche inductive pour découvrir des modèles et des thèmes à partir des données brutes (Lindstedt, 2019).

## 3.2 Application de la méthode de l'Allocation de Dirichlet latente

### 3.2.1 Prétraitement des données

Avant de produire l'ADL, le corpus de texte doit être préparé, nettoyé et prétraité afin d'obtenir une meilleure qualité des données et une plus grande efficacité du processus d'analyse (Denny & Spirling, 2018 ; Alam & Yao, 2019). Ce procédé réduit la quantité de termes à étudier en écartant les éléments non pertinents, ce qui diminue le temps de calcul nécessaire à la modélisation et le « bruit » dans les données (Denny & Spirling, 2018). Enfin, il facilite l'interprétation des modèles, puisque les mots qui demeurent sont les plus importants pour l'analyse (Denny & Spirling, 2018). Par ailleurs, les participants ont tendance à rédiger leurs réponses plus rapidement lors d'enquêtes par questionnaire menées en ligne, ce qui provoque davantage de fautes de frappe et une plus grande utilisation d'expressions courtes (Reja et coll., 2003). Le prétraitement des données est donc essentiel pour garantir la qualité des réponses obtenues.

Tout d'abord, les fautes d'orthographe ont été corrigées à l'aide du logiciel *Antidote* et les réponses en français ont été traduites en anglais à l'aide du logiciel *DeepL* (Lane, 2019 ; DeepL, 2021 ; Antidote, 2022). Ensuite, certaines expressions ont été remplacées pour obtenir une certaine uniformisation (ex. : *U.S.*, *U.S.A.* et *America* sont devenus *United States*) (Lane, 2019). Par la suite, certaines formulations fréquentes ont été réduites en un seul mot (ex. : *World Health Organization*<sup>6</sup> est devenu *worldhealthorganization*, et *United States* est

---

<sup>6</sup> Acronyme pour *World Health Organization* ou Organisation mondiale de la Santé (OMS)

devenu *unitedstates*). Les termes remplacés et réduits se trouvent dans l'annexe 3. Les « mots vides » (en anglais *stopwords*), tels que les déterminants (ex. : *the, a, an*) et les prépositions (ex. : *above, across*), ont également été supprimés (NLTK, 2010 ; Denny & Spirling, 2018 ; Lane, 2019).

Les mots de négations, tels que *not* et *don't*, ont également été exclus, en raison de leur nature peu informative à la tâche de catégorisation du texte (Duwairi & El-Orfali, 2014). Dans la présente étude, le but de l'analyse est de détecter les thématiques dans des documents, et ils peuvent interférer avec cette tâche en compliquant l'interprétation des résultats. En effet, il est possible que ces termes apparaissent dans les listes de mots qualifiant les thématiques, alors qu'ils ne contribuent pas à une meilleure compréhension des thèmes. Toutefois, dans le cadre de l'analyse de sentiment (*sentiment analysis*), qui consiste à extraire des émotions positives, négatives ou neutres, inclure les négations revêt une importance particulière (Duwairi & El-Orfali, 2014). L'annexe 4 présente la liste des mots vides supprimés. La ponctuation et les nombres ont été retirés, et tous les caractères ont été mis en minuscule (Lane, 2019). Les termes qui n'apparaissent que très rarement dans les réponses des participants ont été conservés (Denny & Spirling, 2018). En effet, dans le cadre de la présente étude, écarter les mots qui apparaissent dans moins de 0,5 % à 1 % des documents réduirait considérablement la taille du corpus de réponses à analyser (Denny & Spirling, 2018). Celui-ci contient un grand nombre de termes différents et conserver ces termes rares permet d'éviter une perte d'information importante dans les listes de mots qui composent les thématiques.

Finalement, les termes conservés ont été « lemmatisés », c'est-à-dire qu'ils ont été ramenés à leur base significative (ex. : *changes, changing* et *changed* sont devenus *change*) (Lane, 2019). Dans le cadre de la présente étude, les expressions *Chinese* et *China* n'ont pas été lemmatisées afin de conserver les nuances liées à l'utilisation de ces deux termes et en raison de la présence

récurrente de ces deux mots dans les documents. En effet, le terme *China* fait référence au pays, alors que *Chinese* peut référer à différents éléments, tels que le gouvernement chinois (*chinese government*), les marchés aux animaux vivants (*chinese wet markets*), les villes chinoises (*chinese towns*) et les laboratoires chinois (*chinese laboratory*). Les étapes énumérées ci-dessus ont été réalisées à l'aide des bibliothèques *textmineR* et *tidyverse* de la version 4.0.3 de l'environnement *RStudio* (Jones, 2021 ; Harel et coll., 2023). **L'illustration 2** contient des exemples de documents avant le prétraitement des données et **l'illustration 3** présente ces mêmes exemples après le prétraitement.

**Illustration 2. Exemples de documents avant le prétraitement des données**

Document1	unsafe sanitary conditions dealing with live animals in a marketplace in a different country
Document2	the consumption of certain animals in China
Document3	China not being open about the seriousness of the problem at the start.
Document4	not knowing that eating parts of certain animals might cause Covid

**Illustration 3. Exemples de documents après le prétraitement des données**

Document1	unsafe sanitary conditions dealing live animals marketplace different country
Document2	consumption certain animals China
Document3	China open seriousness problem start.
Document4	knowing eating parts certain animals cause Covid

Le corpus prétraité a ensuite été converti en une matrice des termes-documents pour les fins de l'analyse. En effet, celle-ci est utilisée en entrée pour produire l'ADL. Dans cette matrice, les lignes représentent chaque document (une réponse ouverte par participant) et les colonnes correspondent à chaque mot

du corpus (Westrupp et coll., 2022). La matrice des termes-documents distingue tous les ensembles de termes présents dans les documents (1) et qui en sont absents (0), ce qui peut générer une matrice creuse (*sparse matrix*) contenant beaucoup de zéros (voir l'exemple de **l'illustration 4**) (Crain et coll., 2012). En effet, la plupart des documents ne comportent qu'une fraction de la totalité des mots présents dans l'ensemble des textes analysés.

**Illustration 4. Représentation d'un DTM**

	animal	certain	China	consumption	Food	Market	unsafe	Wet
Doc 1	1	0	0	0	0	0	1	0
Doc 2	1	1	1	1	0	0	0	0
Doc 3	0	0	1	0	0	0	0	0
Doc 4	1	1	0	0	0	0	0	0

L'ADL est une technique de réduction de dimensionnalité, c'est-à-dire qu'il transforme des données textuelles d'un espace de grande dimension à un espace de plus petite dimension (Crain et coll., 2012). Cela permet de convertir une matrice creuse en une représentation plus synthétique et significative. Dans le cadre de l'étude de texte, il s'agit de regrouper des mots similaires qui correspondent aux concepts déterminants de l'ensemble des documents examinés. Ce processus élimine le bruit dans les données afin de conserver le contenu sémantique le plus important. Il est effectué d'une manière probabiliste, où « le document  $d$  est représenté par sa distribution des thématiques » (Crain et coll., 2012, p.150).

### 3.2.2 Modélisation avec l'ADL

Une fois le prétraitement des données effectué, la prochaine étape a consisté à produire l'ADL (Lane, 2019). Un script optimisé par Harel et coll. (2023) dans l'environnement *RStudio* a été utilisé à cette fin. Des intervalles de valeur ont été spécifiés dans le code pour paramétrer alpha ( $\alpha$ ), bêta ( $\beta$ ) et le nombre d'itérations de Gibbs ( $N$ ). Une méthode d'optimisation combinant l'algorithme génétique avec la fonction objective « cohérence » a été employée (Harel et coll., 2023). L'algorithme génétique permet de sélectionner le modèle le plus performant, et la fonction objective « cohérence » est une mesure de la qualité et de la précision du modèle. La cohérence d'une thématique indique à quel point les mots qui le composent sont liés entre eux, et la cohérence globale permet d'évaluer la qualité de l'ensemble des thématiques générées (Harel et coll., 2023). Ces étapes ont été effectuées à l'aide de la librairie *textmineR* et de diverses fonctions et utilités de l'ADL fournies par le script (Harel et coll., 2023)<sup>7</sup>.

L'ADL a été programmée pour produire 25 itérations à chaque temps de mesure. Le choix du modèle optimal s'est appuyé sur une valeur de stabilité des itérations de l'ADL, c'est-à-dire la capacité à reproduire des distributions semblables en générant d'autres modèles (Harel et coll., 2023). Si de grands corpus nécessitent de 50 à 100 thématiques pour couvrir la variété du contenu, les réponses à une question ouverte n'en nécessitent que quelques-uns (Roberts et coll., 2014).

---

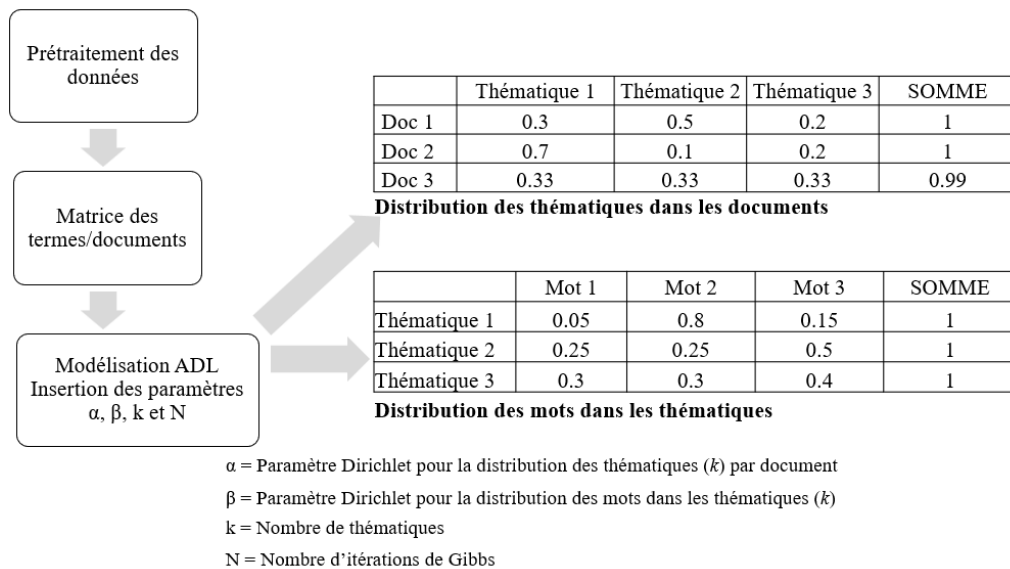
<sup>7</sup> Le tableau 9 présente les intervalles de valeurs utilisés dans le cadre de cette étude pour chaque paramètre, alors que le tableau 10 présente les paramètres des modèles sélectionnés (nombre de thématiques, nombre d'itérations de Gibbs, alpha et bêta) et les valeurs de stabilité. Ces tableaux ont été mis en annexe en raison de leur nature technique.



Pour déterminer le nombre de thématiques ( $k$ ), la procédure suivante a été appliquée. Tout d'abord, un intervalle de valeurs précis a été spécifié pour ce paramètre. Cet intervalle a été fixé entre 7 et 20 en raison de la taille du corpus. Ensuite, la valeur de  $k$  ayant la stabilité la plus élevée a été utilisée pour déterminer le nombre de thématiques final (Harel et coll., 2023). Finalement, le nombre optimal de thématiques pour chaque temps de mesure a été établi à  $k=9$ , basé sur la valeur de stabilité des modèles. Les thématiques ont été également nommées en fonction des listes de termes qui les qualifient. Pour les fins de l'analyse, une réponse ouverte d'un participant à un temps de mesure est considérée comme un document. Un document représente donc une phrase ou une expression écrite par un répondant, à tel temps de mesure.

**L'illustration 5** présente un résumé des étapes d'application de l'ADL. Les données sont d'abord prétraitées en suivant différentes procédures, telles que la traduction du texte, la correction des erreurs d'orthographe, le remplacement de termes, la suppression des mots vides, la réduction des caractères en minuscule et la lemmatisation. Le corpus prétraité est ensuite converti en une matrice des termes-document (DTM). L'étape subséquente correspond à insérer les paramètres  $\alpha$ ,  $\beta$ ,  $k$  et  $N$  et d'appliquer le code. Par la suite, l'ADL produit deux matrices : la distribution des thématiques dans les documents et la distribution des mots dans les thématiques.

**Illustration 5. Résumé des étapes d'application de l'ADL<sup>8</sup>**



<sup>8</sup> Adapté de Shao et coll. (2021)

### **3.3 Enquête sur les conséquences sociales de la crise de la COVID-19**

Les données proviennent d'une enquête longitudinale intensive par questionnaire intitulée « COVID-19 Canada : La fin du monde tel qu'on le connaît ? » menée par le Laboratoire sur les changements sociaux et l'identité de l'Université de Montréal. Le but de la recherche est d'étudier les conséquences sociales de la pandémie de la COVID-19 auprès de la population canadienne (de la Sablonnière et coll., 2020). Les participants de l'enquête ont été sélectionnés sur la base de quotas d'âge, d'identité de genre et de province de résidence (de la Sablonnière et coll., 2020). Ceux-ci avaient la possibilité de répondre au questionnaire en français ou en anglais. Ils ont signé un formulaire de consentement avant d'y avoir accès (voir annexes 5 et 6). Remplir le questionnaire prenait entre 15 et 20 minutes et l'enquête se faisait en ligne sur un ordinateur, un téléphone ou une tablette (de la Sablonnière et coll., 2020). Les participants recevaient une compensation d'environ 2,50 \$ par enquête complétée, sous la forme de points échangeables dans la chaîne de magasins de leur choix (de la Sablonnière et coll., 2020). Les individus ayant répondu à un premier questionnaire étaient ensuite invités à remplir les prochains questionnaires (de la Sablonnière et coll., 2020).

La collecte de données a été réalisée à l'aide du panel Web *AskingCanadians* de Delvinia. C'est un panel de type *opt-in* de participants canadiens qui permet d'intégrer des critères de centaines de variables « démographiques, psychographiques, comportementales et attitudinales afin de répondre aux besoins spécifiques des chercheurs en matière d'échantillonnage » (*AskingCanadians*, n.d.) L'enquête a été effectuée avec un design de type « roulant ». Chaque jour, un sous-échantillon d'environ 250 individus a été contacté pour remplir le questionnaire jusqu'à la fin de la vague de collecte (de

la Sablonnière et coll., 2020). Cette méthode permet de produire des analyses qui « capturent les effets des événements en temps réel » (de la Sablonnière et coll., 2020, p.11), dans un contexte où la situation liée à la pandémie a évolué rapidement.

Parmi une série de questions portant sur les expériences et les attitudes des individus dans le contexte de la pandémie, les participants ont répondu à la question ouverte « Qui ou quoi est principalement responsable de la crise actuelle de la COVID-19 ? » Cette question a été posée à six reprises durant l'étude longitudinale (aux temps 1, 3, 5, 8, 10 et 12). Les informations sur les temps de mesure se trouvent dans le **tableau 1**. L'enquête a été menée auprès d'un échantillon représentatif de 3617 Canadiens au premier temps de mesure. La taille de l'échantillon diminue par la suite.

**Tableau 1. Informations sur les temps de mesure des enquêtes**

<b>Temps de mesure</b>	<b>Dates de l'enquête</b>	<b>Taille de l'échantillon</b>	<b>Nombre de réponses à la question ouverte</b>	<b>Durée qui sépare les différents temps de mesure</b>
1	6 avril – 6 mai 2020	3 617	2 869	2 semaines
2	21 avril – 13 mai 2020	2 282	-	2 semaines
3	4 mai – 25 mai 2020	2 369	2 230	2 semaines
4	18 mai – 10 juin 2020	2 296	-	2 semaines
5	1 <sup>er</sup> juin – 23 juin 2020	2 154	1 999	2 semaines
6	15 juin – 13 juillet 2020	2 116	-	2 semaines
7	13 juillet – 8 août 2020	2 072	-	4 semaines
8	17 août – 13 septembre 2020	1 871	1 778	5 semaines
9	21 septembre – 19 octobre 2020	1 821	-	5 semaines
10	26 novembre – 29 décembre 2020	1 883	1 758	9 semaines
11	13 avril – 31 mai 2021	2 002	-	20 semaines
12	2 mars – 19 avril 2022	1 672	1 596	46 semaines

Le format ouvert de cette question a été privilégié afin de recueillir des réponses spontanées sur le sujet sans être limité ou orienté par des choix prédéfinis (Fink, 2003). La question ouverte représente un moyen de collecter des informations détaillées dans les propres mots du participant (Fink, 2003) et d’avoir un aperçu de leur pensée (Roberts et coll., 2014). Elle offre donc la possibilité d’obtenir des réponses plus diversifiées que la question fermée (Reja et coll., 2003). La question ouverte réduit les biais potentiels causés par les choix de réponses préconçus, qu’il s’agisse de l’ordre de présentation, de la formulation ou du nombre de choix proposés (Reja et coll., 2003). De plus, ce format a été sélectionné en raison de sa capacité à explorer un sujet nouveau et complexe (Fink, 2003), en l’occurrence, l’attribution de la responsabilité pour la crise de la COVID-19. En effet, la question ouverte permet de découvrir des aspects insoupçonnés ou méconnus qui pourraient être absents des questions fermées.

## **4. Résultats**

### **4.1 Caractéristiques sociodémographiques des répondants**

Le **tableau 2** présente la répartition des répondants à la question ouverte et dans l’échantillon total selon le pourcentage de femmes et la moyenne d’âge. Les proportions de femmes sont similaires à chaque temps de mesure pour chacun des échantillons. La moyenne d’âge des participants augmente à chaque temps de mesure, et demeure plus élevée chez les participants à la question ouverte que dans l’échantillon total. Pour ce qui est des catégories d’âge, les proportions des participants de 18 à 24 ans diminuent entre les temps 1 et 12, au profit des participants âgés de 65 ans et plus (voir le **tableau 11** en annexe).

**Tableau 2. Répartition des répondants selon le pourcentage de femmes et la moyenne d'âge (Question ouverte et échantillon total)**

Temps de mesure	Question ouverte		Échantillon total	
	% Femmes	Moyenne d'âge	% Femmes	Moyenne d'âge
1	50,9 %	48,6	50,5 %	47,6
3	50,0 %	49,5	49,2 %	48,8
5	49,8 %	50,2	48,7 %	49,3
8	49,9 %	51,1	49,4 %	50,4
10	49,2 %	51,3	48,4 %	50,3
12	49,8 %	52,9	50,0 %	53,0

En raison de l'utilisation d'un design de type « roulant », il n'a pas été possible d'ajouter le niveau d'éducation dans les quotas de représentativité (de la Sablonnière et coll., 2020). Cela signifie que les participants titulaires d'un baccalauréat ou d'un diplôme supérieur sont surreprésentés dans l'échantillon par rapport à la population canadienne. Le **tableau 3** présente la répartition des répondants de l'enquête COVID-19 selon le plus haut niveau d'éducation atteint et les proportions correspondantes à la population canadienne selon le recensement de Statistique Canada en 2016 (Statistique Canada, 2019). Il convient de souligner que certains participants de l'enquête (2,1 %) ont indiqué n'avoir aucune scolarité, mais ont tout de même répondu à un questionnaire écrit.

**Tableau 3. Répartition des répondants de l'enquête COVID-19 selon le plus haut niveau d'éducation atteint et recensement de 2016**

Niveau d'éducation	Enquête COVID-19	Statistique Canada 2016
Aucune scolarité	2,1 %	11,5 %
École secondaire terminée	15,2 %	23,7 %
Collège ou CÉGEP	21,0 %	33,2 %
Quelques études universitaires	7,5 %	3,1 %
Baccalauréat	33,6 %	20,9 %
Maîtrise	13,3 %	5,9 %
Diplôme professionnel ou doctorat	7,4 %	1,7 %

Pour conclure cette section sur les caractéristiques sociodémographiques, le **tableau 4** présente le nombre de participants ayant répondu à l'enquête, le nombre de participants n'ayant pas répondu à la question ouverte et le nombre d'individus qui n'ont pas répondu à l'enquête, par temps de mesure. Au fil du temps, il y a une diminution progressive de la participation à l'enquête longitudinale, et par extension, à la question ouverte. Ce sont les individus plus jeunes (entre 18 et 24 ans) qui semblent être les moins enclins à revenir aux enquêtes ultérieures (voir le **tableau 11**). Il y a également une baisse du nombre de participants qui répondent au questionnaire, mais qui ne répondent pas à la question ouverte, entre le temps 1 et le temps 3. Cette tendance peut être attribuée au fait que les participants qui reviennent aux enquêtes ultérieures sont les plus motivés et les plus intéressés, ce qui augmente la propension à répondre à la question (Goulias et coll., 1991 ; Salthouse, 2014). Le taux d'attrition par temps de mesure varie entre 22,3 % au temps 1 et 1,1 % au temps 10 et le taux d'attrition cumulatif est de 43,5 % au temps 12.



**Tableau 4. Participation à l'enquête et taux d'attrition**

Temps de mesure	A répondu à l'enquête	N'a pas répondu à la question ouverte	N'a pas répondu à l'enquête	Taux d'attrition par temps de mesure	Taux d'attrition cumulatif
1	2869	746	0	-	-
3	2230	57	1328	22,3%	22,3 %
5	1999	65	1551	10,3 %	30,3 %
8	1778	39	1798	11,1 %	38,0 %
10	1759	43	1813	1,1 %	38,6 %
12	1620	33	1962	7,9 %	43,5 %

## **4.2 Quelques événements importants liés à la pandémie de la COVID-19**

Le **tableau 5** présente les événements importants de la pandémie de la COVID-19 en fonction des temps de mesure durant lesquels la question ouverte a été posée. Avant le premier temps de mesure (du 6 avril au 6 mai 2020), plusieurs événements importants ont eu lieu. Le premier cas de COVID-19 s'est déclaré à Wuhan, en Chine, le 8 décembre 2019 (Shangguan et coll., 2020). Le 11 mars 2020, l'Organisation mondiale de la Santé a décrété l'état de pandémie (Labbé et coll., 2022). Entre le 14 et le 23 mars 2020, les gouvernements de l'Ontario, du Québec, du Manitoba, de la Saskatchewan et de la Colombie-Britannique ont pris la décision de fermer les écoles (Canadian Institute for Health Information, 2020). Des restrictions sur les voyages non essentiels ont été émises le 18 mars 2020 par le gouvernement canadien, excluant initialement les États-Unis. Puis les voyages à des fins récréatives et touristiques entre le Canada et les États-Unis ont été interdits le 21 mars 2020 (Canadian Institute for Health Information, 2020). Au même moment, au Québec et en Ontario, les gouvernements ont prescrit la limitation des rassemblements entre individus. Ensuite, entre le 20 et le 27 mars 2020, les gouvernements ontariens, québécois,

britannico-colombiens et saskatchewanais fermaient les commerces non essentiels (Canadian Institute for Health Information, 2020).

**Tableau 5. Quelques événements importants de la pandémie de la COVID-19 selon les temps de mesure**

Temps de mesure	Dates de l'enquête	Événements importants
1	6 avril – 6 mai 2020	Près d'un million de travailleurs ont perdu leur emploi en mars 2020 <sup>9</sup> .
3	4 mai – 25 mai 2020	Les commerces non essentiels rouvrent dans la plupart des provinces canadiennes <sup>10</sup>
5	1 <sup>er</sup> juin – 23 juin 2020	Le 18 juin 2020, le Canada a enregistré 100 000 cas de coronavirus <sup>11</sup> .
8	17 août – 13 septembre 2020	Le 24 août 2020, le port du masque devient obligatoire dans les lieux publics et les transports au Québec <sup>12</sup> .
10	26 novembre – 29 décembre 2020	Limitation des rassemblements et fermeture des commerces non essentiels durant les festivités de fin d'année dans la plupart des provinces canadiennes <sup>13</sup>
12	2 mars – 19 avril 2022	La campagne de vaccination contre le virus de la COVID-19 est bien amorcée au Canada <sup>14</sup> .

Entre le temps 10 et le temps 12, quelques événements importants se sont également produits. Par exemple, des couvre-feux ont été instaurés au Québec en janvier et février 2021 et janvier 2022 (Canadian Institute for Health Information, 2020). Le passeport vaccinal a été mis en place dans plusieurs provinces canadiennes entre les mois de septembre 2021 jusqu'au début de l'année 2022

<sup>9</sup> Statistique Canada, 2020

<sup>10</sup> Canadian Institute for Health Information, 2020

<sup>11</sup> CBC News, 2020.

<sup>12</sup> Canadian Institute for Health Information, 2020

<sup>13</sup> Ibid.

<sup>14</sup> Ibid.

(voir la section « Discussion » pour plus d'informations sur le passeport vaccinal) (Canadian Institute for Health Information, 2020).

Ces événements ont pu contribuer à façonner l'attribution du blâme et la désignation de boucs émissaires durant la crise de la COVID-19. Avant mars 2020, la plupart des événements majeurs liés à la pandémie se déroulaient essentiellement en Chine. Plusieurs décisions majeures des autorités, telles que la fermeture des écoles, des commerces non essentiels et des frontières, ainsi que la limitation des rassemblements, sont survenues tout juste avant le temps 1. Le port du masque obligatoire ainsi que l'instauration du passeport vaccinal sont des mesures sanitaires qui se sont également immiscées au fur et à mesure de la propagation du virus à partir du temps 8. Puis le début de la campagne de vaccination de masse a débuté aux alentours de mai 2021, avant le temps 12 (Canadian Institute for Health Information, 2020).

### 4.3 Statistiques descriptives

Le **tableau 6** présente les statistiques descriptives concernant le nombre de mots (moyenne, médiane, minimum et maximum) pour chaque temps de mesure. La moyenne varie entre 3,5 au temps 5 et 4,6 au temps 1, et la médiane fluctue entre 2 et 3 mots. La valeur médiane est plus basse que la moyenne, ce qui signifie que le nombre de mots élevé dans certaines réponses tire la moyenne vers le haut. Le maximum varie entre 57 au temps 5 et 196 au temps 3. Enfin, le nombre de mots minimum est de 1 à chaque temps de mesure.

**Tableau 6. Statistiques descriptives du nombre de mots**

<b>Statistiques descriptives</b>	<b>Temps 1</b>	<b>Temps 3</b>	<b>Temps 5</b>	<b>Temps 8</b>	<b>Temps 10</b>	<b>Temps 12</b>
Moyenne	4,6	4,1	3,5	4,2	4,2	4,0
Médiane	3,0	2,0	2,0	2,0	2,0	2,0
Minimum	1	1	1	1	1	1
Maximum	75	196	57	78	92	61

Le **tableau 7** présente la fréquence et la proportion de réponses qui ne comportent qu'un seul mot à chaque temps de mesure. Ce pourcentage fluctue entre un mot sur cinq (22,0 %, au temps 1) et un mot sur quatre (25,8 %, au temps 5). Le **tableau 8** présente les mots seuls les plus récurrents, par temps de mesure. Le mot le plus fréquent est *China*, avec une proportion allant de 9,1 % au temps 1 à 4,8 % au temps 12. Les autres mots (*government, nature, WHO, nobody, chinese, everyone*) comportent des pourcentages beaucoup plus faibles, variant entre 0,1 % et 2,4 %.

**Tableau 7. Pourcentage des réponses ne comportant qu'un seul mot**

<b>Temps de mesure</b>	<b>Fréquence</b>	<b>Fréquence totale</b>	<b>Pourcentage</b>
1	632	2 869	22,0 %
3	540	2 230	24,2 %
5	516	1 999	25,8 %
8	422	1 778	23,7 %
10	419	1 759	23,8 %
12	365	1 620	22,5 %

**Tableau 8. Mots seuls les plus fréquents et leurs proportions sur le nombre total de réponses**

<b>Mots seuls</b>	<b>Temps 1</b>	<b>Temps 3</b>	<b>Temps 5</b>	<b>Temps 8</b>	<b>Temps 10</b>	<b>Temps 12</b>
<i>China</i>	9,1 %	6,1 %	7,4 %	6,2 %	5,7 %	4,8 %
<i>government</i>	1,3 %	0,5 %	1,5 %	1,6 %	1,6 %	1,5 %
<i>Nature</i>	1,2 %	1,7 %	1,8 %	1,3 %	1,8 %	1,2 %
<i>WHO</i>	0,9 %	0,6 %	1,0 %	0,6 %	0,2 %	0,1 %
<i>Nobody</i>	0,8 %	1,3 %	1,5 %	2,0 %	1,3 %	2,4 %
<i>Chinese</i>	0,7 %	1,3 %	1,0 %	1,3 %	1,2 %	0,7 %
<i>Everyone</i>	0,3 %	0,5 %	0,7 %	1,1 %	1,3 %	0,6 %

### 4.3 Principaux mots mentionnés

Les **graphiques 2 à 7** (en annexes) présentent les 10 principaux mots mentionnés par les participants pour chaque temps de mesure. Dans ces graphiques, la fréquence des termes diminue progressivement, puisqu'il y a de moins en moins de répondants à la question ouverte (voir le **tableau 1**). Les mots *government*, *China* et *Chinese* sont récurrents, ce qui indique qu'ils sont fréquemment cités comme responsables de la crise de la COVID-19. Les données de ces graphiques permettent également d'observer une augmentation de la mention du terme *people* à chaque temps de mesure. Au temps 10, d'autres expressions font leur apparition, telles que *follow*, *health* et *guideline*. Ceci suggère que les participants blâment davantage les individus qui ne suivent pas les protocoles sanitaires à des phases ultérieures de la pandémie. Au temps 12, les termes *follow* et *people* sont également cités par les participants. De plus, le mot *unvaccinated* fait son apparition durant cette période. Le temps 12 de

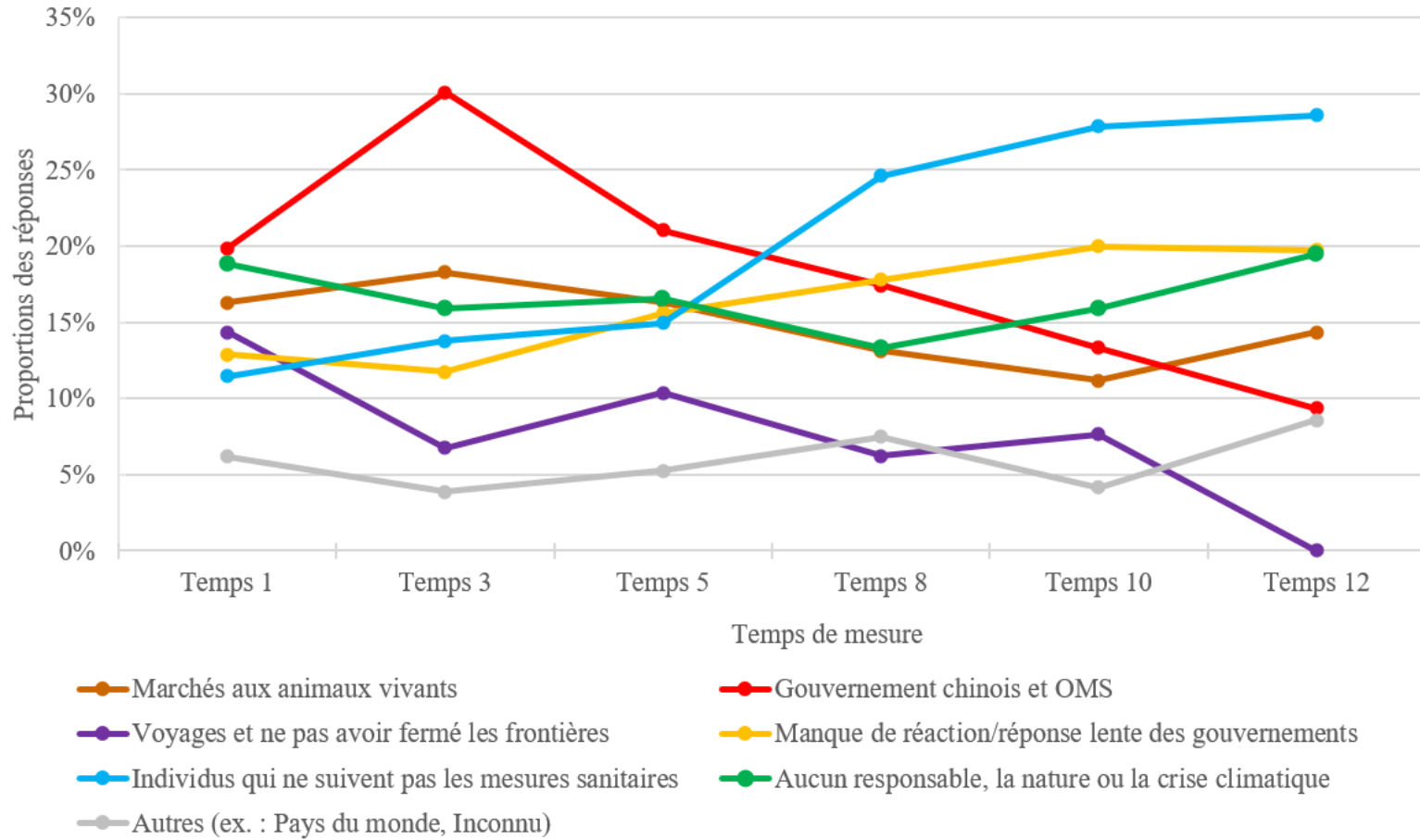
l'enquête a eu lieu entre mars et avril 2022, durant une période où la vaccination de la population canadienne avait débuté depuis un peu plus d'un an. L'Organisation mondiale de la Santé (OMS) se trouve parfois dans les 10 mots les plus mentionnés par les répondants, sous l'expression *WHO*.

#### **4.4 Proportion des thématiques de l'ADL et sommaire des thématiques**

Le **graphique 1** présente la prévalence des thématiques générées par l'ADL pour les différents temps de mesure. Six thématiques sont récurrentes : 1) Gouvernement chinois et OMS ; 2) Marchés aux animaux vivants ; 3) Voyages et ne pas avoir fermé les frontières ; 4) Aucun responsable, responsabilité attribuée à la nature ou à la crise climatique ; 5) Manque de réaction et réponse lente des gouvernements ; 6) Individus qui ne respectent pas les mesures sanitaires. Au temps 12, une sous-catégorie du thème « Individus qui ne respectent pas les mesures sanitaires », portant sur les non-vaccinés, a émergé. Les **tableaux 7 à 12** (en annexes) décrivent le sommaire des thématiques par temps de mesure. Comme mentionné dans la section « Méthodologie », ils ont été nommés en fonction de leur liste de mots correspondants (étiquette). Les quinze termes les plus fréquents pour chaque thématique sont inclus dans ces tableaux. Ces sommaires présentent les mots avec les probabilités les plus élevées d'appartenir à une thématique.

Les prochains paragraphes abordent une analyse détaillée des six thématiques récurrentes de chaque temps de mesure et de la sous-catégorie du temps 12, à partir des informations provenant du **graphique 1** et des **tableaux 12 à 17**. Les propos de différents participants sont rapportés entre guillemets et en italique, et les temps de mesure concernés se trouvent entre parenthèses. Un numéro aléatoire est attribué à chaque participant dont la réponse est citée.

**Graphique 1. Proportion des thématiques de l'ADL**



## 1. Gouvernement chinois et Organisation mondiale de la Santé (OMS)

Ces deux entités sont les plus fréquemment tenues comme responsables pour la crise de la COVID-19 jusqu'au temps 8. La proportion de réponses varie entre 19,9 % et 30,1 % aux temps 1 à 5, puis diminue pour se situer à 9,3 % au temps 12. Le gouvernement chinois et l'OMS ont été regroupés en une seule thématique, ce qui suggère que les participants les citent souvent ensemble dans leur réponse. Autrement dit, lorsque les répondants mentionnent le gouvernement chinois, ils sont plus susceptibles d'évoquer l'OMS, et ainsi de suite : *“Government of China for not containing the virus sooner and advising WHO earlier.”* (participant 1201 ; temps 1), *“WHO for not responding quickly enough China for their hiding the origin and seriousness of the outbreak as I presently understand it”* (participant 250 ; temps 3), *“the chinese government for not informing the world about the virus, and the world health organization for not labelling it as a pandemic sooner”* (participant 1336 ; temps 5) et *“China not advising the WHO earlier of the spread and only advising when it has already spread badly in the country”* (participant 2455 ; temps 8).

Le gouvernement chinois est également pointé du doigt pour sa dissimulation des informations concernant la COVID-19 : *“The People's Republic of China (government) for withholding information to the public and being deceptive when they finally did admit there was a problem and the ongoing deception.”* (participant 910 ; temps 5), *“The Government of China for not alerting the world to the severity and spreading of COVID-19. It was know in 2019 and it seems the rest of the world wasn't made aware until 2020.”* (participant 1574 ; temps 8) et *“I do hold the Chinese government responsible. Safeguards weren't met and they knew about this outbreak long before they let the world know about it”* (participant 799 ; temps 12).



L’OMS est, quant à elle, blâmée pour le manque de rapidité à reconnaître la sévérité du virus et à agir : “*WHO, for downplaying the disease and its ease of spread, and for not declaring this a global pandemic weeks earlier than they had.*” (participant 36 ; temps 1), “*The WHO not declaring this a pandemic until it spread to the world. This could have been stopped or limited in February.*” (participant 103 ; temps 3), “*Global governments and institutions such as WHO that, although being aware of the likelihood of a pandemic, failed to plan for the eventuality and chose to be reactive once the acknowledgment was made of the virus’s existence.*” (participant 3474 ; temps 5), “*Lack of containment of the virus, due to the reluctance of the WHO to react to the worsening situation in China last winter*” (participant 1042 ; temps 8).

## 2. Marchés aux animaux vivants

Si elle est parmi les thématiques prédominantes aux temps 1 à 5, sa prévalence diminue à travers le temps jusqu’au temps 10. Elle augmente ensuite légèrement au temps 12. La proportion de réponses se situe entre 16,3 % et 18,3 % aux temps 1 à 5, puis passe sous le seuil des 15 % aux temps 8 et 10. Cette proportion augmente ensuite pour se situer à 14,4 % au temps 12. Les participants font surtout référence aux animaux qui y sont vendus et décrivent ces lieux comme insalubres ou malsains. Par exemple : “*China’s lack of hygiene. China accepting of wet markets and eating wild animals that may or may not have been infected with a virus.*” (participant 509 ; temps 1), “*wet markets in China that have no sanitary measures and on which the chinese government has no control.*” (participant 337 ; temps 5), et “*The wet market in Wuhan China, greater health measures are needed*” (participant 1601 ; temps 12).

### 3. Voyages et ne pas avoir fermé les frontières

La prévalence de cette thématique diminue progressivement, passant de 14,3 % au temps 1 à 7,6 % au temps 10. Cette thématique disparaît ensuite au temps 12. Celui-ci fait référence au maintien des voyages et à l'ouverture des frontières comme responsables de la crise de la COVID-19. Les participants identifient notamment les voyages en lien avec le manque de précautions sanitaires : *“As the virus is only transmitted from person to person, the magnitude of the epidemic is mainly due to all those people who travel without any restriction from country to country for their own selfish pleasure.”* (participant 1060 ; temps 3), *“International travel without adequate health and safety precautions”* (participant 1227 ; temps 5) et *“No one really knows where it came from. But I do believe the travellers are the one spreading it from country to country. We all need to stay in our own countries now. Even province to province can be a mistake”* (participant 1701 ; temps 10).

De plus, les participants reprochent au gouvernement de ne pas avoir agi assez rapidement, par exemple en fermant les frontières : *“Not shutting down the country fast enough to international travel.”* (participant 2756 ; temps 1), *“Once an outbreak happened travel should have been stopped to minimize the impact. However we were warned a pandemic would happen, and governments did not listen”* (participant 2220 ; temps 3), *“The federal government for not closing the borders to all air traffic from foreign countries when covid-19 was first discovered or at the very least as soon as the WHO declared a pandemic. They also needed to institute mandatory 2 weeks quarantine for all Canadians returning to Canada.”* (participant 579 ; temps 8) et *“The Federal government did not close the border or the airport to foreigners soon enough to limit the entrance of viruses from outside of the country. Same for the provincial government that did not enforce restrictions sooner.”* (participant 1500 ; temps 10).

#### 4. Aucun responsable, responsabilité attribuée à la nature ou à la crise climatique

La prédominance de cette thématique fluctue à travers le temps, passant de 18,8 % au temps 1 à 15,9 % au temps 10 et à 19,4 % au temps 12. Il regroupe les sous-thèmes suivants : 1) aucun responsable précis, 2) la nature et l'environnement, et 3) la crise climatique. Certains participants mentionnent que personne n'est à blâmer ou qu'il s'agit d'une coïncidence : “*There is no one responsible ..we can't dwell on this question ..it's not the time to look for the who and why ..anyway we will probably never know.*” (participant 10 ; temps 1), “*Fate - A new virus can strike at any time and we cannot always be perfectly prepared.*” (participant 854 ; temps 3), et “*Nobody is responsible... It is a pandemic difficult to control.*” (participant 1382 ; temps 12).

D'autres pointent du doigt la crise climatique : “*The destruction and devastation of the environment and the intrusion on the natural habitat of many animals because of this destruction.*” (participant 703 ; temps 1), “*Climate change and rapid urbanization leading to overcrowded urban centres.*” (participant 2168 ; temps 3) et “*The human devastation of the environment forcing animals into unfamiliar terrain and increasing contact with humans due to our pollution and unending consumerism.*” (participant 1024 ; temps 8). Certains participants combinent plusieurs thèmes dans leurs réponses, tels que : “*no one, it is just what it is. Mother nature trying to correct a balance.*” (participant 281 ; temps 8) et “*Not sure any one thing or person is responsible. A perfect storm of a global society, probably fuelled by one step too far of misuse of the environment and wildlife interaction. Pushing the lines for decades - something was going to give somewhere along the way.*” (participant 128 ; temps 12).

## 5. Manque de réaction et réponse lente des gouvernements

L'importance de cette thématique augmente progressivement à travers le temps, passant de 12,9 % au temps 1 à 19,7 % au temps 12. Celui-ci correspond au manque de réaction et à la réponse lente des gouvernements comme responsables de la crise de la COVID-19. Les participants soulignent le manque de rapidité, de préparation ou la lenteur à réagir face à la pandémie : *“Both the federal and provincial governments’ slow response to activities that were going on abroad. At the very least, back in early January, PPE<sup>15</sup> stocks and hospital/LTCH<sup>16</sup> procedures should have been reviewed.”* (participant 1753 ; temps 1), *“Governments who failed to prepare ahead of time and therefore did not have needed supplies”* (participant 549 ; temps 8) et *“Government flip flop decision in policy. Starting from Federal to Provincial, no one ready like to take full responsibility on dealing with the crisis.”* (participant 309 ; temps 12).

## 6. Individus qui ne suivent pas les mesures sanitaires

La prévalence de cette thématique augmente progressivement entre les temps 1 et 5. Par la suite, il devient dominant entre les temps 8 et 12. La proportion de réponses passe de 11,4 % au temps 1 à 28,6 % au temps 12. Celui-ci fait référence aux individus qui ne suivent pas les règles sanitaires, telles que la distanciation sociale, la limite des rassemblements ou le port du masque : *“The virus itself, people who won’t wear a mask, self-isolate or follow rules during this pandemic. Basically anyone who is selfish that puts their WANTS in front of everyone’s safety and health.”* (participant 902 ; temps 3), *“People that are not social distancing, especially people in the park and young people that I saw on the internet in Toronto gathering together and not wearing masks.”*

---

<sup>15</sup> ‘Personal protective equipment’ ou équipement de protection individuelle.

<sup>16</sup> ‘Long-term care home’ ou établissement de soins de longue durée.

(participant 1643 ; temps 5) et *“Anyone not following the rules and regulations and the government for not getting tougher because it has been proven that some people don’t listen. Much higher consequences for anyone and not just a slap on the hand. Even the fines are not nearly high enough and everyone involved should be fined. Should be mandatory for everyone to wear masks in public and shopping.”* (participant 1952 ; temps 8).

### 6.1 Non-vaccinés

Cette thématique, qui a émergé au temps 12, est une sous-catégorie du thème « Individus qui ne suivent pas les mesures sanitaires ». Celui-ci fait référence aux personnes qui sont non-vaccinées en tant que responsables de la crise de la COVID-19. La proportion de réponses faisant référence à cette thématique au temps 12 est de 22,0 %. Certains répondants assimilent les non-vaccinés aux individus qui ne suivent pas les mesures sanitaires : *“Unvaccinated people who also don’t wear masks and go out in huge groups and don’t take precautions to limit the spread”* (participant 636) et *“Those who are unvaccinated and do not practise recommendations, particularly masking”* (participant 13). D’autres les décrivent comme ne croyant pas en la science de la vaccination : *“The unvaccinated ignorant vocal minority who refuse to believe the science and dragged the crisis out longer than it needed to be”* (participant 990) et *“Unvaccinated, especially those who think Covid is a hoax or vaccines are unscientific”* (participant 1210). Enfin, certains répondants les décrivent comme étant égoïstes : *“Selfishness of those who do not vaccinate”* (participant 100) et *“Unvaccinated misinformed people who are selfish and do not care about protecting their fellow friends family or neighbours”* (participant 342).

## 5. Discussion

La crise de la COVID-19 peut être considérée comme un changement social dramatique (de la Sablonnière, 2017 ; de la Sablonnière et coll., 2020). Elle a entraîné des transformations sociales soudaines et profondes parmi les populations du monde entier. Elle a également provoqué des sentiments de confusion et d'ambiguïté chez les individus qui, de surcroît, n'exercent aucun contrôle sur la situation. Le manque de clarté et de contrôle sur la situation les incite ainsi à « chercher des responsables ». Les perceptions du public quant aux responsables de l'origine et de la propagation de la COVID-19 sont influencées par certains facteurs, tels que la proximité géographique de l'épidémie et les actions des individus (Mayor et coll., 2012 ; Roy et coll., 2020). Plus précisément, le blâme et la désignation de boucs émissaires ont tendance à fluctuer selon l'état de l'épidémie ; lorsque la maladie est considérée comme étant éloignée, il se concentre sur les collectivités distantes, et au moment où elle est considérée comme étant proche, il est transféré aux collectivités locales (Mayor et coll., 2012 ; Roy et coll., 2020). Le premier objectif de ce mémoire est d'analyser l'évolution de la dynamique du blâme et de la désignation de boucs émissaires durant la pandémie de la COVID-19 auprès de la population canadienne à travers les termes utilisés par le public. En effet, l'étude du langage permet d'obtenir une compréhension plus approfondie des perceptions et des préoccupations des individus. Le deuxième objectif est d'évaluer l'intérêt d'utiliser le modèle de mélange génératif bayésien de l'ADL (Allocation de Dirichlet latente) dans l'analyse de questions ouvertes. Les données proviennent d'une enquête longitudinale sur les conséquences sociales de la COVID-19 administrée auprès d'un échantillon représentatif de la population canadienne.

## **5.1 Considérations théoriques**

### **5.1.1 Attribution du blâme durant la COVID-19**

Les résultats de recherche démontrent que le blâme se concentre majoritairement sur les communautés distantes et se déplace ensuite vers les communautés locales. Six thématiques récurrentes ont été identifiées à chaque temps de mesure : 1) Gouvernement chinois et OMS (Organisation mondiale de la Santé) ; 2) Marchés aux animaux vivants ; 3) Voyages et ne pas avoir fermé les frontières ; 4) Aucun responsable, responsabilité attribuée à la nature ou à la crise climatique ; 5) Manque de réaction et réponse lente des gouvernements ; et 6) Individus qui ne respectent pas les mesures sanitaires. Une sous-catégorie de la thématique « Individus qui ne respectent pas les mesures sanitaires », qui porte sur les non-vaccinés, a émergé au temps 12. Ces thématiques seront abordées de manière plus détaillée dans les paragraphes suivants.

#### 1. Gouvernement chinois et OMS

Le gouvernement chinois et l'OMS sont parmi les premiers acteurs concernés par la crise de la COVID-19 ; les premiers cas du virus ont été détectés dans la ville de Wuhan, en Chine, et l'OMS est l'organisation qui a déclaré la COVID-19 comme étant une pandémie (Barreneche, 2020 ; de Rosa & Mannarini, 2020 ; Labbé et coll., 2022). Des recherches précédentes ont démontré que l'OMS a été blâmée pour la censure d'informations concernant le virus (Choli & Kuss, 2021), conformément aux résultats de la présente étude. L'OMS a également été blâmée pour des crises de santé publique antérieures, telles que l'épidémie de grippe A (H1N1) en 2009 (Wagner-Egger et coll., 2011 ; Mayor et coll., 2012). En effet, l'OMS était perçue comme étant l'un des principaux acteurs responsables de la lutte contre la maladie (Wagner-Egger et coll., 2011 ; Mayor et coll., 2012). Selon Wagner-Egger et coll. (2011), les

opinions du public à l'égard des organisations internationales, telles que l'OMS, sont partagées. Certains individus considèrent que leurs actions sont utiles, alors que d'autres expriment leur méfiance (Wagner-Egger et coll., 2011). Durant l'épidémie d'Ebola, l'OMS a notamment été accusée d'avoir menti à propos des modes de transmission du virus et d'avoir agi trop lentement (Roy et coll., 2020).

Conformément aux résultats de la présente recherche, des études précédentes mentionnent que le gouvernement chinois a été accusé d'avoir dissimulé ou censuré des renseignements concernant le virus de la COVID-19 (Hardy et coll., 2021 ; Nguyen et coll., 2021). En effet, d'après Shangguan et coll. (2020), la restriction de l'information par le gouvernement chinois a conduit à l'ignorance et au manque de préparation de la population face à la crise de la COVID-19. L'auteur explique qu'au début de la pandémie, le contrôle prolongé de l'information s'est produit en raison du climat politique et de la concomitance entre la première éclosion du virus et l'arrivée du Nouvel An chinois (Shangguan et coll., 2020). Cette gestion préalable de la pandémie a provoqué de la méfiance au sein du public (Choli & Kuss, 2021).

Les participants de la présente recherche ont indirectement désigné les personnes d'ascendance asiatique comme responsables de la crise de la COVID-19 en pointant du doigt certains lieux associés à des pratiques culturelles, tels que les marchés aux animaux vivants (voir le prochain paragraphe). Cette perception peut s'inscrire dans une vision essentialiste de la culture asiatique, où celle-ci est réduite à quelques aspects stéréotypés (Hardy et coll., 2021). Cela peut contribuer à la stigmatisation des communautés asiatiques et renforcer des préjugés. Ces résultats diffèrent quelque peu des conclusions de Ferrante et coll. (2022) selon lesquelles la majorité des Canadiens ont de faibles niveaux de préjugés envers les Asiatiques tout au long de la pandémie. Bien qu'ils aient tendance à résister au changement, l'intensité des préjugés peut augmenter durant les situations de crise (Allport, 1979). D'autre part, les recherches antérieures ont plutôt conclu



qu'il y a eu une augmentation des préjugés envers les Asiatiques, plus particulièrement de la part des politiciens et des médias de droite (de Rosa & Mannarini, 2020 ; Dionne & Turkmen, 2020 ; Choli & Kuss, 2021 ; Ittefaq et coll., 2022). La population se déclarant politiquement de droite a également une propension à attribuer la responsabilité de la pandémie aux Asiatiques (Ferrante et coll., 2022, Hardy et coll., 2021). Par ailleurs, l'étude de Nguyen et coll. (2021), démontre que les Asiatiques sont surtout blâmées pour son apparition. Cela suggère qu'ils sont pointés du doigt pour son origine plutôt que pour sa propagation.

## 2. Marchés aux animaux vivants

Les marchés aux animaux vivants (*wet markets*) sont des lieux où des produits alimentaires sont vendus. Historiquement, la consommation d'animaux sauvages vendus dans ces marchés est une pratique jugée « malsaine » et « primitive » par de nombreux Occidentaux (Labbé et coll., 2022). L'identification de ces lieux en tant que responsables de la crise remonte aux débuts de la pandémie, lorsque l'origine du virus a été attribuée à un marché de la ville de Wuhan en Chine. Les traditions culinaires asiatiques ont également été pointées du doigt comme étant la source de la maladie (Barreneche, 2020), ce qui a mené une partie de la population à éviter les restaurants chinois et asiatiques (Labbé et coll., 2022). Selon Hardy et coll. (2021), certains individus qui blâment les marchés aux animaux vivants font référence à la consommation d'animaux sauvages et en parlent comme une habitude qu'ils trouvent répugnante.

Si le public évoque les animaux qui y sont vendus et ses aspects insalubres ou malsains, il existe différentes variétés de marchés aux animaux vivants qui n'impliquent pas nécessairement leur commercialisation (Lin et coll., 2021). En effet, les marchés d'animaux vivants et les marchés d'animaux exotiques sont souvent confondus (Lin et coll., 2021), un phénomène qui a été observé chez les

participants de la présente recherche et de l'étude de Hardy et coll. (2021). Cette confusion peut amplifier les attitudes xénophobes à l'égard des cultures asiatiques (Lin et coll., 2021).

### 3. Voyages et ne pas avoir fermé les frontières

Ces deux thèmes (voyages et ne pas avoir fermé les frontières) ont été générés sous une seule thématique par l'ADL. Ceci peut s'expliquer par le fait que les participants les mentionnent souvent ensemble dans leur réponse. Lorsqu'ils abordent les voyages, ils font notamment référence au comportement individuel des voyageurs. Puis quand ils mentionnent la non-fermeture des frontières, ils font principalement référence à des décisions du gouvernement. Cette thématique comprend donc à la fois le blâme adressé aux individus et aux institutions.

Les voyages sont notamment blâmés en raison de leur possibilité de faire entrer le virus dans le pays (Choli & Kuss, 2021 ; Labbé et coll., 2022). Les participants de l'étude de Nguyen et coll. (2021) ont identifié le tourisme et les voyages internationaux en tant que principales raisons de la propagation de la COVID-19. Les voyageurs ont également été représentés comme responsables d'avoir propagé le virus de la COVID-19 (Labbé et coll., 2022). Le blâme était à la fois attribué aux individus qui revenaient au Canada et aux personnes qui ont enfreint les restrictions sur les voyages (Labbé et coll., 2022). Certains ont reproché aux voyageurs de ne pas avoir fait preuve de solidarité avec le reste de la population, qui devait faire face aux restrictions sanitaires (Labbé et coll., 2022).

Au cours de la pandémie de la COVID-19, de nombreuses personnes ont demandé à leurs gouvernements respectifs de fermer les frontières, et plus particulièrement les frontières chinoises (Choli & Kuss, 2021). La gestion des

points d'entrée internationaux était également un enjeu crucial lors de l'épidémie d'Ebola. Plusieurs individus ont critiqué leurs gouvernements pour le manque de contrôle ; ils ont condamné la décision de les maintenir partiellement ouvertes (Roy et coll., 2020). Abeysinghe (2016) affirme que les frontières tiennent un rôle historique important dans la gestion d'une épidémie, puisqu'elles représentent une « barrière géopolitique et symbolique entre les individus malades et les individus en bonne santé » (p.464). Leur fermeture créerait une sorte de « barrière de protection » auprès d'individus en proie à des peurs irrationnelles de l'étranger et du malade. Dans cette perspective, la nation incarnerait un espace de protection idéalisé et rassurant face à des épidémies menaçantes (Abeysinghe, 2016 ; Roy et coll., 2020). Les gouvernements qui ne préconisent pas un contrôle aux frontières ou un confinement de la nation seraient perçus comme incompetents et fautifs (Roy et coll., 2020).

#### 4. Aucun responsable, responsabilité attribuée à la nature ou à la crise climatique

Certains participants de l'étude n'ont pas attribué la responsabilité pour la crise de la COVID-19 à un quelconque facteur humain ou l'ont attribué à la crise climatique. Ils ont identifié soit 1) aucun responsable précis, soit 2) la nature et l'environnement, et 3) la crise climatique. Ces thèmes ont également émergé dans d'autres enquêtes auprès de populations, telles que Nguyen et coll. (2021), Hardy et coll. (2021) et Rateau et coll. (2021). Ils n'ont toutefois pas émergé dans d'autres types d'études, telles que les analyses des réseaux sociaux (Choli & Kuss ; 2021) et les analyses de discours médiatiques et institutionnels (de Rosa & Mannarini, 2020 ; Ittefaq et coll., 2022 ; Labbé et coll., 2022). Ce sont pourtant des thèmes fréquents en ce qui concerne l'attribution de la responsabilité durant la COVID-19. Quand il s'agit de ce thème, selon les études de Nguyen et coll. (2021) et Rateau et coll. (2021), soit le public n'impute pas le blâme à un quelconque facteur humain, soit le public croit que l'origine du virus est non humaine et non intentionnelle. D'après l'enquête de Hardy et coll. (2021), ces

tendances varient également selon l'orientation politique : les individus s'identifiant comme étant de gauche ou apolitiques sont plus enclins à affirmer qu'il n'y a aucun responsable pour la crise de la COVID-19 que les autres.

### 5. Manque de réaction et réponse lente des gouvernements

Dans la présente recherche, l'attribution du blâme a évolué au fur et à mesure de la progression de la pandémie et de sa propagation localement, pour se concentrer entre autres sur le manque de réaction et la réponse lente des gouvernements. De précédentes études soulignent également que le blâme durant une crise sanitaire évolue, passant de communautés distantes aux autorités et élites politiques locales (Mayor et coll., 2012 ; Roy et coll., 2020 ; Labbé et coll., 2022). Celles-ci sont notamment pointées du doigt à cause de leur inefficacité à combattre le virus ou d'une détérioration de la situation (Joffe, 2011 ; de Rosa & Mannarini, 2020). Dans l'étude de Choli & Kuss (2021), les gouvernements ont été blâmés en raison de leur manque d'action et de leur approche désordonnée face à la pandémie. Ils ont également été désignés comme responsables à cause de leur gestion imparfaite de crises de santé publique précédentes. Au cours de l'épidémie de grippe A (H1N1), Mayor et coll. (2012) ont constaté que le public exprimait des points de vue divergents concernant les autorités. Près de la moitié des participants exprimaient de la méfiance envers les gouvernements à des stades avancés de la pandémie (Mayor et coll., 2012). Le même phénomène peut être observé à des phases ultérieures de l'épidémie d'Ebola. Les autorités étaient fréquemment blâmées pour l'origine et la mauvaise gestion de la maladie, ainsi que pour la corruption et la dissimulation des problèmes qui ont contribué à sa propagation (Mondragon et coll., 2017 ; Roy et coll., 2020).

Conformément à ce qui a été mentionné auparavant, l'identification des autorités publiques en tant que responsables d'une crise peut suggérer la présence d'une dynamique de méfiance envers les gouvernements, dont les actions

seraient perçues comme étant dommageables à la population (Mayor et coll., 2012 ; Choli & Kuss, 2021 ; Hardy et coll. 2021). Choli et Kuss (2021) expliquent que la méfiance croissante à l'égard des autorités, et la politisation des épidémies précédentes pourraient expliquer leur prédominance dans l'attribution du blâme. Ils ajoutent que les individus peuvent ressentir un sentiment de contrôle accru lorsqu'ils désignent les gouvernements comme responsables de la crise, puisqu'ils ont la possibilité de leur demander des comptes (Choli & Kuss, 2021). S'il est possible qu'une préparation adéquate et une réponse efficace de la part des autorités diminuent la prévalence de ce phénomène, une nouvelle rhétorique de blâme est toujours susceptible d'émerger en raison de l'incertitude inhérente à des événements soudains et inattendus (Longstaff & Yang, 2008 ; Nguyen et coll., 2021). L'attribution du blâme peut détourner l'attention des gens par rapport aux efforts gouvernementaux et collectifs pour gérer la crise, ce qui fait que ces efforts semblent moins efficaces (Longstaff & Yang, 2008).

## 6, Individus qui ne suivent pas les mesures sanitaires

Les résultats de la présente étude sont cohérents avec des recherches antérieures qui affirment que les individus ne respectant pas les mesures sanitaires sont majoritairement identifiés comme responsables de la crise de la COVID-19 au cours des phases ultérieures de la pandémie (Nguyen et coll., 2021 ; Labbé et coll., 2022). Ces individus ont été présentés comme étant « immoraux », « stupides » et « égocentriques » dans des caricatures éditoriales canadiennes (Labbé et coll., 2022). Le néologisme « covidiot », résultat d'une combinaison entre « covid » et « idiot », est apparu durant la pandémie pour qualifier les individus qui ne suivent pas les mesures sanitaires (Labbé et coll., 2022). Selon Martikainen et Sakki (2021), les « fêtards imprudents » sont l'une des représentations visuelles construites par les médias de masse lors de la crise de la COVID-19. Ceux-ci sont souvent incarnés par des foules de jeunes qui ne

pratiquent pas la distanciation sociale et qui ne portent pas de masque de protection (Martikainen & Sakki, 2021).

Durant des périodes de crise, un langage militaire impliquant une construction narrative de « héros », de « vilains » et de « victimes » a souvent été utilisé sur la place publique. La pandémie de la COVID-19 ne fait pas exception. Des métaphores de guerre ont été utilisées par les leaders politiques et les médias de masse ; des expressions telles que « nous avons besoin de combattre le virus, de gagner la bataille contre le virus » sont devenues courantes (Labbé et coll., 2022). Les individus ne respectant pas les mesures sanitaires ont été pointés du doigt comme des « vilains » en raison de leur refus de faire leur part pour l'intérêt collectif (Barreneche, 2020 ; Casciano, 2020 ; Labbé et coll., 2022). La condamnation des actions de ces personnes témoigne « d'un discours sur le sacrifice pour le bien commun, typique des situations de guerres » (Labbé et coll., 2022, p.9). Enfin, selon Douglas (1992), le blâme envers les individus qui ne suivent pas les protocoles en place correspond à une stratégie pour les encourager à se conformer aux règles.

Par ailleurs, dans la société contemporaine occidentale, le maintien de la santé représente une valeur fondamentale, voire une métaphore de la conduite vertueuse (Brandt & Rozin, 1997). Sa « maximisation » est considérée comme un élément essentiel du bien-être et doit constituer un objectif à atteindre (Labbé et coll., 2022). Dans ce contexte, le blâme envers les individus qui ne suivent pas les règles sanitaires peut révéler une éthique d'autodiscipline, d'autorégulation et de responsabilité envers la santé (Petersen & Lupton, 2000 ; Labbé et coll., 2022). Cette attitude peut conduire à l'intolérance, l'exclusion ou la persécution des individus qui ne veulent pas s'engager dans des activités comme étant bénéfiques à la santé, puisqu'ils sont perçus comme étant responsables de leur condition (Petersen & Lupton, 2000 ; Labbé et coll., 2022). Selon Petersen et Lupton (2000) : *“The ‘healthy’ body has become an increasingly important signifier of*

*moral worth, a mark of distinction that serves to delineate those who deserve to succeed from those who will fail” (p. 25).*

### 6.1 Non-vaccinés

La sous-catégorie « non-vacciné » est apparue au temps 12 de la présente enquête sur la COVID-19, entre le 10 mars et le 9 avril 2022. Durant cette période, au Canada, la vaccination contre le virus était déjà bien répandue ; au moins deux doses de rappel ont été administrées à une grande partie de la population (Agence de la santé publique du Canada, 2022). Leur mise en service n'est généralement approuvée que deux ans après des études poussées avec des données probantes (Bardosh et coll., 2022). Toutefois, compte tenu de la situation urgente engendrée par la pandémie, les vaccins ont été administrés dès la fin de l'année 2020 (Bardosh et coll., 2022). Ils ont représenté une « intervention critique » pour éviter la morbidité et la mortalité au cours de la pandémie (Bardosh et coll., 2022, p. 2).

Pendant plusieurs mois entre 2021 et 2022, plusieurs pays ont exigé une preuve de vaccination (ou passeport vaccinal) à la population pour accéder à certains lieux publics comme mesure pour lutter contre la COVID-19 (Bardosh et coll., 2022 ; Voigt, 2022). La majorité des provinces et territoires canadiens ont exigé une telle preuve, sauf les Territoires du Nord-Ouest et le Nunavut (Retail Council of Canada, 2022). Au Canada, les emplacements concernés par cette mesure sont les restaurants, les centres de divertissement, les installations destinées aux sports, les musées, les théâtres, les salles de spectacle, les cinémas et les salles de conférence (Retail Council of Canada, 2022). L'accès aux commerces essentiels (épiceries et pharmacies), aux édifices religieux, aux lieux extérieurs, aux banques et aux commerces de détail n'était toutefois pas affecté par ces mesures (Retail Council of Canada, 2022). Pour obtenir une telle preuve,

il fallait notamment avoir reçu deux doses de vaccin contre la COVID-19 (Gouvernement du Canada, 2023).

Les études de Voigt (2022) et Bardosh et coll. (2022) démontrent que la mise en place du passeport vaccinal a eu pour effet de percevoir les individus non-vaccinés en termes négatifs. En effet, le qualificatif « non-vacciné » a été associé à des préoccupations de contagion et de dangerosité (Voigt, 2022 ; Bardosh et coll., 2022). Ils ont notamment été blâmés pour la prolongation de la pandémie, la pression accrue subie par les hôpitaux, l'émergence de nouveaux variants du virus et le maintien des mesures sanitaires (Bardosh et coll., 2022). Un discours moralisateur s'est construit autour de la vaccination, considérée comme un « devoir social » (Voigt, 2022). Les non-vaccinés refuseraient d'accomplir cette responsabilité, ce qui justifie « le fait qu'ils doivent supporter une charge plus importante que les vaccinés » (Voigt, 2022, p.57). Ils sont donc devenus les boucs émissaires de la pandémie, stigmatisés et discriminés à travers une rhétorique politique qui visait à augmenter le taux de vaccination (Bardosh et coll., 2022). Lors d'une allocution, le premier ministre du Canada, Justin Trudeau, a déclaré ceci à leur égard : *“They are extremists who don't believe in science, they're often misogynists, also often racists... It's a small group that muscles in, and we have to make a choice, as a leader and as a country: Do we tolerate these people?”* (Bardosh et coll., 2022, p.5).

Au cours de la pandémie, la perception négative des non-vaccinés est devenue un phénomène socialement accepté dans la population, pour laquelle la vaccination est considérée comme une obligation morale (Bardosh et coll., 2022). Les individus qui refusent de se faire vacciner contre la COVID-19 ont été pointés du doigt en tant qu'obstacle au « retour à la normale » et comme la source des variants du virus (Labbé et coll., 2022). Ceci a renforcé la polarisation et la division sociale, orientant les attitudes vers une dynamique « nous vs eux ». De plus, les individus non-vaccinés sont généralement perçus de façon négative par



la population (Jedwab et coll., 2021). Selon Bor et coll. (2022), les personnes vaccinées ont également une forte antipathie envers les non-vaccinés. Les non-vaccinés sont considérés comme indignes de confiance en raison de leur refus de contribuer au bien collectif (en se faisant vacciner) et incompetents puisqu'ils croient à de fausses informations concernant la vaccination (Bor et coll., 2022).

### **5.1.2 Évolution du blâme durant la pandémie**

L'attribution du blâme est un phénomène dynamique qui fluctue selon les développements de la pandémie de la COVID-19 et de son emplacement géographique. Si le gouvernement chinois, l'OMS et les marchés aux animaux vivants sont d'abord désignés comme principaux responsables de la crise de la COVID-19 aux temps 1 à 5 (première vague du virus<sup>17</sup>), les individus qui ne respectent pas les mesures sanitaires sont ensuite pointés du doigt aux temps 8 et 10 (deuxième vague de la pandémie<sup>18</sup>). Cette tendance s'est également maintenue au temps 12 de l'étude, durant la sixième vague du virus<sup>19</sup>. Avant le temps 8 (août et septembre 2020), les mesures sanitaires mises en place pour faire face à la COVID-19 étaient principalement axées sur le confinement et la distanciation sociale. À partir du temps 8, d'autres mesures sanitaires se sont ajoutées, telles que le port du masque obligatoire, le passeport vaccinal et le couvre-feu. Au temps 10 (novembre et décembre 2020), les gouvernements provinciaux ont fortement limité les rassemblements sociaux des festivités du temps des fêtes, ce qui a provoqué déception et tristesse parmi la population (Aubin, 2020). Cette période de restrictions a probablement contribué à renforcer l'attribution du blâme envers les personnes qui ne respectent pas les mesures sanitaires, car elles étaient considérées comme potentiellement responsables de la persistance de la

---

<sup>17</sup> La 1<sup>e</sup> vague a débuté en février 2020 et s'est terminée en juillet 2020 (INSPQ, 2022).

<sup>18</sup> La 2<sup>e</sup> vague a débuté en août 2020 et s'est terminée en mars 2021 (INSPQ, 2022).

<sup>19</sup> La 6<sup>e</sup> vague a débuté en mars 2022 et s'est terminée en mai 2022 (INSPQ, 2022).

pandémie et du risque de nouvelles vagues d'infections. Enfin, au temps 12, à mesure que la couverture vaccinale augmentait, la stigmatisation envers les non-vaccinés s'intensifiait.

Ceci vient confirmer l'hypothèse selon laquelle le blâme se concentre d'abord majoritairement sur les communautés distantes (temps 1, 3 et 5), pour ensuite se déplacer vers les communautés locales (temps 8, 10 et 12) au fur et à mesure que les effets de la pandémie se font ressentir à proximité. En effet, la désignation du blâme change notamment en fonction de la proximité de la menace ou du risque de propagation international (Roy et coll., 2020 ; Choli & Kuss, 2021 ; Labbé et coll., 2022). La présente étude a également permis de relever que le blâme se déplace selon les périodes clés de la pandémie, et particulièrement l'implantation de nouvelles mesures sanitaires par le gouvernement. Lorsque de nouvelles mesures sont mises en place, certaines personnes peuvent être pointées du doigt si elles ne les respectent pas.

Mayor et coll. (2012) relèvent que, durant l'épidémie de grippe A (H1N1), le déplacement du blâme s'est également produit à travers la couverture médiatique. Au début de l'épidémie, les pays éloignés étaient plus représentés, puis lorsque la maladie s'est propagée dans le pays, les mentions des populations à risque et des médecins sont devenues plus fréquentes (Mayor et coll., 2012). Les auteurs suggèrent que le public peut être influencé par le traitement médiatique de la maladie sur le plan de l'attribution de la responsabilité (Mayor et coll., 2012).

Il n'est pas possible de déterminer quelles sont les tendances dans l'attribution du blâme avant la mise en place des mesures sanitaires par les autorités gouvernementales. En effet, l'enquête canadienne a débuté en avril 2020, alors que les premières mesures visant à lutter contre le virus ont été implantées à la mi-mars 2020 (INSPQ, 2022). Avant cette période, il est possible

que les tendances à identifier des communautés distantes comme responsables de la crise aient été plus élevées, comme l'ont démontré Hardy et coll. (2021).

## 5.2 Considérations méthodologiques

La présente recherche vise également à évaluer l'intérêt d'utiliser le modèle de mélange génératif ADL (Allocation de Dirichlet latente) dans l'étude de questions ouvertes. Les résultats obtenus démontrent que cette méthode détient un grand potentiel d'analyse du langage naturel et des mots. La majorité des thématiques extraites sont pertinentes à la question de recherche. Les groupes de mots de la majorité des thématiques ont été aisément interprétables. Grâce à sa capacité à détecter les associations sémantiques entre les mots, l'ADL s'est révélée être un outil puissant pour comprendre la complexité des discours entourant le blâme pendant la crise de la COVID-19. L'ADL a réduit un corpus de réponses « désordonné » et complexe en catégories tangibles et compréhensibles (Nelson, 2020). Cela permet aux chercheurs de consacrer davantage de temps à l'analyse et à l'interprétation des résultats plutôt qu'à la tâche fastidieuse du codage manuel (Nelson, 2020). En effet, l'étape de catégorisation liée au codage manuel, où les chercheurs associent les textes aux thématiques, requiert une analyse minutieuse et réfléchie pour construire des thèmes pertinents dans lesquels classer les réponses. Cette étape peut être complexe et exigeante, car elle implique de regrouper les réponses en ensembles cohérents et significatifs, ce qui nécessite une compréhension approfondie du contenu des réponses et de la question de recherche (Nelson, 2020). Par ailleurs, le codage manuel est sujet à des erreurs humaines potentielles, telles que des jugements biaisés ou des interprétations subjectives lors de la catégorisation des réponses (Pietsch & Lessman, 2018). Cette limite peut seulement être palliée par une mesure de la concordance qui vise à évaluer la fiabilité des décisions prises par les codeurs lors de l'attribution de catégories au texte (Pietsch & Lessman, 2018). En revanche, l'ADL offre une approche plus objective et reproductible

dans la catégorisation des réponses, réduisant ainsi le risque de biais ou d'incohérences pouvant survenir lors du codage manuel (Nelson, 2020).

L'ADL a permis d'économiser du temps d'analyse qui aurait été consacré au codage manuel (Pietsch & Lessman, 2018). En effet, cette dernière nécessite un investissement de temps important en raison de la lecture et du traitement manuel de chaque document, en particulier lorsque le corpus de texte est volumineux (Nelson, 2020). En comparaison, l'ADL offre une approche plus efficace et rapide, car le processus de catégorisation est automatisé (Pietsch & Lessman, 2018). Cela permet aux chercheurs de se concentrer davantage à l'analyse des résultats. En somme, l'ADL a rendu possible l'extraction de thématiques plus riches et ancrées dans les données (Nelson, 2020), puisqu'elles reflètent les tendances relevées auprès des participants.

Les études en sciences sociales bénéficieraient grandement des méthodes automatisées d'analyse de texte. Les faits sociaux constituent l'objet d'étude de la sociologie, et les méthodes d'analyses automatisées du texte donnent l'opportunité au chercheur de se pencher sur ces phénomènes à grande échelle. Elles offrent diverses perspectives de recherches. Par exemple, il est possible d'étudier la couverture médiatique sur un sujet, des discours et des contenus audio et vidéo, ainsi que des publications ou des commentaires sur les réseaux sociaux. Plus particulièrement, ces derniers représentent des sources d'informations importantes pour connaître les réactions et les préoccupations des individus par rapport à des enjeux collectifs (Choli & Kuss, 2021). En effet, les réseaux sociaux sont des plateformes où les internautes partagent librement leurs opinions et leurs émotions. De plus, dans un contexte où la sociologie vise à découvrir de nouvelles perspectives et à prendre de la distance par rapport à un phénomène, les méthodes automatisées permettent d'obtenir des résultats sous un angle nouveau ou sous des aspects qui peuvent ne pas être décelés par une

analyse manuelle. Toutefois, ces méthodes n'en comportent pas moins certaines limites, qui seront abordées à la prochaine section.

## 6. Limites

Cette recherche comporte plusieurs limites. Tout d'abord, l'identification de responsables de la crise de la COVID-19 a été mesurée par une question ouverte, laquelle comporte plusieurs avantages (évoqués à la section « Méthodologie »), mais également des limites. En effet, le format ouvert est souvent caractérisé par un taux de réponse plus bas que les questions fermées (Reja et coll., 2003). La non-réponse peut à la fois être expliquée par un désintérêt et un manque d'éloquence (Roberts et coll., 2014). En effet, certains individus ne répondent pas aux questions ouvertes « puisqu'ils ne disposent pas de l'outil rhétorique nécessaire » (Roberts et coll., 2014, p.1066). La qualité de l'information obtenue peut donc dépendre de la facilité du participant à s'exprimer par écrit (Fink, 2003). Ce biais de sélection est également lié au niveau d'éducation ; les individus plus éduqués répondent plus souvent et plus longuement aux questions ouvertes que les moins éduqués (Zuell et coll., 2014).

De plus, certaines réponses énoncées par les participants peuvent être vagues et difficiles à interpréter (Fink, 2003). C'est le cas de formulations brèves qui ne comportent qu'un seul mot. Les réponses courtes et vagues ne fournissent pas assez d'information pour être comprises dans les thématiques générées par l'ADL (Crain et coll., 2012 ; Pietsch & Lessman, 2018). En effet, le modèle s'appuie sur la présence simultanée de deux ou de plusieurs mots afin d'obtenir des thématiques significatives (DiMaggio et coll., 2013 ; Nelson, 2020). Il est donc possible que les résultats pour la thématique « gouvernement chinois et OMS » de la présente étude soient sous-estimés, étant donné que beaucoup de réponses comportent uniquement le terme "*China*" (voir les **tableaux 6, 7 et 8**

de la section « Résultats »). Hardy et coll. (2021) ont également remarqué une tendance chez certains répondants à n'écrire que le mot "*China*" comme entité à blâmer pour la crise de la COVID-19. Durant leurs entrevues, ils ajoutent que plusieurs participants ont parlé de la Chine sans en être certains et ont utilisé des expressions comme "*time will tell*" et "*I really don't know*" (Hardy et coll., 2021, p.14).

Ensuite, la formulation de la question ouverte ne permet pas de distinguer entre l'origine et la propagation de la COVID-19. Les réponses des participants pourraient être différentes en fonction de l'origine de la pandémie ou de la propagation du virus. Nguyen et coll. (2021) font une distinction entre ces deux concepts, en posant ces deux questions ouvertes : « Pourquoi la COVID-19 s'est-elle répandue si rapidement en Nouvelle-Zélande ? » et « Qui est à blâmer pour la COVID-19 ? ». Les résultats de leur enquête démontrent que les réponses à ces deux questions sont divergentes. En effet, les participants tendent à pointer du doigt les Asiatiques pour l'origine de la pandémie, et ils désignent le tourisme et les individus qui ne suivent pas les mesures sanitaires comme étant les causes de la propagation du virus (Nguyen et coll. 2021). Hardy et coll. (2021) ont opté pour une seule question : « Qui est responsable de la pandémie de la COVID-19 ? », à un moment où les enjeux d'imputabilité étaient capitaux dans un contexte d'évolution rapide de la situation (Hardy et coll. ; 2021). Poser deux questions aux participants, l'une portant sur l'origine du virus et l'autre portant sur sa propagation, peut permettre d'approfondir le phénomène de l'attribution du blâme durant la COVID-19. Compte tenu des résultats de la présente enquête, il est possible que les participants fassent allusion à l'origine du virus au cours des premières phases de la pandémie et à sa transmission au cours de ses phases ultérieures.

Une autre limite de cette recherche concerne l'utilisation du mot « responsable » dans la question ouverte : « Qui ou quoi est principalement

responsable de la crise actuelle de la COVID-19 ? ». En effet, ce terme a été qualifié d'équivoque, puisqu'il peut être confondu avec d'autres phénomènes ou peut regrouper des conceptions distinctes (Malle et coll., 2014). Cette expression peut comprendre quatre construits sous-jacents : le caractère fautif, la causalité, la prescience et l'intentionnalité (Malle et coll., 2014). Elle est également employée pour désigner l'obligation et le devoir, l'admissibilité à un jugement moral et le blâme (Malle et coll., 2014). Selon Fincham et Schultz (1981), le blâme est un concept différent, étant donné qu'il concerne le degré de responsabilité morale attribué à une personne. L'utilisation du mot « responsable » dans le cadre de la présente étude est tout de même judicieuse, puisqu'il est plus neutre que le terme « blâme ».

La technique de l'ADL comporte également quelques limites. Elles font notamment référence aux choix qui doivent être faits avant la production de l'analyse (Nelson, 2020). Le chercheur doit prendre plusieurs décisions relatives au prétraitement des données, incluant le fait de retirer les mots vides, de lemmatiser les expressions, d'exclure ou de remplacer des termes, ce qui peut produire différents résultats (Nelson, 2020). Par contre, les avis sur la question indiquent que les résultats obtenus doivent être évalués en fonction de leur utilité pour le contexte de recherche (Nelson, 2020). De plus, le chercheur ne connaît techniquement pas le nombre de thématiques qui doit être extrait avant l'analyse (Nelson, 2020). Il existe plusieurs méthodes qui permettent de le déterminer, et celles-ci font l'objet d'un débat entre les spécialistes (Nelson, 2020). Dans le cadre de la présente étude, ce nombre a été déterminé automatiquement à l'aide d'une méthode optimisée pour réaliser l'ADL (Harel et coll., 2023). Finalement, les thèmes qui ne comportent que très peu de réponses ont de faibles probabilités d'apparaître dans des thématiques (Roberts et coll., 2014), contrairement au codage manuel, où toutes les réponses, même les plus rares, sont catégorisées. Par exemple, certains participants ont blâmé les personnes d'origine asiatique pour la crise de la COVID-19. Toutefois, étant donné que peu de répondants sont

concernés, aucune thématique en lien avec les personnes d'origine asiatique n'a émergé.

Dernièrement, une des caractéristiques de l'ADL est qu'elle permet à un mot d'être attribué à plusieurs thématiques simultanément (Miner et coll., 2012 ; Gülzau, 2019). Cela permet de clarifier les diverses utilisations d'un même terme en fonction du contexte dans lequel il apparaît (DiMaggio et coll., 2013 ; Gülzau, 2019). Par exemple, le mot *China*, apparaît dans plus d'une thématique. Il se trouve dans les thématiques sur le gouvernement chinois, sur les voyages et la fermeture des frontières et sur les marchés aux animaux vivants. Toutefois, il se trouve également présent dans d'autres thématiques, tel que *Canadian government response* au temps 10. Cela est dû à la fréquence très élevée de l'utilisation du mot *China* par les participants. Ce phénomène peut rendre l'interprétation des résultats plus complexe et affecter la représentativité des autres mots associés à cette thématique (Kun & Cai, 2017). Une étape supplémentaire de prétraitement des données aurait pu être effectuée afin d'éliminer les mots fréquemment utilisés, tels que *China*. Toutefois, ce terme occupe un rôle central dans l'élaboration des thématiques et ne pouvait être supprimé.

## 7. Conclusion

Cette recherche a pour but de relever des tendances en ce qui concerne l'attribution de la responsabilité pour la crise de la COVID-19 au fur et à mesure de son évolution. Dans les premières phases de la pandémie, les communautés éloignées, telles que le gouvernement chinois, l'Organisation mondiale de la Santé (OMS) et les marchés aux animaux vivants, sont principalement identifiées comme responsables de la crise de la COVID-19. Au fil du temps, le blâme se



déplace vers les communautés locales, telles que les individus qui ne suivent pas les mesures sanitaires.

Cette recherche met en évidence le rôle de la proximité géographique et de la perception du risque dans la manière dont le public perçoit une pandémie. Elle démontre également que le blâme se déplace en fonction de périodes clés durant la pandémie, notamment l'instauration de nouvelles mesures sanitaires par le gouvernement. L'étude souligne donc l'importance d'une communication efficace et transparente de la part des autorités durant des crises de santé publique. En effet, l'incertitude inhérente aux crises sanitaires peut contribuer à la recherche de coupables et à la désignation de boucs émissaires, confirmant la nécessité pour les autorités d'adopter une approche globale et bien coordonnée de la gestion de ces événements. Les résultats de cette étude ont des répercussions sur les futurs événements sanitaires majeurs, car ils mettent en évidence le rôle primordial des stratégies de communication orientées vers la promotion d'un comportement responsable.

De plus, cette recherche vise à explorer l'intérêt des différentes fonctionnalités de l'ADL dans l'analyse de réponses ouvertes. L'utilisation de l'ADL permet d'extraire automatiquement des thématiques pertinentes à partir de grandes quantités de données textuelles, réduisant ainsi le temps et les efforts nécessaires consacrés habituellement au codage manuel. Elle offre également une approche plus objective et reproductible dans la catégorisation des réponses, éliminant le risque de biais ou d'incohérences qui pourraient survenir lors du codage manuel. La pertinence de cette étude est accentuée par l'importance de faire connaître cette méthode, encore peu utilisée en sociologie, ainsi que son potentiel d'analyse, ses avantages et ses limites (Lindstedt et coll., 2019). En effet, l'ADL permet d'ouvrir les possibilités de recherche à l'étude de plateformes riches en information textuelle, telles que les réseaux sociaux, les forums et les articles de presse. Il offre également de nouvelles perspectives pour

comprendre les phénomènes sociaux complexes qui sont souvent difficiles à quantifier.

Finalement, les enquêtes qui comprennent des questions ouvertes comportent des informations d'intérêt pour les chercheurs en sciences sociales. Ceux-ci s'intéressent particulièrement au rôle des variables sociodémographiques dans l'étude d'un phénomène, telles que le sexe, l'âge et l'orientation politique. Dans le cadre d'une enquête, ces informations peuvent influencer le contenu d'une réponse à une question ouverte. L'inclusion de ces variables dans le modèle permettrait d'obtenir une meilleure estimation des thématiques d'un point de vue statistique (Roberts et coll., 2014). Il serait également possible d'étudier comment les proportions des thématiques varient en fonction de ces variables (Roberts et coll., 2014). Les futures recherches devraient reproduire une procédure automatisée d'analyse de texte à l'aide de modèles qui permettent l'inclusion de covariables, telles que le *structural topic model* et le *dynamic topic model* (Lindstedt, 2019 ; Roberts et coll., 2014).

## Références

- Abeyasinghe, S. “Ebola at the Borders: Newspaper Representations and the Politics of Border Control.” *Third World Quarterly* 37, n° 3 (2016) : 452-67. <https://doi.org/10.1080/01436597.2015.1111753>.
- Agence de la santé publique du Canada. “Mises à jour : Doses de vaccin contre la COVID-19 administrées au Canada ». *Ensembles de données ; statistiques ; éducation et sensibilisation*. 2022. <https://sante-infobase.canada.ca/covid-19/vaccins-administres/>.
- Alam, S. & N. Yao. “The impact of preprocessing steps on the accuracy of machine learning algorithms in sentiment analysis.” *Computational and Mathematical Organization Theory* 25 (2019). <https://doi.org/10.1007/s10588-018-9266-8>.
- Allport, G. W. *The nature of prejudice*. New York Basic Books, 1979.
- Alvarez-Melis, D. & M. Saveski. “Topic Modeling in Twitter : Aggregating Tweets by Conversations.” *Proceedings of the International AAAI Conference on Web and Social Media* 10 n° 1 (2019) : 519-22. <https://doi.org/10.1609/icwsm.v10i1.14817>.
- « Antidote », 2022. <https://www.antidote.info/fr>.
- AskingCanadians (Delvinia) “Sample Profiling”, <https://corporate.askingcanadians.com/panel/sample-profiling/>
- Attema, A.E., O. L’Haridon, J. Raude & V. Seror. “Beliefs and Risk Perceptions About COVID-19: Evidence from Two Successive French Representative Surveys During Lockdown.” *Frontiers in Psychology* 12 (2021) : 1-16. <https://doi.org/10.3389/fpsyg.2021.619145>.
- Aubin, E. « Comment allez-vous célébrer Noël cette année ? » *TVA Nouvelles*, 24 décembre 2020. <https://www.tvanouvelles.ca/2020/12/24/et-vous-comment-allez-vous-celebrer-noel-cette-annee-1>.

- Banerjee, D., R.A. Kallivayalil & T.S.S. Rao. “The “othering” in pandemics: Prejudice and orientalism in COVID-19.” *Indian Journal of Social Psychiatry* 36 (2020): S102-106. <https://doi.org/10.4103/ijsp.ijsp.26120>.
- Barde, B.V. & A.M. Bainwad. “An overview of topic modeling methods and tools.” In *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)* (2017): 745-50. <https://doi.org/10.1109/ICCONS.2017.8250563>.
- Bardosh, K., A. de Figueiredo, R. Gur, E. Jamrozik, James Doidge, et al. “The Unintended Consequences of COVID-19 Vaccine Policy: Why Mandates, Passports and Restrictions May Cause More Harm than Good.” *BMJ Global Health* 7 (2022) : 1-14. <https://doi.org/doi:10.1136/bmjgh-2022-008684>.
- Baron, R.A., D. Byrne & G. Watson. *Exploring Social Psychology*. Second Canadian Edition. Scarborough, Ontario : Allyn and Bacon Canada, 1997.
- Barreneche, S.M. “Somebody to Blame: On the Construction of the Other in the Context of the COVID-19 Outbreak.” *Society Registry* 4, n° 2 (2020): 19-32, <https://doi.org/10.14746/sr.2020.4.2.02>.
- Bernstein, B. *Class, Codes and Control. Vol. 1 : Theoretical Studies Towards a Sociology of Language*. Routledge. 2003.
- Blei, D.M., A.Y. Ng & M.I. Jordan. “Latent Dirichlet Allocation.” *Journal of Machine Learning Research*. 3 (2003): 993-1022.
- Blei, D.M. “Probabilistic Topic Models.” *Communications of the ACM* 55, n° 4 (2012): 77-84. <https://doi.org/Doi:10.1145/2133806.2133826>.
- Bliss, M. *Montréal au temps du grand fléau : l'histoire de l'épidémie de 1885*. Montréal : Libre expression, 1996. 352 p.
- Bor, A., F.J. Jørgensen & M.B. Petersen. “Prejudice Against the Vaccinated and the Unvaccinated During the COVID-19 Pandemic: A Global Conjoint Experiment.” *PsyArXiv*, 2022. <https://doi.org/10.31234/osf.io/t2g45>.

- Boudon, R. « Le modèle des classes latentes ». *Revue française de sociologie* 3, n° 3 (1962) : 259-89. <https://www.jstor.org/stable/3319968>.
- Bottemanne, H., H. Y. Longuet et C. Gauld. « L'esprit prédictif : introduction à la théorie du cerveau bayésien ». *L'Encéphale* (2021) : 1-9. <https://doi.org/10.1016/j.encep.2021.09.011>.
- Bouguettaya, A. "Social and Cognitive Psychology Theories in Understanding COVID-19 as the Pandemic of Blame." *Frontiers in Psychology* 12, n° 672395 (2022): 1-7. <https://doi.org/10.3389/fpsyg.2021.672395>.
- Brandt, A.M. & P. Rozin, éd. *Morality and Health*. New York: Routledge, 1997.
- Bucher, R. "Blame and Hostility in Disaster." *American Journal of Sociology* 62, n° 5 (1957): 467-75, <https://doi.org/10.1086/222076>.
- Canadian Institute for Health Information. « Canadian COVID-19 Intervention Timeline | CIHI ». Consulté le 30 juillet 2023. <https://www.cihi.ca/en/canadian-covid-19-intervention-timeline>.
- Caron-Diotte, M., R. de la Sablonnière & N. Sadykova. "The malleability of collective memories: One year after the Tulip Revolution in Kyrgyzstan." *British Journal of Social Psychology* 61 (2022) : 192-213. <https://doi.org/DOI:10.1111/bjso.12476>.
- Casciano, D. "COVID-19, Discipline and Blame: From Italy with a Call for Alternative Futures." *Journal of Extreme Anthropology* 4, n° 2 (2020): E18-24. <https://doi.org/10.5617/jea.7864>.
- CBC News. "More than 100,000 Reported Coronavirus Cases in Canada | CBC News". 2020. <https://www.cbc.ca/news/canada/canada-coronavirus-covid-19-cases-1.5617042>.
- Cheng, X., Q. Cao & S. Shaoyi Liao. "An Overview of Literature on COVID-19, MERS and SARS: Using Text Mining and Latent Dirichlet Allocation." *Journal of Information Science* 48, n° 3 (2022) : 304-20. <https://doi.org/DOI : 10.1177/0165551520954674>.

- Choli, M. & D.J. Kuss. “Perceptions of Blame on Social Media during the Coronavirus Pandemic.” *Computers in Human Behavior* 124, n° 106895 (2021) : 1-12. <https://doi.org/10.1016/j.chb.2021.106895>.
- Crain, S.P., K. Zhou, S.-H. Yang & H. Zha. “Dimensionality Reduction and Topic Modeling: From Latent Semantic Indexing to Latent Dirichlet Allocation and Beyond.” In *Mining Text Data*. Boston, MA : Springer, 2012. [https://doi.org/10.1007/978-1-4614-3223-4\\_5](https://doi.org/10.1007/978-1-4614-3223-4_5).
- DeepL Traduction. <https://www.DeepL.com/translator>.
- de la Sablonnière, R., L. French Bourgeois & M. Najih. “Dramatic Social Change : A Social Psychological Perspective.” *Journal of Social and Political Psychology* 1, n° 1 (2013) : 253-72. <https://doi.org/doi:10.5964/jspp.v1i1.14>.
- de la Sablonnière, R. “Toward a Psychology of Social Change: A Typology of Social Change.” *Frontiers in Psychology* 8, n° 397 (2017) : 1-20. <https://doi.org/doi:10.3389/fpsyg.2017.00397>.
- de la Sablonnière, É. Lacourse, R., A. Dorfman, M. Pelletier-Dumas, J.-M. Lina, S. Dietlind, et coll. « Présentation de l’Enquête COVID-19 (Rapport technique N° 1) ». Université de Montréal, 2020.
- Denny, M.J., et A. Spirling. “Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It.” *Political Analysis* 26, n° 2 (2018) : 168-89. <https://doi.org/10.1017/pan.2017.44>.
- de Rosa, A.S. & T. Mannarini. “The ‘Invisible Other’: Social Representations of COVID-19 Pandemic in Media and Institutional Discourse.” *Papers on Social Representations* 29, n° 2 (2020) : 1-35.
- Deerwester, S., S.T. Dumais, G.W. Furnas, T.K. Landauer & R. Harshman. “Indexing by Latent Semantic Analysis.” *Journal of the American Society for Information Science* 41, n° 6 (1990) : 391-407.

- de Saussure, F. *Course in General Linguistics*. The Library of the University of California. 1983.
- DiMaggio, P., M. Nag & D. Blei. “Exploiting Affinities between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of U.S. Government Arts Funding.” *Poetics* 41 (2013): 570-606. <http://dx.doi.org/10.1016/j.poetic.2013.08.004>.
- Dionne, K.Y. & F.F. Turkmen. “The Politics of Pandemic Othering: Putting COVID-19 in Global and Historical Context.” *International Organization* 74, n° S1 (2020) : E213-30. <https://doi.org/10.1017/S0020818320000405>.
- Douglas, M. *Risk and Blame: Essays in Cultural Theory*. Routledge. 1992.
- Duwairi, R., et M. El-Orfali. « A Study of the Effects of Preprocessing Strategies on Sentiment Analysis for Arabic Text ». *Journal of Information Science* 40, n° 4 (2014) : 501-13. <https://doi.org/10.1177/0165551514534143>.
- Ferrante, V., É. Lacourse, A. Dorfman, M. Pelletier-Dumas, J.-M. Lina, D. Stolle & R. de la Sablonnière. “COVID-19, Economic Threat and Identity Status: Stability and Change in Prejudice Against Chinese People within the Canadian population” [Article accepté], *Frontiers in Psychology*, 2022.
- Finch, W. Holmes, M.E. Hernández Finch, C.E. McIntosh & C. Braun. “The Use of Topic Modeling with Latent Dirichlet Analysis with Open-Ended Survey Items.” *American Psychological Association* 4, n° 4 (2018): 403-24. <http://dx.doi.org/10.1037/tps0000173>.
- Fincham, F.D. & T.R. Shultz. “Intervening Causation and the Mitigation of Responsibility for Harm.” *British Journal of Social Psychology* 20 (1981): 113-20. <https://doi.org/10.1111/j.2044-8309.1981.tb00483.x>.
- Fink, A. *The Survey Handbook*. Second Edition. USA : SAGE, 2003.
- Forester, J., et G. McKibbin. “Beyond Blame: Leadership, Collaboration and Compassion in the Time of COVID-19.” *Socio-Ecological Practice Research* 2, n° 3 (2020): 205-16. <https://doi.org/10.1007/s42532-020-00057-0>.

- Goulias, Konstadinos G., R.M. Pendyala, et R. Kitamura. “Updating a Panel Survey Questionnaire.” *UC Berkeley: University of California Transportation Center*, 1991. <https://escholarship.org/uc/item/2g93m0gg>.
- Gouvernement du Canada. « COVID-19 : Preuve de vaccination au Canada et pour voyages à l'étranger pour les Canadiens ». Règlements, 2023. <https://www.canada.ca/fr/sante-publique/services/maladies/maladie-coronavirus-covid-19/vaccins/preuve-vaccinale.html>.
- Gülzau, F. « A Paradigm Shift in German Family Policy: Applying a Topic Model to Map Reform and Public Discourse, 1990–2016 ». *European Policy Analysis*, 2019, 1-20. <https://doi.org/10.1002/epa2.1072>.
- Hardy, L.J., A. Mana, L. Mundell, M. Neuman, S. Benheim & E. Otenyo. “Who Is to Blame for COVID-19? Examining Politicized Fear and Health Behavior through a Mixed Methods Study in the United States.” *PLOS ONE* 16, n° 9 (2021) : 23. <https://doi.org/10.1371/journal.pone.0256136>.
- Harel, S.-M., S. Ratté & L. Mottron. “Combinations of new heuristics in the automatic search for near-optimal Latent Dirichlet Allocation parameters by evaluating their stability through the clustering of repeated runs.” [Article soumis], 2023, 1-28.
- Hofmann, T. “Probabilistic latent semantic analysis.” In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, 289-96. UAI'99. San Francisco, CA, USA : Morgan Kaufmann Publishers Inc., 1999.
- Hoppe, T. “Spanish Flu: When Infectious Disease Names Blur Origins and Stigmatize Those Infected.” *American Journal of Public Health* 108, n° 11 (2016) : 1462-64. <https://doi.org/10.2105/AJPH.2018.304645>.
- Institut national de santé publique du Québec (INSPQ). « Données COVID-19 par vague selon l'âge et le sexe au Québec ». 2022. <https://www.inspq.qc.ca/covid-19/donnees/age-sexe>.



- Ittefaq, M., M. Abwao, A. Baines, G. Belmas, S.A. Kamboh & E. J. Figueroa. "A pandemic of hate: Social representations of COVID-19 in the media." *Analyses of Social Issues and Public Policy* (2022): 1-28. <https://doi.org/10.1111/asap.12300>.
- Jackman, S. *Bayesian Analysis for the Social Sciences*. Wiley. 2009.
- Jedwab, J. "Views on Vaccinated, Unvaccinated and Others and Intergroup Relations." *Association for Canadian Studies and Metropolis Canada*, 2021.
- Jensen, G. *The Path of the Devil: Early Modern Witch Hunts*. Rowman & Littlefield Publishers, Inc. United States, 2007.
- Joffe, H. "Public Apprehension of Emerging Infectious Diseases: Are Changes Afoot?" *Public Understanding of Science* 20, n° 4 (2011): 446-460. <https://doi.org/10.1177/0963662510391604>.
- Johns Hopkins Coronavirus Resource Center. "COVID-19 Map." 2022. <https://coronavirus.jhu.edu/map.html>.
- Jones, T., W. Doane & M. Attbom. *textmineR: Functions for Text Mining and Topic Modeling*, 2021. <https://CRAN.R-project.org/package=textmineR>.
- Kelly, M.P., N.M. Kriznik, A.L. Kinmonth & P.C. Fletcher. "The Brain, Self and Society: A Social-Neuroscience Model of Predictive Processing." *Social Neuroscience* 14, n° 3 (2019): 266-76. <https://doi.org/DOI:%2010.1080/17470919.2018.1471003>.
- Labbé, F., C. Pelletier, J. A. Bettinger, J. Curran, J. E. Graham, D. Greyson, N.E. MacDonald, et al. "Stigma and Blame Related to COVID-19 Pandemic: A Case-Study of Editorial Cartoons in Canada." *Social Science* n° 296 (2022) : 1-11. <https://doi.org/10.1016/j.socscimed.2022.114803>.
- Landau, M.J., A.C. Kay & J.A. Whitson. "Compensatory Control and the Appeal of a Structured World." *Psychological Bulletin* 141, n° 3 (2015) : 694-722. <http://dx.doi.org/10.1037/a0038703>.

- Lane, H. C. Howard & H.M. Hapke. *Natural Language Processing in Action: Understanding, Analyzing, and Generating Text with Python*. New York: Manning Publications Co, 2019.
- Liddy, E.D. “Natural Language Processing.” In *Encyclopedia of Library and Information Science*, 2nd Edition., 15. New York: Marcel Decker, Inc., 2001.
- Lin, B. M.L. Dietrich, R.A. Senior & D.S. Wilcove. “A Better Classification of Wet Markets Is Key to Safeguarding Human Health and Biodiversity.” *The Lancet Planet Health* 5, n° 6 (2021): e386-94. [https://doi.org/10.1016/S2542-5196\(21\)00112-1](https://doi.org/10.1016/S2542-5196(21)00112-1).
- Lindstedt, N.C. “Structural Topic Modeling for Social Scientists : A Brief Case Study with Social Movement Studies Literature, 2005–2017.” *Social Currents* 6, n° 4 (2019): 307-318. <https://doi.org/10.1177/2329496519846505>.
- Liu, Z., M. Li, Y. Liu & M. Ponraj. “Performance evaluation of Latent Dirichlet Allocation in text mining.” In *2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)* 4 (2011): 2695-98. <https://doi.org/10.1109/FSKD.2011.6020066>.
- Longstaff, P. & S.-U. Yang. “Communication Management and Trust: Their Role in Building Resilience to ‘Surprises’ Such as Natural Disasters, Pandemic Flu, and Terrorism.” *Ecology and Society* 13, n° 1 (2008): 3. <https://doi.org/10.5751/ES-02232-130103>.
- Lynch, S.M. & B. Bartlett. “Bayesian Statistics in Sociology: Past, Present, and Future.” *Annual Review of Sociology* 45, n° 1 (2019): 47-68. <https://doi.org/10.1146/annurev-soc-073018-022457>.
- Lyons, J. “Natural, Non-Natural and Unnatural Languages: English, Urdu and Other Abstractions.” In *Natural Language and Universal Grammar: Essays in Linguistic Theory*, 1<sup>st</sup> ed. Cambridge University Press, 1991.
- Malle, B.F., S. Guglielmo & A. E. Monroe. “A Theory of Blame.” *Psychological Inquiry* 25, n° 2 (2014): 147-86. <https://doi.org/10.1080/1047840X.2014.877340>.

- Martikainen, J. & I. Sakki. “How Newspaper Images Position Different Groups of People in Relation to the COVID-19 Pandemic: A Social Representations Approach.” *Journal of Community & Applied Social Psychology*, n° 31 (2021): 465-94. <https://doi.org/10.1002/casp.2515>.
- Mayor, E., V. Eicher, A. Bangerter, I. Gilles, A. Clémence & E.G.T. Green. “Dynamic Social Representations of the 2009 H1N1 Pandemic: Shifting Patterns of Sense-Making and Blame.” *Public Understanding of Science* 22, n° 8 (2012): 1011-24. <https://doi.org/10.1177/0963662512443326>.
- McCutcheon, Allan L. *Latent Class Analysis*. Newbury Park, Calif.: Sage Publications, 1987.
- Miner, G., D. Delen, J. Elder, A. Fast et coll., éd. “Chapter 13 - Clustering Words and Documents.” In *Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications*, 959-66. Boston : Academic Press, 2012. <https://doi.org/10.1016/B978-0-12-386979-1.00041-4>.
- Mohr, J.W. & P. Bogdanov. “Introduction—Topic Models : What They Are and Why They Matter.” *Poetics* 41 (2013): 545-69. <http://dx.doi.org/10.1016/j.poetic.2013.10.001>.
- Mondragon, N.I., L. Gil de Montes & J. Valencia. “Ebola in the Public Sphere: A Comparison Between Mass Media and Social Networks.” *Science Communication* 39, n° 1 (2017): 101-24. <https://doi.org/10.1177/1075547016688908>.
- Müller, A.C. & S. Guido. *Introduction to Machine Learning with Python: A Guide to Data Scientists*. O’Reilly. United States, 2016.
- Nelkin, D. & S.L. Gilman. “Placing Blame for Devastating Disease.” *Social Research* 55, n° 3 (1988): 361-72. <http://www.jstor.org/stable/40970508>.
- Nelson, L.K. “Computational Grounded Theory: A Methodological Framework.” *Sociological Methods & Research* 49, n° 1 (2020): 3-42. <https://doi.org/10.1177/0049124117729703>.

- Nguyen, T., S.M. Croucher, A. Diers-Lawson & E. Maydell. “Who’s to Blame for the Spread of COVID-19 in New Zealand?” *Communication Research and Practice* (2021): 1-18. <https://doi.org/10.1080/22041451.2021.1958635>.
- Niraula, N., R. Banjade, D. Ștefănescu & V. Rus. “Experiments with Semantic Similarity Measures Based on LDA and LSA.” In *Statistical Language and Speech Processing*, edited by A.-H. Dediu, C. Martín-Vide, R. Mitkov, et B. Truthe, 188-99. Berlin : Springer, 2013. [https://doi.org/10.1007/978-3-642-39593-2\\_17](https://doi.org/10.1007/978-3-642-39593-2_17).
- NLTK’s List of English Stopwords. <https://gist.github.com/sebleier/554280>.
- Panichella, A. “A Systematic Comparison of Search-Based Approaches for LDA Hyperparameter Tuning.” *Information and Software Technology* 130 (2021): 106411. <https://doi.org/10.1016/j.infsof.2020.106411>.
- Petersen, A.R. & D. Lupton. *The New Public Health: Health and Self in the Age of Risk*. Reprinted. London : Sage Publications, 2000.
- Pietsch, A.-S. & S. Lessman. “Topic modeling for analyzing open-ended survey responses.” *Journal of Business Analytics* 1, n° 2 (2018): 93-116. <https://doi.org/10.1080/2573234X.2019.1590131>.
- Proulx, T. & M. Inzlicht. “The Five ‘A’s of Meaning Maintenance: Finding Meaning in the Theories of Sense-Making.” *Psychological Inquiry* 23, n° 4 (2012): 317-35. <https://doi.org/10.1080/1047840X.2012.702372>.
- Rateau, P., J.-L. Tavani & S. Delouvé. “Social Representations of the Coronavirus and Causal Perception of Its Origin: The Role of Reasons for Fear.” *Health* (2023) : 1-20. <https://doi.org/10.1177/13634593211005172>.
- Reja, U., K.L. Manfreda, V. Hlebec & V. Vehovar. “Open-ended vs. Close-ended Questions in Web Questionnaires.” *Developments in Applied Statistics* 19 (2003): 159-77.

- Retail Council of Canada. “Proof of Vaccination Requirements by Province.” 2022. <https://www.retailcouncil.org/coronavirus-info-for-retailers/vaccination-requirements-by-province/>.
- Roberts, M.E., B.M. Stewart, D. Tingley, C. Lucas, J. Leder-Luis, S. K. Gadarian, et al. “Structural Topic Models for Open-Ended Survey Responses.” *American Journal of Political Science* 58, n° 4 (2014): 1064-82. <https://doi.org/DOI: 10.1111/ajps.12103>.
- Rouquette, M. -L. « Une taxinomie des peurs collectives ». Dans *Les peurs collectives*, 17-31. Société. Toulouse : Érès, 2013. <https://doi.org/10.3917/eres.delou.2013.01.0017>.
- Roy, M., N. Moreau, C. Rousseau, A. Mercier, A. Wilson & L. Atlani-Duault. “Ebola and Localized Blame on Social Media: Analysis of Twitter and Facebook Conversations During the 2014–2015 Ebola Epidemic.” *Culture, Medicine, and Psychiatry* 44, n° 1 (2020): 56-79. <https://doi.org/10.1007/s11013-019-09635-8>.
- Salthouse, T.A. “Selectivity of Attrition in Longitudinal Studies of Cognitive Functioning.” *The Journals of Gerontology: Series B* 69, n° 4 (2014): 567-74. <https://doi.org/10.1093/geronb/gbt046>.
- Scarpino, I., C. Zucco & M. Cannataro. “Characterization of Long COVID using text mining on narrative medicine texts.” In *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 2022-27*, 2021. <https://doi.org/10.1109/BIBM52615.2021.9669651>.
- Schutz, A. *The Phenomenology of the Social World*. Northwestern University Press., 1967 [1932].
- Shangguan, Z., M.Y. Wang & W. Sun. “What Caused the Outbreak of COVID-19 in China: From the Perspective of Crisis Management.” *International Journal of Environmental Research and Public Health* 17, n° 9 (2020): 3279. <https://doi.org/10.3390/ijerph17093279>.

- Shao, Y., X. Li, Y. Chen, L. Yu & B. Cui. “Sys-TM : A Fast and General Topic Modeling System.” *IEEE Transactions on Knowledge and Data Engineering* 33, n° 6 (2021): 2790-2802. <https://doi.org/10.1109/TKDE.2019.2956518>.
- Stinchcombe, A.L. *Constructing Social Theories*. New York: Harcourt, Brace & World. 1968.
- Strong, P. “Epidemic Psychology: A Model.” *Sociology of Health & Illness* 12, n° 3 (1990): 249-59. <https://doi.org/10.1111/1467-9566.ep11347150>.
- Tausczik, Y.R. & J.W. Pennebaker. “The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods.” *Journal of Language and Social Psychology* 29, n° 1 (2010): 24-54. <https://doi.org/DOI:10.1177/0261927X09351676>.
- Tudoran, A.A. “Why do internet consumers block ads? New evidence from consumer opinion mining and sentiment analysis.” *Internet Research* 29, n° 1 (2018): 144-66. <https://doi.org/10.1108/IntR-06-2017-0221>.
- Voigt, K. “COVID-19 Vaccination Passports : Are They a Threat to Equality?” *Public Health Ethics* 15, n° 1 (2022): 51-63. <https://doi.org/10.1093/phe/phac006>.
- Wagner, R.H. “The Theory of Games and the Balance of Power.” *World Politics* 38, n° 4 (1986): 546-76. <https://doi.org/doi:10.2307/2010166>.
- Wagner-Egger, P., A. Bangerter, I. Gilles, E. Green, D. Rigaud, F. Krings, et al. “Lay Perceptions of Collectives at the Outbreak of the H1N1 Epidemic: Heroes, Villains and Victims.” *Public Understanding of Science* 20, n° 4 (2011): 461-76. <https://doi.org/10.1177/0963662510393605>.
- Western, B. “Bayesian Analysis for Sociologists” *Sociological Methods & Research* 28, n° 1 (1999) : 7-34.
- Westrupp, E.M., C.J. Greenwood, M. Fuller-Tyszkiewicz, T.S. Berkowitz, L. Hagg & G. Youssef. “Text Mining of Reddit Posts.” *PLOS One* 17, n° 2 (2022). <https://doi.org/10.1371/journal.pone.0262529>.

White, A.I.R. “Historical Linkages: Epidemic Threat, Economic Risk, and Xenophobia.” *The Lancet* 395, n° 10232 (2020): 1250-51. [https://doi.org/10.1016/S0140-6736\(20\)30737-6](https://doi.org/10.1016/S0140-6736(20)30737-6).

Williams, T. & J. Betak. “A Comparison of LSA and LDA for the Analysis of Railroad Accident Text.” *Procedia Computer Science*, 130 (2018): 98-102. <https://doi.org/10.1016/j.procs.2018.04.017>.

Zuell, C., N. Menold, S. Körber. “The Influence of the Answer Box Size on Item Nonresponse to Open-Ended Questions in a Web Survey.” *Social Science Computer Review* 33, n° 1 (2014): 115-122.

## Annexe 1. Tableaux

**Tableau 9. Paramètres de l'ADL et leurs intervalles**

Paramètre	Description	Intervalle
$k$	Nombre de thématiques à extraire	[7, 20]
Alpha ( $\alpha$ )	Distribution des thématiques par document	[0. 0000001, 1]
Beta ( $\beta$ )	Distribution des mots par thématique	[0. 0000001, 1]
$N$	Nombre d'itérations de Gibbs	[200, 300]

**Tableau 10. Paramètres des modèles sélectionnés et scores de stabilité**

Temps de mesure	Nbre de thématiques ( $k$ )	Nbre d'itérations de Gibbs ( $N$ )	Alpha ( $\alpha$ )	Beta ( $\beta$ )	Score de stabilité
1	9	226	0,406 610 481	0,874 262 131	77,334 350 3
3	9	239	0,434 906 447	0,658 273 509	73,801 008 14
5	9	262	0,340 118 37	0,687 295 257	77,753 752 02
8	9	294	0,437 232 698	0,463 194 471	74,870 195 44
10	9	299	0,512 057 261	0,285 698 616	68,198 308 21
12	9	229	0,538 491 039	0,302 442 267	70,887 718 51

**Tableau 11. Répartition des répondants selon les catégories d'âge**

Catégories d'âge	Question ouverte		Échantillon total	
	Temps 1	Temps 12	Temps 1	Temps 12
18 à 24 ans	9,8 %	5,0 %	11,0 %	6,9 %
25 à 34 ans	15,4 %	10,9 %	16,6 %	12,6 %
35 à 44 ans	16,5 %	14,8 %	17,0 %	16,4 %
45 à 54 ans	16,6 %	17,3 %	16,0 %	16,5 %
55 à 64 ans	18,6 %	21,4 %	17,6 %	20,9 %
65 ans et plus	23,1 %	30,5 %	21,8 %	30,5 %



**Tableau 12. Sommaire des thématiques au temps 1**

Thématique	Étiquette	15 principaux mots
1	Chinese Government, WHO	government, chinese, China, worldhealthorganization, canadian, information, federal, outbreak, hide, chinese, communistparty, official, health, inaction, unitedstates, authority
2	Wet markets	China, market, wet, animal, wuhan, live, chinese, wild, food, eat, condition, open, unsanitary, bat, sell
3	Travel, Not closing Borders	government, China, enough, world, travel, border, soon, spread, virus, act, early, country, close, worldhealthorganization, federal
4	Lack of reaction/slow government response	lack, poor, pandemic, health, response, slow, country, global, China, government, virus, communication, food, preparation, hygiene
5	People not following guidelines	people, take, distance, virus, social, follow, seriously, spread, home, stay, go, everyone, rule, travel, isolate
6	No one is responsible	noone, responsible, hold, anyone, crisis, happen, nature, virus, particular, people, blame, person, good, nobody, everyone
7	China not taking the pandemic seriously	China, worldhealthorganization, virus, response, slow, world, government, country, information, pandemic, early, spread, unitedstates, start, trump
8	Virus, Nature, Animals	virus, human, animal, nature, bat, eat, China, unknown, transmission, mother, cause, wild, natural, come, new
9	Unknown	unknown, virus, country, blame, think, good, know, China, many, canada, come, unitedstates, make, much, unsure

**Tableau 13. Sommaire des thématiques au temps 3**

<b>Thématique</b>	<b>Étiquette</b>	<b>15 principaux mots</b>
1	Chinese Government, WHO	government, chinese, worldhealthorganization, soon, communist, unitedstates, trump, chinese communist party, canadian, world, virus, China, unknown, noone, cover
2	Wet markets	China, market, wet, animal, eat, bat, wuhan, food, open, wild, live, laboratory, virus, sell, meat
3	People not following guidelines	people, spread, virus, follow, take, measure, distance, covid, social, travel, think, seriously, believe, everyone, rule
4	Lack of reaction/slow government response	lack, government, response, health, public, federal, slow, early, preparedness, country, unknown, action, leader, level, preparation
5	Chinese Government slow response	China, government, world, enough, worldhealthorganization, virus, noone, act, react, early, quickly, slow, unitedstates, outbreak, response
6	No one is responsible	noone, virus, responsible, unknown, hold, everyone, think, China, blame, anyone, believe, global, point, us, particular
7	Virus, nature, animal	nature, happen, noone, mother, nobody, pandemic, answer, time, virus, prefer, unknown, nothing, blame, unsure, prepare
8	Travel, Not closing borders	virus, human, travel, country, transmission, pandemic, animal, China, close, spread, people, different, life, time, unsure
9	World countries	China, government, country, virus, border, answer, travel, worldhealthorganization, home, close, trump, unknown, soon, unitedstates, prefer

**Tableau 14. Sommaire des thématiques au temps 5**

<b>Thématique</b>	<b>Étiquette</b>	<b>15 principaux mots</b>
1	Chinese Government, WHO	government, chinese, worldhealthorganization, trump, China, unknown, canadian, world, slow, communist, federal, virus, begin, act, Trudeau
2	Wet markets	China, market, wet, animal, wuhan, open, live, eat, virus, unknown, sell, wild, chinese, communistparty, bat, Chinese
3	People not following guidelines	people, follow, virus, social, take, distance, measure, health, isolate, rule, seriously, travel, mask, think, covid
4	Travel, Not closing borders	China, government, enough, travel, border, close, worldhealthorganization, soon, slow, act, action, Trudeau, early, know, unknown
5	Lack of reaction/slow government response	virus, lack, China, government, spread, health, world, public, response, country, unknown, poor, information, slow, global
6	No one is responsible	noone, responsible, virus, unsure, hold, anyone, happen, nobody, unknown, particular, crisis, everyone, person, blame, pandemic
7	Nature, Nobody	nature, human, mother, noone, bat, world, unknown, animal, natural, eat, people, nobody, virus, good, coronavirus
8	Poor government response	lack, poor, country, government, virus, China, get, unknown, bad, noone, people, good, unitedstates, global, know
9	Virus	China, virus, spread, worldhealthorganization, travel, world, unknown, government, nobody, time, country, covid, unitedstates

**Tableau 15. Sommaire des thématiques au temps 8**

<b>Thématique</b>	<b>Étiquette</b>	<b>15 principaux mots</b>
1	People not following guidelines	people, follow, mask, rule, distance, wear, health, guideline, social, public, spread, home, measure, virus, stay
2	Chinese government, WHO	government, chinese, worldhealthorganization, China, canadian, federal, communist, unknown, slow, trudeau, response, poor, severity, scientist, situation
3	No one is responsible	noone, responsible, virus, nature, happen, think, nobody, anyone, blame, particular, unknown, hold, country, mother, person
4	Travel, Not closing borders	China, travel, market, international, country, wet, world, virus, unknown, everyone, poor, know, trump, global, hygiene
5	Wet markets	China, market, animal, wet, human, food, virus, eat, bat, people, live, chinese, wuhan, nobody, unknown
6	Lack of government early measures	enough, government, soon, worldhealthorganization, unknown, China, border, take, close, early, act, pandemic, virus, measure, everyone
7	Lack of reaction/slow government response	government, lack, canada, slow, action, unitedstates, response, federal, nature, country, China, traveller, early, unknown, virus
8	Virus	virus, China, spread, world, know, unknown, start, pandemic, country, laboratory, people, come, nobody, disease, stop
9	People	people, go, virus, trump, government, party, canada, everyone, unknown, say, like, come, unitedstates, world, get

**Tableau 16. Sommaire des thématiques au temps 10**

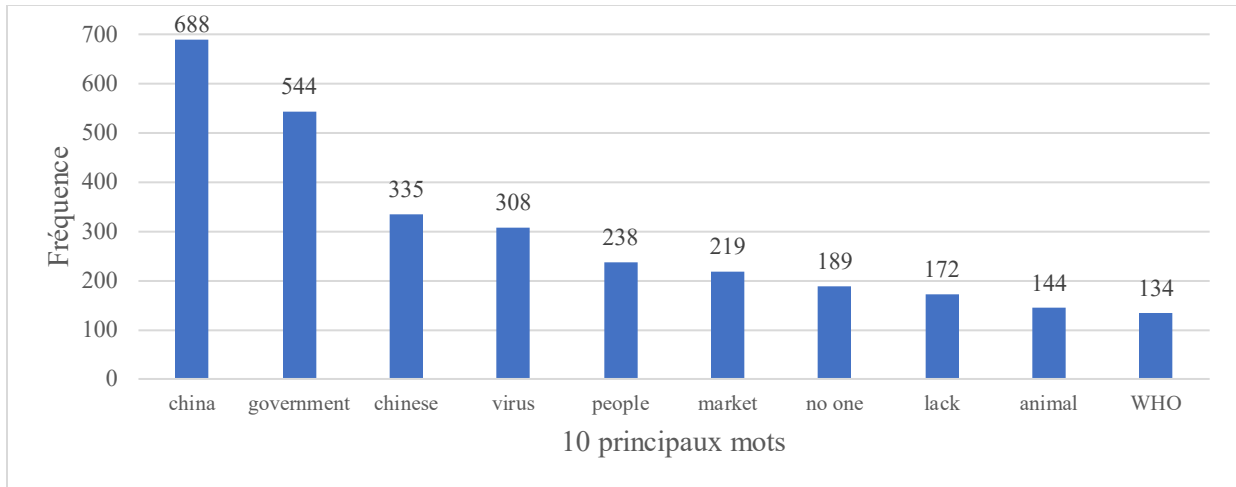
<b>Thématique</b>	<b>Étiquette</b>	<b>15 principaux mots</b>
1	People not following guidelines	people, follow, guideline, rule, health, mask, distance, wear, social, public, covid, measure, believe, refuse, irresponsible
2	Canadian government response	government, virus, federal, provincial, country, enough, measure, act, canadian, border, unknown, restriction, China, know, soon
3	Wet markets	China, market, animal, wet, human, virus, bat, eat, chinese, wuhan, live, wild, open, unknown, meat
4	Chinese government, WHO	government, chinese, worldhealthorganization, trump, China, federal, trudeau, unsure, American, border, unitedstates, unknown, information, nobody
5	Lack of reaction/slow government response	virus, China, lack, spread, action, slow, control, start, early, begin, crisis, worldhealthorganization, government, reaction, prevent
6	No one is responsible	noone, responsible, hold, anti, blame, virus, anyone, masker, unknown, crisis, pandemic, particular, person, think, nobody
7	Travel, Not closing borders	take, travel, world, seriously, people, everyone, pandemic, virus, enough population, currently, unsure, hygiene, isolation, international
8	Nature	nature, mother, politician, unknown, response, society, crisis, everyone, think, health, nobody, system, feel, science, life
9	Unknown	get, need, us covid, unknown, like, people, party, political, trudeau, China, keep, stupid, help, go

**Tableau 17. Sommaire des thématiques au temps 12**

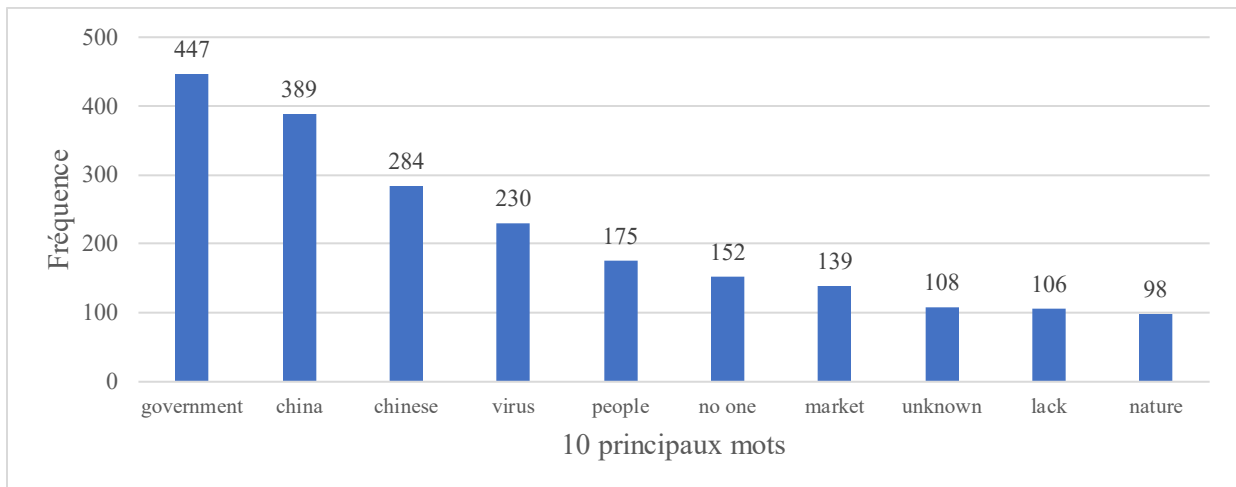
<b>Thématique</b>	<b>Étiquette</b>	<b>15 principaux mots</b>
1	People not following guidelines, unvaccinated	people, unvaccinated, follow, vaccinate, get, refuse, mask, rule, covid, spread, guideline, wear, care, other, measure
2	Canadian and Chinese governments	government, chinese, federal, provincial, canadian, enough, act, early, border, begin, China, slow, world, nobody, communist
3	Wet markets, Laboratory	China, market, laboratory, virus, animal, wuhan, Chinese, wet, worldhealthorganization, human, bat, start, nobody, food, unsure
4	Lack of reaction/slow government response	government, lack, health, restriction, public, slow, early, official, travel, information, response, pandemic, healthcare, begin, measure
5	Virus	virus, China, spread, world, country, know, take, travel, allow, noone, open, stop, covid, nobody, unknown
6	Unknown	unknown, responsible, noone, hold, anyone, virus, pandemic, think, unsure, make, start, politician, blame, crisis, person
7	Anti-vaccine, government	unknown, anti, medium, vaxxers, vaccine, good, right, government, Trudeau, poor, go, trump, worldhealthorganization, prepare, report
8	Nature	nature, mother, virus, human, life, like, global, travel, know, vaccine, unsure, pandemic, noone, earth, back
9	No one is responsible	noone, happen, nobody, everyone, particular, unknown, blame, person, pandemic, nothing, world, virus, natural, group, unsure

## Annexe 2. Graphiques

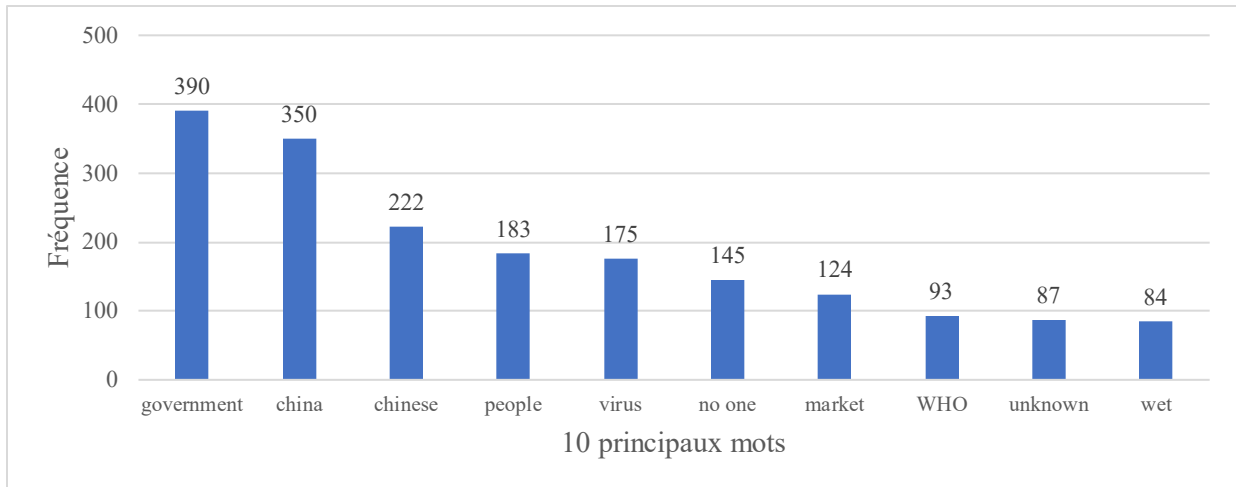
**Graphique 2. 10 principaux mots mentionnés au temps 1**



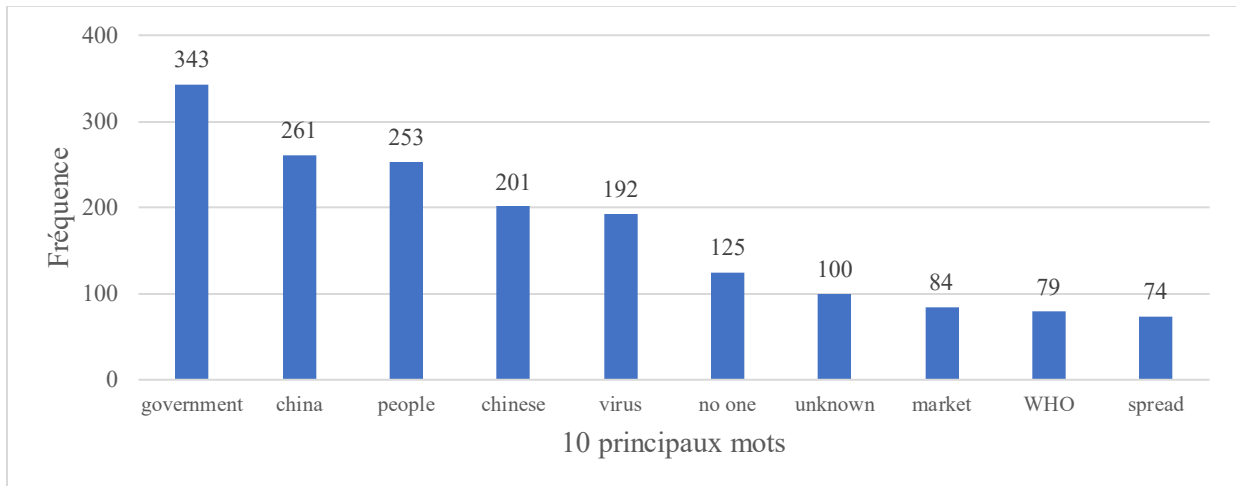
**Graphique 3. 10 principaux mots mentionnés au temps 3**



**Graphique 4. 10 principaux mots mentionnés au temps 5**

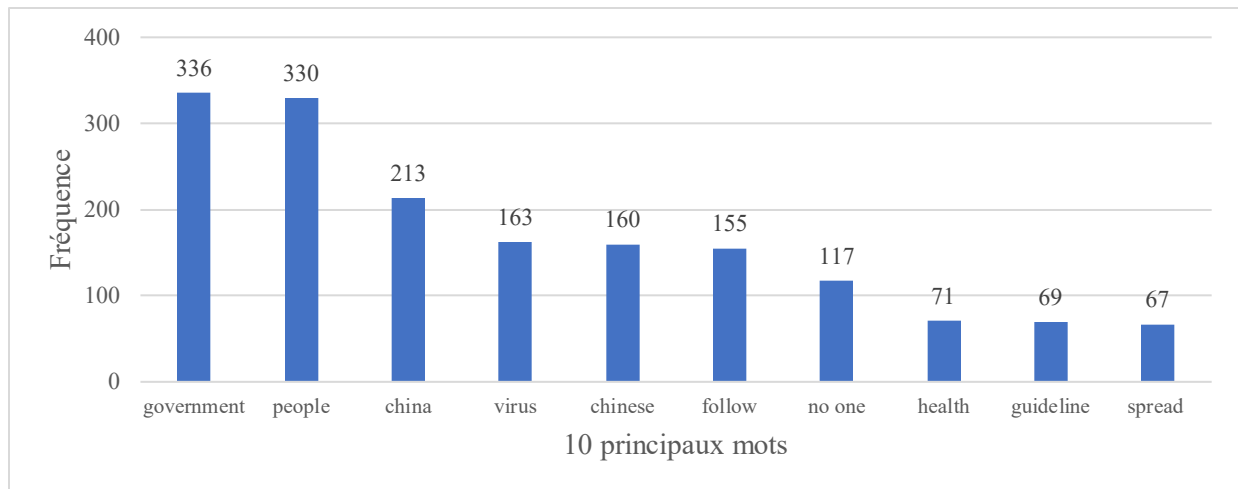


**Graphique 5. 10 principaux mots mentionnés au temps 8**

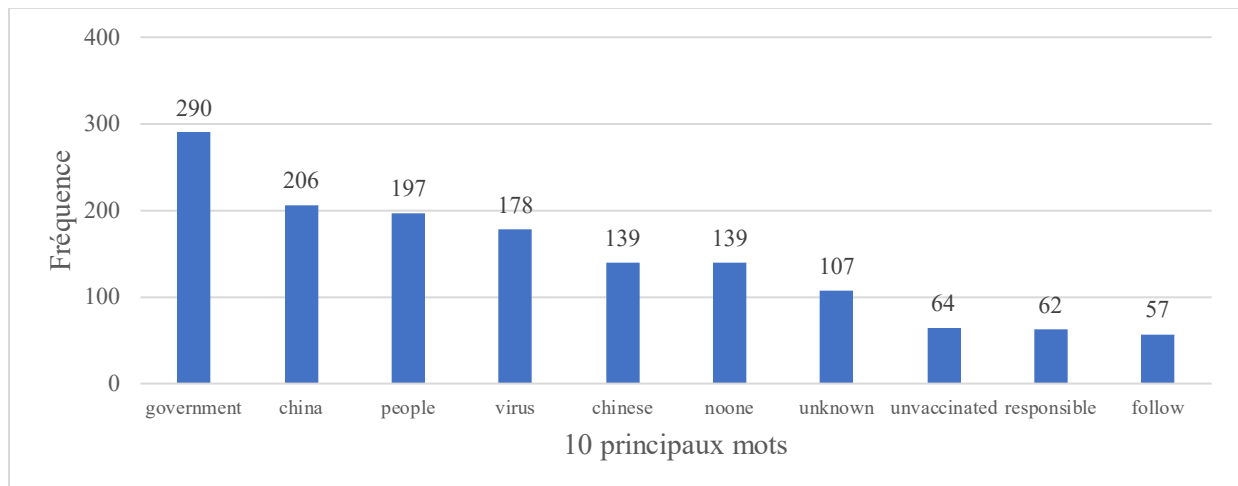




**Graphique 6. 10 principaux mots mentionnés au temps 10**



**Graphique 7. 10 principaux mots mentionnés au temps 12**



### Annexe 3. Termes remplacés ou réduits en un mot

```
df $ response <- str_replace_all(df $ response, c ("whole" = " ", "whose" = " ",  
"whom" = " ", "World Health Organization" = "worldhealthorganization",  
"world health organization" = "worldhealthorganization", "World Health  
people" = "worldhealthorganization people", "World Health organization" =  
"worldhealthorganization", "World health organization" =  
"worldhealthorganization", "WHO" = "worldhealthorganization", "THOSE  
worldhealthorganization" = "THOSE WHO", "PEOPLE  
worldhealthorganization" = "PEOPLE WHO", "the who" = "the  
worldhealthorganization", "look for the worldhealthorganization" = "look for the  
who", "Chinese and who" = "Chinese and worldhealthorganization", "China,  
who" = "China, worldhealthorganization", "China and who" = "China and  
worldhealthorganization", "Who and federal" = "worldhealthorganization and  
federal", "worldhealthorganization don't" = "WHO don't",  
"worldhealthorganization DID" = "WHO DID", "GUY  
worldhealthorganization" = "GUY WHO", "U.S.A." = "unitedstates", "U.S." =  
"unitedstates", "united states" = "unitedstates", "UNITED STATES" =  
"unitedstates", "United States" = "unitedstates", "United States of America" =  
"unitedstates", "Runitedstates" = "Russian", "Nunitedstates" = "NURSE",  
"Punitedstates" = "PUSSY", "ISSunitedstatesWITH" = "ISSUES WITH",  
"QunitedstatesION" = "QUESTION", "C.C.C.P." = "chinesecommunistparty",  
"C.C.P." = "chinesecommunistparty", "CCP." = "chinesecommunistparty",  
"CCP" = "chinesecommunistparty", "CCp" = "chinesecommunistparty", "c.c.p."  
= "chinesecommunistparty", "USA" = "unitedstates", "thounitedstatesnds" =  
"thousands", "C.C.P.," = "chinesecommunistparty,", "Chinese Communist  
Party" = "chinesecommunistparty", "Chinese Communist party" =  
"chinesecommunistparty", "Chinese communist party" =  
"chinesecommunistparty", "Chinese communist Party" =  
"chinesecommunistparty", "114heresa communist party" =  
"chinesecommunistparty", "P.R.C." = "peoplerepublicofChina", "People's  
Republic of China" = "peoplerepublicofChina", "people's Republic of China" =  
"peoplerepublicofChina", "people's republic of China" =  
"peoplerepublicofChina", "(People's republic)" = "(peoplerepublicofChina)",  
"PM" = "primeminister", "premiers" = "primeministers", "Premier" =  
"primeminister", "premier" = "primeminister", "Prime Minister" =
```

“primeminister”, “prime minister” = “primeminister”, “prime Minister” = “primeminister”, “U.K.” = “unitedkingdom”, “(UK)” = “unitedkingdom”, “FUnitedkingdomNG” = “Fucking”, “COCKSUnitedkingdomRS” = “COCKSUCKERS”, “LUnitedkingdomMOST” = “LUCK MOST”, “COVID-19” = “covid”, “Covid-19” = “covid”, “covid-19” = “covid”, “PPE” = “personalprotectiveequipment”, “PEEs” = “personalprotectiveequipment”, “Hapersonal protective equipmentN” = “HAPPEN”, “STOpersonal protective equipmentD” = “STOPPED”, “preventative” = “preventive”, “rest of 115heres, 115heresa” = “rest of 115heres, unitedstates”, “evil 115heresa” = “evil unitedstates”, “America and their 115heres” = “unitedstates and their 115heres”, “fed government” = “federal government”, “Fed and” = “Federal and”, “fed and” = “federal and”, “U.N.” = “unitednations”, “RETUnitednationsNG” = “RETURNING”, “JUSTIN TRUDEAUnitednations” = “justin 115heresa”, “pamers” = “pampers”, “U.C.P.” = “united conservative party”, “rUnited” = “reunited”, “retUnited” = “reunited”, “ppl” = “people”, “115heresa115115ion” = “organization”, “kingdomng” = “kingdom”, “kingomrs” = “kingdom”, “fUnited” = “united”, “Feds” = “federal”, “feds” = “federal”, “vs” = “versus”, “jan” = “115heresa”, “Jan-“ = “115heresa”, “feb” = “115heresa115”, “Feb.” = “February”, “januaryFebruary” = “115heresa 115heresa115”, “februaryuary” = “115heresa115”, “Februaryuary” = “115heresa115”, “individual” = “person”, “Individual” = “person”, “personly” = “individually”, “persons” = “people”, “All of us” = “everyone”, “all of us” = “everyone”, “no one” = “115heres”, “No one” = “115heres”, “No One” = “115heres”, “Not one” = “115heres”, “NO one” = “115heres”, “NO ONE” = “115heres”, “not one thing” = “nothing”, “not sure” = “unsure”, “Not sure” = “unsure”, “Not sure” = “unsure”, “not really sure” = “unsure”, “Not really sure” = “unsure”, “no idea” = “unknown”, “No idea” = “unknown”, “NO IDEA” = “unknown”, “No Idea” = “unknown”, “Don’t know” = “unknown”, “don’t know” = “unknown”, “Do not know” = “unknown”, “do not know” = “unknown”, “don’t KNOW” = “unknown”, “don’t really know” = “unknown”, “Don’t really know” = “unknown”, “do not really know” = “unknown”, “don’t Know” = “unknown”, “does not know” = “unknown”, “man-made” = “manmade”, “Man-made” = “manmade”, “LTC” = “chslD”, “chslDH” = “chslD”, “Long-Term Care” = “chslD”, “long term care homes” = “chslD”, “long-term care homes” = “chslD”, “long-term care” = “chslD”, “WWII” = “worldwartwo”, “insanitary” = “unsanitary”, “CHSLDs” = “chslD”, “Donald trump” = “trump”, “Donald Trump” = “trump”, “115heres trump” = “trump”, “Donald Duck” = “trump”, “Justin Trudeau” = “115heresa”, “justin 115heresa”

= "116heresa", "Justin 116heresa" = "116heresa", "idiot Justin" = "idiot 116heresa", "Justin is responsib"e" " "tru"ea" is responsibl", "worldhe"lt"organization k"ows" = "who knows", "TV" = "television", "Theresa Tam" = "tam", "116heresa tam" = "tam", "116heresa Tam" = "tam", "116heresa TAM" = "tam", "Theresa tam" = "tam", "Doug Ford" = "ford", "doug ford" = "ford", "Doug Fraud Ford" = "fraud ford", "Doug ford" = "ford", "Bill Gates" = "billgates", "Fauci and Gates" = "Fauci and billgates", "bill gates" = "billgates", "Anthony Fauci" = "fauci", "Jason Kenney" = "kenney", "British Columbia" = "britishcolumbia", "Xi Jinping" = "jinping", "President Xi" = "jinping", "Bojo the Clown Boris Johnson" = "borisjohnson", "as Johnson" = "as borisjohnson", "Hong Kong" = "hongkong", "New Zealand" = "newzealand", "North America" = "northamerica", "North American" = "northamerican", "South America" = "southamerica", "Francois Legault" = "legault", "we are all" = "everyone is", "WE ARE ALL" = "everyone is", "we all are" = "everyone is", "all the people on the planet" = "everyone on the planet", "All humans are" = "everyone is", "We're all" = "Everyone is", "We are all" = "everyone is", "We are ALL" = "everyone is", "all people. People" = "everyone. People", "CPC" = "Conservative Party of Canada"))

## Annexe 4. Liste des mots vides supprimés

"i", "me", "my", "myself", "we", "our", "ours", "ourselves", "you", "you're", "you've", "you'll", "you'd", "your", "yours", "yourself", "yourselves", "he", "him", "his", "himself", "she", "she's", "her", "hers", "herself", "it", "it's", "its", "itself", "they", "them", "their", "theirs", "themselves", "what", "which", "who", "whom", "this", "that", "that'll", "these", "those", "am", "is", "are", "was", "were", "be", "been", "being", "have", "has", "had", "having", "do", "does", "did", "doing", "done", "a", "an", "the", "and", "but", "if", "or", "as", "of", "at", "by", "for", "with", "to", "from", "up", "in", "out", "on", "off", "then", "once", "here", "there", "when", "where", "why", "how", "any", "both", "le", "such", "no", "nor", "not", "ne", "only", "own", "pa", "so", "too", "very", "s", "t", "can", "can't", "cannot", "will", "just", "don", "don't", "should", "should've", "now", "d", "ll", "m", "o", "re", "ve", "y", "ain", "aren", "aren't", "couldn", "couldn't", "didn", "didn't", "doesn", "doesn't", "hadn", "hadn't", "hasn", "hasn't", "haven", "haven't", "isn", "isn't", "ma", "might", "mightn", "mightn't", "mustn", "mustn't", "needn", "needn't", "shan", "shan't", "shouldn", "shouldn't", "wasn", "wasn't", "weren", "weren't", "won", "won't", "wouldn", "wouldn't", "na", "NA", "also", "etc", "b", "c", "d", "e", "f", "g", "h", "j", "k", "l", "m", "n", "o", "p", "q", "r", "s", "t", "u", "v", "w", "x", "y", "z", "ms", "th", "thereby", "thirdly", "zero", "really", "oh", "se", "hb", "nc", "thing", "yet", "mr", "sorry", "self", "thing", "thus", "non", "first", "one")

## **Annexe 5. Formulaire de consentement (français)**

### **Bienvenue au sondage sur la COVID-19 !**

Si vous acceptez de participer, vous serez invité(e) à remplir un questionnaire en ligne toutes les deux semaines pendant cinq mois. Remplir un questionnaire nécessitera environ 12 minutes de votre temps et pourrez toujours accepter ou refuser de participer. Plus particulièrement, accepter de participer aujourd'hui ne veut pas dire que vous acceptez de remplir tous les questionnaires, mais uniquement que vous acceptez de remplir le questionnaire d'aujourd'hui. Nous communiquerons avec vous plus tard pour remplir le prochain questionnaire. Toutefois, nous vous encourageons à remplir tous les questionnaires au cours des cinq prochains mois. **En remplissant tous les questionnaires, vous nous aiderez considérablement à comprendre les enjeux actuels liés à la maladie du coronavirus (COVID-19) et ceux des maladies dont nous pourrions souffrir à l'avenir.** Enfin, il n'y a pas de bonnes ni de mauvaises réponses aux questions. Notre prétest a révélé la gamme complète des réponses pour chaque question. Merci à l'avance de participer. Vous recevrez vos points de récompense chaque fois que vous remplirez un questionnaire.

### **Politiques publiques entourant la maladie à coronavirus (COVID-19)**

#### **FORMULAIRE D'INFORMATION ET DE CONSENTEMENT**

##### **Chercheurs :**

Roxane de la Sablonnière, professeure titulaire, Département de psychologie, Université de Montréal ;

Jean-Marc Lina, professeur, Département de génie électrique, École de technologie supérieure (ETS) ;

Dietlind Stolle, professeure James McGill, Département de science politique, Université McGill ;

Donald M. Taylor, professeur titulaire, Département de psychologie, Université McGill.

**Vous êtes invité(e)s à participer à un projet de recherche. Avant d'accepter, veuillez prendre le temps de lire ce document présentant les conditions de**

**participation au projet. N'hésitez pas à poser toutes les questions que vous jugerez utiles.**

## **A) RENSEIGNEMENTS AUX PARTICIPANTS**

### **1. Objectifs de la recherche**

Ce projet de recherche vise à mieux comprendre l'influence des politiques publiques entourant la maladie à coronavirus (COVID-19) sur les attitudes et le bien-être de la population.

### **2. Participation à la recherche**

Votre participation consiste à remplir un questionnaire qui nécessitera 12 minutes de votre temps, et ce, toutes les 2 semaines pendant 5 mois (10 fois au total). Vous serez avisé par courriel quand le moment sera venu de remplir le questionnaire de nouveau. Vous aurez une semaine environ pour répondre au questionnaire.

### **3. Risques et inconvénients**

Outre le temps nécessaire pour remplir les questionnaires (environ 12 minutes toutes les 2 semaines pendant 65 mois), il n'y a pas de risques particuliers, connus ou anticipés à participer à ce projet.

### **4. Avantages et bénéfices**

Il n'y a pas d'avantage particulier à participer à ce projet. Vous contribuerez cependant à l'avancement des connaissances en psychologie sociale et à une meilleure compréhension du bien-être des sociétés et des individus.

### **5. Confidentialité**

Les renseignements personnels que vous nous donnerez demeureront confidentiels. Aucune information permettant de vous identifier d'une façon ou d'une autre ne sera publiée. De plus, chaque participant à la recherche se verra attribuer un code et seuls les chercheurs et leur équipe pourront connaître son identité. Les données seront conservées sur un ordinateur sécurisé par un mot de passe. Toute information personnelle sera détruite sept ans après la fin du projet. Seules les données ne permettant pas de vous identifier seront conservées après cette période. Les données récoltées dans le cadre de cette étude (p. ex., les réponses aux questionnaires) seront rendues accessibles à la communauté

scientifique après avoir été rendues complètement anonymes. Puisque les données partagées seront rendues complètement anonymes, aucune information communiquée à d'autres chercheurs ne permettra de vous identifier.

## **6. Compensation**

Vous serez récompensé pour votre temps par l'entremise d'un incitatif nominal.

## **7. Droit de retrait**

Votre participation à ce projet est entièrement volontaire et vous pouvez à tout moment vous retirer de la recherche en cessant de répondre au questionnaire sans devoir justifier votre décision et sans conséquence pour vous. À votre demande, toutes les données qui vous concernent pourront aussi être détruites. Si vous désirez retirer toutes les données vous concernant de la recherche, veuillez communiquer avec le chercheur au numéro de téléphone indiqué ci-dessous. Cependant, une fois les résultats publiés, il sera impossible de retirer vos données des analyses et de modifier les résultats.

Si vous en ressentez le besoin, vous pouvez communiquer avec la chercheuse afin d'être redirigé vers des ressources psychologiques.

Si vous désirez avoir accès aux résultats généraux de l'étude, vous pouvez communiquer avec la chercheuse et un rapport de recherche vous sera envoyé.

Si vous avez des inquiétudes en lien avec vos droits ou les responsabilités des chercheurs, vous pouvez contacter le Comité d'éthique de la recherche en éducation et en psychologie.

## **B) CONSENTEMENT**

### **Déclaration du participant**

- Je comprends que je peux prendre mon temps pour réfléchir avant d'accepter ou de refuser de participer à la recherche.
- Je peux poser des questions à l'équipe de recherche et exiger des réponses satisfaisantes.
- Je comprends qu'en participant à ce projet de recherche, je ne renonce à aucun de mes droits ni ne dégage les chercheurs de leurs responsabilités.
- J'ai pris connaissance du présent formulaire d'information et de consentement et j'accepte de participer au projet de recherche.



**Pour toute question relative à l'étude ou pour vous retirer de la recherche,** veuillez communiquer avec Roxane de la Sablonnière au 514 343-6732 ou à [roxane.de.la.sablonniere@umontreal.ca](mailto:roxane.de.la.sablonniere@umontreal.ca)

Pour toute préoccupation sur vos droits ou les responsabilités des chercheurs concernant votre participation à ce projet, vous pouvez contacter le Comité d'éthique de la recherche en éducation et en psychologie par courriel à [cerep@umontreal.ca](mailto:cerep@umontreal.ca) ou par téléphone au 514 343-6111 ou encore consulter le site Web <http://recherche.umontreal.ca/participants>

Toute plainte relative à votre participation à cette recherche peut être adressée à l'ombudsman de l'Université de Montréal en appelant au 514 343-2100 ou en communiquant par courriel à [ombudsman@umontreal.ca](mailto:ombudsman@umontreal.ca) (**l'ombudsman accepte les appels à frais virés**).

**En répondant au questionnaire suivant,** je déclare avoir pris connaissance des renseignements ci-dessus, savoir que je peux obtenir les réponses à mes questions sur ma participation à la recherche auprès de la chercheuse et comprendre le but, la nature, les avantages, les risques et les inconvénients de cette recherche. Je consens librement à prendre part à cette recherche. Je sais que je peux cesser de participer en tout temps sans préjudice et sans devoir justifier ma décision.

J'accepte de participer à cette recherche.

Oui

Non

## **Annexe 6. Formulaire de consentement (anglais)**

### **Welcome to the COVID-19 Survey!**

If you agree to participate, you will be invited to fill out an online questionnaire every two weeks for a period of five months. It will take approximately 12 minutes to answer each questionnaire. Each time, you will have the option of accepting or refusing to participate. More specifically, accepting to participate today does not mean that you agree to fill out all the questionnaires. It means that you agree to complete the current one and to be contacted again for the next questionnaire. However, we encourage everyone to fill out all the questionnaires over the next five months. **By completing all the questionnaires, you will have made a very important contribution to our understanding of the current coronavirus (COVID-19) challenges and those that are sure to hit us in the future.** Lastly, there are no right or wrong answers to individual questions. Our pretesting revealed the complete range of answers for each question. Thank you so much for your help. You will receive your reward points every time you fill out a questionnaire.

### **Public Policies Regarding the Coronavirus (COVID-19)**

#### **INFORMATION AND CONSENT FORM**

##### **Researchers:**

Roxane de la Sablonnière, full professor, Psychology Department, Université de Montréal ;

Jean-Marc Lina, professor, Electrical Engineering Department, École de technologie supérieure (ETS) ;

Dietlind Stolle, James McGill Professor, Political Science Department, McGill University; Donald M. Taylor, full professor, Psychology Department, McGill University.

**You are invited to participate in a research project. Before agreeing, please read the conditions of participation below. Feel free to ask any question you find relevant.**

#### **A) INFORMATION FOR PARTICIPANTS**

### **1. Research objective**

This research project aims to better understand the impact of public policies regarding the coronavirus (COVID-19) on people's attitudes and well-being.

### **2. Participation**

Your participation entails filling out a questionnaire, which will take approximately 12 minutes of your time every 2 weeks for 5 months (10 questionnaires in total). You will receive an email informing you of the deadline for the next questionnaire. You will have about one week to answer the questionnaire.

### **3. Risks**

Other than the time needed to answer the questionnaires (about 12 minutes every 2 weeks for 5 months), there are no known risks of your participation in this study.

### **4. Benefits**

Although there are no particular concrete benefits for participating in our study, note that you will be playing a role in advancing our knowledge in social psychology and in better understanding society's and individual's well-being.

### **5. Confidentiality**

Personal information will remain confidential. None of the information that could lead to discovering your identity will be published. Furthermore, each participant will be assigned a code and the research team alone will know the participant's identity. Data will be kept on a password-secured computer. Any personal information will be destroyed seven years after the project has ended. Only data that cannot be traced back to participants will be kept after the seven-year period. After being rendered anonymous, the research data collected in this study (i.e. the answers to the questionnaire) will be made available to the scientific community. As the shared data will be completely anonymous, no information that could be used to identify you will be communicated to other researchers.

### **6. Compensation**

Your participation will be recognized through a nominal incentive.

### **7. Right to withdraw**

Your participation is fully voluntary. You can withdraw from the study at any moment by simply not answering the questionnaires any longer, without having to provide any explanation or justification. Your decision will not result in any repercussions. Upon your request, all the information you have provided will be destroyed. If you decide to withdraw the data related to you collected through the study, please inform the main researcher at the phone number below. However, it will be impossible to remove your data from the analyses and modify the results after the publication process.

If needed, you can contact the main researcher to be referred to psychological resources.

If you want to access the general results of the study, contact the researcher and a research report will be sent to you.

If you have concerns regarding your rights or researchers' responsibilities, please contact the Comité d'éthique de la recherche en éducation et en psychologie.

## **B) CONSENT**

### **Participant statement**

- I understand that I can take my time to consider participating in the study before giving my consent.
- I can ask the research team questions and demand satisfactory answers.
- I understand that my participation in this research project does not negate my right to withdraw and does not release researchers from their responsibilities.
- I have read and understood the information in this consent form and I agree to participate in the research project.

**For any question regarding the study or to withdraw from the research project**, please contact Roxane de la Sablonnière at 514-343-6732 or email her at [roxane.de.la.sablonniere@umontreal.ca](mailto:roxane.de.la.sablonniere@umontreal.ca)

If you have concerns regarding your rights or researchers' responsibilities, please contact the Université de Montréal's Comité d'éthique de la recherche en

éducation et en psychologie via email [cerep@umontreal.ca](mailto:cerep@umontreal.ca), by phone at 514-343-6111 or via their website at <http://recherche.umontreal.ca/participants>

Complaints related to your participation can be addressed to the ombudsman of the Université de Montréal by calling 514-343-2100 or via email at [ombudsman@umontreal.ca](mailto:ombudsman@umontreal.ca) (**the ombudsman accepts collect calls**).

By answering the following question, I declare that I have read the information above. I am aware that I can demand answers to my questions concerning my participation in the research project and I understand the goal, nature, benefits, risks and inconveniences of the research project. I know that I can withdraw my participation at any time without any consequences or justifications.

I agree to participate in this research project.

Yes

No