

Université Paris-Saclay - Institut Polytechnique de Paris (ENSAE)

Master 2 Sociologie Quantitative et Démographie & Master Data Sciences et
Sciences Sociales

De l'intime au numérique :

Étude des témoignages de violences sexuelles dans
le mouvement #MeToo sur Twitter en France

— Clara Le Gallic-Ach —

Mémoire de recherche soutenu en octobre 2023

Sous la direction de :

Alice Debauche (Université de Strasbourg) et Etienne Ollion (CNRS - Polytechnique)

Jury :

Paola Tubaro (CNRS - CREST), Alice Debauche (Université de Strasbourg), Etienne
Ollion (CNRS - Polytechnique)

Sommaire

Introduction	4
1 Partie 1 : Emergence des violences sexistes et sexuelles comme question politique et scientifique	8
1.1 Chapitre 1 : Une chronologie contemporaine de l'étude des VSS .	8
1.1.1 Un élan féministe aux avancées juridiques et historiques sur la question du viol	8
1.1.2 Les données administratives disponibles : évolutions et limites	10
1.1.3 Les enquêtes de victimation : une autre approche statistique	13
1.1.4 D'autres espaces d'expression investis par les victimes et la recherche	16
1.2 Chapitre 2 : Le mouvement #MeToo et l'espace Twitter	21
1.2.1 Émergence et évolution du mouvement #MeToo	21
1.2.2 Le corpus numérique de #Metoo ou la constitution d'archives inédites	24
2 Partie 2 : Méthodologie de collecte et d'analyse des données	27
2.1 Chapitre 3 : Les données de l'étude	27
2.1.1 Comprendre la structure de Twitter et de ses contenus . .	27
2.1.2 L'utilisation de cette structure dans le mouvement #Metoo	28
2.1.3 Deux formes de collecte : la base INA et l'API Twitter . .	29
2.2 Chapitre 4 : La méthodologie	32
2.2.1 La classification automatique de données textuelles	33

2.2.2	D'autres méthodes pour l'analyse des données textuelles	40
2.2.3	L'analyse des métadonnées	42
2.3	Chapitre 5 : Les mesures de la qualité des méthodes	42
3	Partie 3 : Nature, contenu et dynamique des témoignages sur Twitter	48
3.1	Chapitre 6 : Temporalité et persistance de #Metoo sur Twitter	48
3.1.1	Chronologie et volumétrie	48
3.1.2	Les types de contenu partagés et leurs évolutions	52
3.2	Chapitre 7 : Les spécificités des récits numériques étudiés	56
3.2.1	Des formes de témoignages propres à la plateforme	56
3.2.2	Des pratiques de publication propres aux témoignages	60
4	Partie 4 : Analyse comparative de différents espaces de déclaration	67
4.1	Chapitre 8 : Comparaison entre déclarations numériques et statistiques nationales	68
4.1.1	Prévalence des violences anciennes dans les témoignages numériques : une confirmation des tendances observées depuis #MeToo	68
4.1.2	Différences de contextes de violences : des écarts qui interrogent les dynamiques de déclaration	72
4.2	Chapitre 9 : Comparaison entre les déclarations en ligne et hors ligne	76
4.2.1	De la porosité entre les espaces numériques, médiatiques et politiques : les cas de #BalanceTonPorc et #MetooInceste	76
4.2.2	Entre démocratisation et <i>silence des dominés</i>	79

Conclusion	85
5 Annexes	88

Introduction

Le mouvement #Metoo a marqué un tournant dans la manière dont les violences sexuelles et sexistes (VSS) sont exprimées, perçues et comprises au sein de la société contemporaine. Ce mouvement, apparu en 2017, s’inspirant du travail de l’activiste Tarana Burke qui avait initialement créé l’expression aux États-Unis en 2006 pour soutenir les victimes de violences sexuelles, a permis de mettre en lumière l’ampleur des VSS grâce aux nombreux témoignages partagés sur les réseaux sociaux, notamment Twitter¹. En France, le mouvement a pris un ancrage spécifique avec une forte mobilisation numérique, de nombreuses affaires médiatisées et la création de nouveaux hashtags tels que #Balancetonporc ou #Metooinceste, révélant des mécanismes et contextes de violences parfois mal connus. Cette mobilisation numérique a également mis en évidence la tension entre l’intime et le public, impliquant parfois les expériences personnelles dans une réflexion collective, médiatique ou politique.

Ce mémoire s’inscrit dans une réflexion statistique et sociologique sur la parole des victimes, leurs dynamiques et la connaissance que l’on en a. Twitter constitue un accès inédit à une pratique numérique du témoignage qui permet d’enrichir des connaissances partielles sur ces violences et sur leur expression. Le recours à une plateforme numérique telle que Twitter permet de collecter des témoignages non filtrés, spontanés, qui échappent en partie aux contraintes institutionnelles. C’est dans ce contexte que cette recherche cherche à analyser l’émergence et la dynamique des témoignages de VSS sur Twitter, cinq ans après le lancement de #Metoo.

1. Le changement de nom de Twitter à X n’est pas intégré dans ce travail.

Dans cette étude, le terme "Violences Sexuelles et Sexistes" (VSS) sera utilisé pour définir un large spectre de violences, ciblant disproportionnellement les femmes, les minorités de genre et de sexualité. La littérature s'accorde pour inclure les VSS au sein d'un éventail de faits devant inclure le contexte et l'accumulation à leur interprétation. Les violences physiques et sexuelles y sont pensées dans la continuité de violences psychologiques ou économiques, poursuivant les travaux de Liz Kelly sur le continuum des violences (Kelly, 1987). Ma démarche n'est jamais celle de caractérisation juridique et admet que les violences peuvent s'inscrire dans un contexte ou une période de violences. Aussi, les données de cette étude n'émettent aucune supposition sur le genre ou l'orientation sexuelle des personnes concernées. Pourront donc être visibles dans les données : différentes formes de violences sexuelles, des cas de multivictimation, ainsi que des facteurs contextuels tels que le mauvais accueil de la parole des victimes, des menaces, des traumatismes. La notion de gravité perçue sera néanmoins abordée, et les termes "violences", "violences sexuelles" ou "VSS" seront utilisés tout au long du mémoire.

Les enjeux de ce travail sont multiples. D'abord, il s'agit d'un enjeu d'archivage : les témoignages publiés sur les réseaux sociaux constituent une archive précieuse des mobilisations contre les VSS, d'autant plus pertinente que les plateformes comme Twitter modifient régulièrement leurs règles d'utilisation, restreignant l'accès aux données. Ensuite, des enjeux méthodologiques se posent : comment recueillir, analyser et préserver ces données de manière rigoureuse ? Ce mémoire propose une approche mixte, combinant des outils de collecte a posteriori via l'API de Twitter et l'utilisation de la base de données journalières de l'Institut National de l'Audiovisuel (INA). Ces deux sources ont permis de constituer un

corpus de plus de 330 000 tweets uniques, recueillis entre 2017 et 2022, incluant des témoignages qui ne sont plus accessibles en raison de suppressions de comptes ou de contenus.

La méthodologie employée pour cette recherche repose sur des techniques de traitement automatique du langage naturel (TAL ou NLP) afin de classifier et analyser le contenu textuel des témoignages¹. L'utilisation de modèles de classification automatique permet d'identifier des catégories de discours (témoignages, opinions, soutiens) et de distinguer les contextes des violences rapportées. En parallèle, l'analyse des métadonnées associées aux tweets (telles que les dates de publication, le nombre de likes ou la structure des réseaux sociaux des auteurs) permet d'apporter une dimension complémentaire pour comprendre les dynamiques de diffusion et d'engagement. L'évaluation de la qualité des méthodes employées fait également partie intégrante de cette étude afin d'assurer la fiabilité des résultats.

La question centrale qui guide cette recherche est la suivante : comment les témoignages numériques partagés dans le cadre de #Metoo en France peuvent-ils enrichir notre compréhension des VSS, en explorant la parole des victimes, leur rapport à leur expérience, et la tension entre l'intime, le collectif et le public ? Pour y répondre, ce mémoire se structure en quatre grandes parties. La première partie explore l'émergence des VSS comme question politique et scientifique. La deuxième partie présente la méthodologie de collecte et d'analyse des données, en décrivant en détail les outils employés. La troisième partie s'intéresse au contenu

1. Tous les tweets cités dans ce travail sont pseudonymisés : aucun nom d'utilisateur n'est cité et le contenu textuel est modifié ou tronqué.

et à la dynamique des témoignages sur Twitter, en analysant les spécificités de ces récits numériques. Enfin, la quatrième partie propose une analyse comparative entre les témoignages en ligne, les données statistiques nationales, ainsi que d'autres espaces médiatiques, afin de mettre en évidence les différences et similitudes dans les dynamiques de déclaration des victimes.

Ainsi, ce mémoire contribue à enrichir la compréhension des formes contemporaines de prise de parole sur les violences, en examinant le rôle des réseaux sociaux dans la visibilité des VSS, en analysant ce que disent les témoignages sur les besoins et le parcours des victimes, ainsi que sur la réflexion collective et les dynamiques de pouvoir à l'œuvre, en particulier dans l'imbrication du numérique avec d'autres espaces de prise de parole. De plus, ce travail propose des pistes de réflexion méthodologiques et statistiques sur les données actuellement disponibles en matière de VSS, sur la place des données numériques, et sur le rôle des témoignages dans l'action publique.

Partie 1 : Emergence des violences sexistes et sexuelles comme question politique et scientifique

Chapitre 1 : Une chronologie contemporaine de l'étude des VSS

1.1.1 Un élan féministe aux avancées juridiques et historiques sur la question du viol

En France, les études portant sur les violences sexuelles ont largement émergé à la suite des mobilisations féministes des années 1970 qui ont permis de faire reconnaître progressivement la réalité des violences vécues. En 1980, la nouvelle loi sur le viol permet d'en moderniser la législation, le reconnaissant comme un crime quand les attentats à la pudeur (dits agressions sexuelles aujourd'hui) sont reconnus comme des délits. L'exception conjugale est abandonnée et peu à peu d'autres nuances sont ajoutées à la définition pénale du viol, comme la menace en 1992 qui s'ajoute à la surprise, la contrainte et la violence. La culpabilité change alors légalement de camp et la victime peut être reconnue en tant que telle juridiquement. Ces décennies sont aussi celles de la reconnaissance et le traitement des violences comme une question qui doit légitimement être traitée pour les politiques publiques ([Dauphin, 2023](#)).

Ces avancées sont accompagnées par des travaux de recherche historiques, sociologiques, politiques ou psychologiques. Il faut alors reconnaître que les productions scientifiques sur les violences sexuelles sont principalement portées par des féministes ([Corrin et al., 1997](#)) et la question est souvent abordée par le

prisme des études sur les rapports de genre et le contrôle social (Hanmer, 1977) (Guillaumin, 1978), dans la continuité des *feminist legal studies* et *gender studies* aux Etats-Unis (Schwartz, 2002). Nous verrons aussi la réalité de cette approche féministe dans les enjeux d'archivage et de mémoire collective qui sont au coeur des témoignages des violences sexuelles sur Twitter.

Les mouvements féministes des années 1970 participent à libérer les paroles autour des sexualités, de leurs tabous et des problèmes intraséquelement liés à la notion de domination tels que les violences sexuelles. Les premiers articles contemporains qui abordent ces questions sont surtout des travaux historiques sur la condition et la place des femmes ou des rapports judiciaires et travaux juridiques sur la violence. Les violences sexuelles sont rarement au centre des recherches mais apparaissent de plus en plus dans les mondes universitaire et judiciaire. Ainsi Chaperon (2002), dans *L'histoire contemporaine des sexualités en France*, évoque un "chantier en plein essor" des violences sexuelles dans la recherche. Les approches historiques se développent aussi et nous voyons Louis (1994) faire du droit de cuissage le sujet de son ouvrage, Vigarello (1999) traiter de l'histoire du viol en insistant sur la complexité de cette violence souvent tolérée. Celui-ci révèle aussi la difficulté de traiter de ce sujet dont les données quantitatives sont minces et met notamment en avant "un véritable blocage du processus judiciaire" sur le sujet. Ce constat déplorant le manque de matière pour traiter de l'histoire des femmes contemporaines et particulièrement de leur victimation était déjà ouvertement revendiqué par Sohn (1981) qui dresse un état des lieux de la vie domestique des femmes du début du 20e siècle dans la sphère privée. Cet article est un des points de départ de la recherche interdisciplinaire sur les violences sexuelles puisqu'il aborde par une approche quantitative les violences domestiques

(familiales et conjugales) par un recueil et traitement de 354 archives judiciaires. En formant un échantillon détaillé, Anne-Marie Sohn apporte une analyse éclairée sur les crimes et le regard des intermédiaires sur ceux-ci, en scrutant les mots employés par les policiers, témoins et agresseurs pour remarquer la faible place laissée aux femmes et aux victimes, en tant que victimes. Elle dresse également une cartographie sociale et géographique des archives, une rigueur essentielle et objective qui éclaire ses données plutôt pauvres et biaisées. Elle conclut en mettant en évidence la limite des données et outils dont elle dispose tout en dénonçant les biais et contraintes des archives judiciaires.

Les conclusions de ce travail seront partiellement remis en cause par l'élargissement de la base de données, mais non le postulat initial : l'histoire des femmes doit s'appuyer sur les sources manuscrites pour se dégager de la subjectivité.

Anne-Marie Sohn, 1981

La conclusion d'Anne-Marie Sohn soulève des enjeux clefs des données sur les violences sexuelles : la nécessité d'élargir nos sources de données et d'en améliorer la qualité. En effet, la restitution des témoignages par les services ministérielles sont souvent biaisés par la réinterprétation des faits par la victime, l'interlocuteur ou l'interlocutrice et les contraintes légales. Pour autant, depuis les années 1980, les données sur les violences sexuelles se sont déjà améliorées et sont à la fois plus nombreuses, mieux recueillies et mieux restituées.

1.1.2 Les données administratives disponibles : évolutions et limites

Jusqu'en 2006, le ministère de l'Intérieur publiait chaque année un recueil statistique dans le but déclaré de fournir des moyens d'analyse de la criminalité et

de guider les services chargés de la lutte contre la délinquance. Depuis 2009, cette tâche incombe à l'Observatoire national de la délinquance et des réponses pénales (ONDRP). Ces données administratives reposent sur des critères d'enregistrement inchangés depuis 1972 : seuls les faits constatés par procès-verbal transmis au parquet sont enregistrés. Cependant, il convient de noter que les statistiques policières ne reflètent pas les mêmes réalités que celles du ministère de la Justice ([Interstats Méthode N° 16, 2019](#)), notamment car la majorité des signalements n'auront pas de traitement judiciaire. Les méthodes, protocoles, qualifications et objectifs de publications posent des problèmes pour la retranscription et la comparaison des données ministérielles. De plus, l'évolution des lois et des définitions pénales en termes de violences sexuelles pose aussi le problème de suivi longitudinal. Depuis 2013, la Mission interministérielle pour la protection des femmes contre les violences et la lutte contre la traite des êtres humains ([MIPROF](#)) a été désignée comme l'observatoire national des violences faites aux femmes. La MIPROF a pour responsabilité de rassembler, analyser et diffuser les informations et données relatives aux violences faites aux femmes. Elle contribue également à la réalisation d'études, de travaux de recherche et d'évaluation dans le domaine de la protection des femmes victimes de violences.

Un effort a donc été fait sur la production de données nationales annuelles, avec des statistiques précises et des protocoles détaillés publiés. Cependant, les limites de ces données résident dans leurs contextes de production. Différents freins biaisent la réalité des données ministérielles. Premièrement, seule une minorité de victimes fait porter leur agression à la connaissance de la police ou gendarmerie. Dans les cas de viols, tentatives de viol et agressions sexuelles, seules 6% des victimes auraient plainte en 2022 ([SSMSI, 2022](#)). Ensuite, moins de 3 plaintes sur

10 font l'objet de poursuite parmi lesquelles environ 35% pourront bénéficier d'un non-lieu (Infostat Justice n°160, 2018).

Un enjeu important se joue lors de l'accueil de la parole des victimes puisque celles-ci ont souvent peur de la police, honte ou sont en fort niveau de stress (Linnet and Nizard, 2004). Les policiers ont souvent été accusés d'encourager les victimes à renoncer à déposer plainte, de leur témoigner de l'hostilité et de remettre en question la crédibilité de leurs témoignages (Delage, 2017; Mossuz-Lavau, 2002). De plus, le récit qui doit ensuite être recueilli est soumis à des obligations protocolaires et à l'interprétation des intermédiaires policiers. En effet, l'accueil de cette parole se fait par *"la production, par les policiers, d'un « script pénal de sexualité contrainte », c'est-à-dire d'un récit détaillé de l'interaction, ordonné suivant une chronologie, informant sur des caractéristiques sociales des victimes et des mis en cause"* (Perona, 2022) qui sera ensuite considéré retranscription officielle du vécu de la victime. L'approche des policiers est souvent très subjective, teintée de préjugés et de clichés autour des violences sexuelles et des victimes (Parratt and Pina, 2017), et dégradée par des jugements sociaux ou raciaux sur les victimes ou les coupables ainsi que des niveaux de connaissances sur les violences sexuelles très disparates et parfois très limités (Perona, 2022). Ces éléments ont ainsi des conséquences sur le dépôt de plainte mais également sur la qualification des faits, qui est parfois méconnue par les victimes mais aussi par les policiers (Perona, 2022). Ces récits officiels (quand ils sont en effet rédigés) sont alors teintés des normes contemporaines et individuelles des intermédiaires judiciaires. Ces archives sont alors le reflet de récits personnels maquillés par le contexte dans lequel il ont été recueillis (Martin, 1996).

Finalement, si les méthodes statistiques et la transparence des données des

institutions ministérielles sont satisfaisantes, le contexte de production et l'origine de ces données fournissent une version tronquée et biaisée de la réalité des violences sexuelles.

1.1.3 Les enquêtes de victimation : une autre approche statistique

L'une des solutions pour compléter les données judiciaires sont les enquêtes de victimation. L'un de leurs buts est d'informer sur toutes les infractions inconnues de la police et de la gendarmerie. [Zauberman \(2015\)](#) explique : *"il s'agit d'appliquer la technique des sondages, c'est-à-dire des enquêtes sur échantillon, à un phénomène social qui n'avait traditionnellement été mesuré qu'à travers des données administratives"*. [Robert et al. \(1999\)](#) appuient l'intérêt de ces enquêtes en mettant en évidence de forts écarts entre les conclusions ministérielles et les conclusions d'enquêtes. Ils mettent notamment en avant l'intérêt de ces travaux pour les violences sexuelles en raison des difficultés de qualifications pénales et des évolutions du vocabulaire pénal et du vocabulaire usité par les policiers et les victimes.

Les choix méthodologiques des enquêtes de victimation permettent aussi de combler certaines limites des données judiciaires puisque dans certains cas les questions formulées s'écartent des discours et catégories juridiques. Ce principe veut à la fois permettre à tous et toutes les enquêtées de se reconnaître dans des faits peu importe leur niveau de connaissances et de vocabulaire mais également de réduire l'émotion que pourrait provoquer l'évocation de certains termes. Nous pouvons citer certaines enquêtes telles que l'ACSF de [Bajos and Spira \(1991\)](#), les [Baromètres Santé](#) ou les enquêtes annuelles [Cadre de Vie et Sécurité](#) du Ministère de l'intérieur. Les deux enquêtes majeures qui ont largement amélioré les

connaissances concernant les violences subies par les Françaises et les Français sont : l'*Enquête nationale sur les violences envers les femmes en France* (Enveff) réalisée en 2000 par le Centre de recherche de l'institut démographique de l'Université Paris 1 (Cridup) (Jaspard, 2001) et l'enquête *Violences et rapports de genre* (Virage) réalisée en 2015 par réalisée par l'Institut national d'études démographiques (Ined) (Brown et al., 2020). Cette dernière est la seule enquête de victimation réalisée en France répondant aux standards internationaux édités par l'ONU en matière de mesure des violences à l'encontre des femmes. De plus, elle contextualise précisément les violences, leurs contextes et leurs conséquences. Cette enquête a permis de fournir des statistiques précieuses et représentatives de la population française donnant des clefs de compréhension sur la complexité et la multiplicité de ces violences à l'échelle nationale.

Les enquêtes de victimation ont donc permis des avancées majeures en termes de connaissances de l'étendue du problème des VSS et des vécus des victimes. Les personnes enquêtées dans les enquêtes citées sont principalement contactées et interrogées par téléphone. Ce mode de contact offre de nombreux avantages puisqu'il permet une certaine distance avec la victime qui peut parfois plus facilement se livrer, en particulier concernant les violences sexuelles (Sipes et al., 2022). L'intermédiaire est aussi partiellement gommé ce qui améliore l'introspection et diminue le sentiment de jugement (Lévy-Guillain et al., 2022). Aussi l'anonymat peut être protégé, ce qui aide également les victimes à se déclarer comme telles. Les jugements sociaux et raciaux sont également plus limités par rapport aux entretiens judiciaires en face à face, l'expérience émotionnelle et la rigidité des procédures sont minimisées.

Un autre avantage méthodologique est celui de la possibilité de mettre en avant

la multivictimation dans ces enquêtes ([Debarbieux, 2004](#)) et donc la répétition voire la régularité des violences, éléments difficilement mesurables mais crucial pour saisir la systématisation des violences sexuelles ([Kelly, 1987](#); [Le Goaziou, 2011](#)).

Les enquêtes de victimation présentent cependant certaines limites méthodologiques. Tout d'abord, la fiabilité des données peut être affectée par des problèmes de mémoire et de perception, ce qui rend difficile pour les répondants et répondantes de se souvenir avec précision des événements violents, en particulier ceux qui ont eu lieu dans un passé lointain ([Salmona, 2018](#)). Dans le cas de violences sexuelles, on constate aussi de la sous-déclaration ou de la minimisation des violences subies ([Jaspard, 2011](#)). Ces enquêtes dépendent donc de la volonté et de la capacité des victimes à se déclarer, mais aussi du champ de l'enquête et de l'échantillon choisi ce qui peut entraîner une sous-représentation des cas les plus graves ainsi que des groupes de population marginalisés. Les victimes marginalisées peuvent d'ailleurs représentées les cas les plus graves de cumul des violences sexuelles et sexistes. La marginalisation peut être une cause et une conséquence des événements vécus ([Ingenito, 2019](#); [Salmona, 2022](#)). Il est alors très difficile de les détecter par des outils statistiques.

Les données ministérielles sur les violences sexuelles occupent une place essentielle dans la compréhension et la prise de conscience de l'ampleur de ce problème. Elles sont considérées comme légitimes, sérieuses et représentatives, et jouent un rôle crucial dans la rendue publique de ces violences ([Fouquet*, 2010](#)) et la constitution de la question en problème public ([Dauphin, 2023](#)). Cependant, il est important de reconnaître les limites inhérentes à ces données.

1.1.4 D'autres espaces d'expression investis par les victimes et la recherche

Pour compléter l'étude des données administratives sur les violences sexuelles, il est également essentiel de s'intéresser à d'autres approches possibles. En effet, la nécessité de contextualiser et de saisir la complexité des VSS encourage les scientifiques à aller au-delà des sources administratives et à élargir les champs de recueil de l'information.

Premièrement, nous pouvons citer **l'étude des données issues d'associations**. Ces données associatives, recueillies par le biais des appels téléphoniques et des services d'aide aux victimes, ont une grande valeur dans l'étude des violences sexuelles. Ces données sont produites dans un espace d'accueil de la parole qui ne vise pas la judiciarisation ou une finalité statistique, permettant ainsi de réduire la subjectivité et les cadres préétablis. Les associations offrent un environnement propice où les victimes peuvent s'exprimer librement, favorisant ainsi une compréhension plus approfondie des violences subies.

Un exemple notable de travail dans ce domaine est celui de [Debauche \(2011\)](#), qui a réalisé une étude quantitative à partir des appels passés au numéro anonyme et gratuit SOS Viols Femmes Informations. Cette étude a permis de construire une classification des appels en utilisant des techniques d'analyse de données, regroupant ainsi les appels présentant des similarités dans les violences décrites et les comportements de recours des victimes. Ses résultats révèlent que les violences sexuelles prennent des formes variées en fonction du contexte où elles se produisent. Les violences sexuelles au sein de la famille sont souvent répétées, commençant généralement par des agressions sexuelles et n'impliquant pas toujours

explicitement l'usage de la violence physique. Les appels relatant des viols conjugaux font souvent état de violences répétées, où l'usage de la violence physique est fréquent. Les autres types de viols sont moins souvent répétés. En somme, l'analyse des données associatives offre un complément essentiel aux données officielles en permettant aussi de saisir les effets de la médiatisation sur les appels et en confrontant les résultats avec les enquêtes menées par les autorités. Cela contribue à une meilleure compréhension des différentes dimensions des violences sexuelles et des besoins spécifiques des victimes.

Les entretiens menés en psychologie, en histoire ou en sciences sociales sont également précieux. Ces approches qualitatives permettent d'explorer en profondeur les histoires individuelles des victimes, en mettant l'accent sur les contextes, les motivations et les conséquences des violences sexuelles. Les entretiens offrent un espace d'expression plus libre et confidentiel, favorisant ainsi la révélation de témoignages plus personnels et détaillés. Finalement, ces données associatives ne visent pas à être représentatives, car elles dépendent de l'initiative des victimes d'appeler. D'ailleurs, ces archives ne sont pas produites dans le but d'être étudiées scientifiquement. Cependant, elles offrent un aperçu des témoignages plus authentiques et complets, permettant aussi d'étudier la dynamique et les motivations entourant la prise de parole.

Dans cette catégorie, il est essentiel de s'attarder sur le travail de Liz Kelly qui a fourni une analyse des vécus de violences sexuelles dès les années 1980 (Kelly, 1987). Son article sur le *continuum* des violences s'appuie ainsi sur des entretiens réalisés avec soixante femmes. Ces femmes ont volontairement choisi de discuter en détail de leurs expériences. Cet échantillon est donc évidemment non représentatif

d'une quelconque population mais de tels entretiens qui ont été longs et réitérés auraient été très pénibles pour une population non avertie et volontaire. L'objectif de cette recherche était tout de même de donner la parole à un éventail plus large de femmes que celles généralement contactées par les instances officielles ou les organisations bénévoles et associations. Un effort a donc été fait pour s'entretenir avec des femmes et profil de femmes qui n'étaient pas ou peu présents dans les statistiques officielles et associatives.

Le concept de *continuum* peut se référer à "une série continue d'éléments ou d'événements qui se confondent et ne peuvent être distingués d'emblée" (définition de l'Oxford English Dictionary). Cette définition reconnaît qu'il n'existe pas de catégories d'analyse clairement définies et délimitées pour rendre compte du vécu des femmes en matière de violence, s'écartant des approches judiciaires et administratives. Il est important d'intégrer ce concept fondamental dans l'étude des VSS et notamment dans l'analyse des discours directs de victimes qui ne se réfèrent pas toujours à une catégorie juridique précise ou qui évoquent une multiplicité de faits.

Le travail de Liz Kelly met donc en avant un avantage des entretiens en convoquant une population précise qui répond un échantillonnage prédéfini par le cadre de recherche. Cette population peut bien sûr exclure certaines victimes puisque dans le cas de Liz Kelly il ne s'agit que de femmes. Les entretiens offrent une perspective complémentaire aux autres sources de données évoquées, telles que les statistiques ministérielles, les enquêtes de victimation et les données associatives. Alors que ces dernières fournissent des informations sur la prévalence des violences

sexuelles et sur les schémas systémiques en étudiant un plus large panel de victimes, les entretiens permettent d'explorer la réalité complexe et subjective des expériences vécues par un plus petit nombre. Ils offrent également un espace pour que les victimes puissent exprimer leur vécu de manière détaillée et contextualisée, en dehors des contraintes juridiques ou statistiques. Plusieurs recherches ont été menées pour étudier des population spécifiques dont voici quelques exemples : des hommes victimes d'inceste ([Wicky, 2020](#)), des détenus et détenues ([Ricordeau*, 2004](#)), des étudiantes ([Savoie et al., 2018](#)). Ces travaux continuent d'éclairer les différents vécus de victimes que ce soit en mettant en lumière des réalités ignorées ou en contextualisant des chiffres déjà connus de la société.

Finalement, il est intéressant de citer des **approches exploratoires analysant différents supports médiatiques**. Des recherches telles que celles de [Lochon \(2021\)](#) et [Ruffio \(2020\)](#) montrent que la presse écrite française a souvent tendance à traiter les VSS sous l'angle des faits divers sensationnels ou du discours sécuritaire. Ce constat souligne le problème de la subjectivité des rédactions et des journalistes qui interviennent comme intermédiaires dans la transmission de l'information. L'analyse de la couverture médiatique, bien qu'elle repose sur un volume substantiel de données, semble donc être plus utile pour étudier les discours et opinions publics, plutôt que de refléter fidèlement la réalité des VSS. Par ailleurs, un autre type de traitement médiatique des VSS peut être mentionné. Il s'agit des émissions et tribunes testimoniales qui permettent aux victimes de s'exprimer directement. [Beccarelli \(2021\)](#) souligne le rôle des radios nocturnes, qui, grâce à leur ambiance intimiste et anonyme, permettent aux victimes de partager leurs expériences. [Salmon \(2017\)](#) mentionne, elle, des courriers des lecteurs

concernant le divorce, qui abordent souvent les scènes de violences conjugales. Ainsi, ces recherches dans les archives permettent de mettre en évidence l'existence et l'accessibilité à des témoignages de victimes sur des médias publics, même avant #MeToo. L'émission télévisée "Ça commence aujourd'hui", qui donne la parole à des victimes de VSS sous divers aspects, tels que l'inceste, le harcèlement, ou les violences subies par les hommes n'a jamais été étudiée mais peut être citée en tant que média testimonial très populaire. Ces médias offrent un accès plus brute aux expériences des victimes, dans toute leur émotion et leur subjectivité.

L'analyse des œuvres littéraires offre également une perspective personnelle unique. Les autobiographies, en particulier, fournissent des témoignages détaillés sur le sujet. Parmi ces œuvres, on peut citer [Thomas \(1986\)](#), [Flament \(2016\)](#), [Springora \(2020\)](#), ou [Kouchner \(2021\)](#). L'étude de [Vernier \(2022\)](#) sur "La petite fille sur la banquise" d'Adélaïde Bon révèle comment cette œuvre dépasse le simple récit autobiographique pour devenir un appel à la prise de conscience collective et un guide pour les victimes. Au-delà de partager son vécu, Bon conseille d'autres victimes de violences sexuelles en détaillant sa lutte et son parcours à travers le système judiciaire. Son récit illustre sa résilience et donne des informations précieuses pour surmonter un traumatisme souvent méconnu. Ces livres autobiographiques sont essentiels en tant que données qualitatives car ils offrent un aperçu profond des expériences et des défis auxquels sont confrontées les victimes de VSS. Ils aident également à sensibiliser et à éduquer le public sur la gravité et l'impact de ces violences. En plus des récits autobiographiques, des œuvres de fiction abordent également les VSS, avec une approche souvent plus audacieuse et critique. L'écriture de Virginie Despentes est un exemple notable.

Comme le souligne Virginie Sauzon, Despentès aborde la violence sans la minimiser et met en évidence la complexité du viol. Son œuvre confère aux personnages féminins une force et une agressivité qui défient les stéréotypes de genre. Elle utilise l'espace fictionnel comme un terrain de démonstration et de création politique pour questionner la domination masculine et encourager l'empathie. Selon Sauzon, l'art de Despentès sert de miroir performant où les lecteurs peuvent trouver la capacité d'action et de réaction. De plus, la reconnaissance de Despentès dans le milieu littéraire et la médiatisation de son travail permettent à sa voix de toucher un public plus large, au-delà des cercles féministes. En conclusion, les œuvres littéraires, qu'elles soient autobiographiques ou fictionnelles, constituent une source précieuse de données qualitatives sur les VSS. Elles permettent de mieux comprendre les expériences des victimes, de sensibiliser le public et d'offrir des perspectives critiques sur les enjeux sociétaux liés aux violences sexistes et sexuelles. Les œuvres littéraires se révèlent être des supports précieux pour l'étude qualitative des violences sexistes et sexuelles. Les recherches qui utilisent ces œuvres comme matériel de base mettent en lumière l'importance d'analyser les mots des victimes eux-mêmes.

Chapitre 2 : Le mouvement #MeToo et l'espace Twitter

1.2.1 Émergence et évolution du mouvement #MeToo

Tweet 1

Sandra Muller - 13/10/2017 : *"#balancetonporc!! toi aussi raconte en donnant le nom et les détails un harcèlement sexuel que tu as connu dans ton boulot. Je vous attends"*

Tweet 2

Alyssa Milano - 15/10/2017 : *"If you've been sexually harassed or assaulted write 'me too' as a reply to this tweet."*

L'expression #Metoo a émergé initialement en 2006 sous l'impulsion de Tarana Burke, une militante américaine des droits civiques, qui a utilisé l'expression "Me too" pour sensibiliser et soutenir les jeunes filles noires victimes de violences sexuelles. C'est en octobre 2017, à la suite des révélations concernant Harvey Weinstein et d'autres personnalités de l'industrie du divertissement, qu'Alyssa Milano a relancé l'expression sur les réseaux sociaux, en appelant les victimes de violences sexuelles à répondre "me too" à son tweet, afin de démontrer l'ampleur de ces violences (cf Tweet 2). Ce message ainsi que l'affaire Weinstein ont rapidement une ampleur mondiale. En France, le mouvement été porté par l'apparition du hashtag #Balancetonporc, créé par Sandra Muller le 13 octobre 2017 (cf Tweet 1). Ce hashtag appelait les victimes à dénoncer publiquement les violences, et a ouvert un espace spécifique d'expression et de dénonciation dans le contexte français. Le mouvement #Metoo a ainsi trouvé un écho particulier en France, où il a été complété par de nombreux autres hashtags plus spécifiques, permettant de mettre en lumière des formes variées de violences sexuelles et sexistes (VSS) dans différents contextes, tels que Metooinceste ou Metoogay.

Les réseaux sociaux, avec en premier lieu Twitter, ont été des plateformes clés pour l'émergence et le développement du mouvement. Les caractéristiques inhérentes des réseaux sociaux ont facilité le partage de témoignages, en offrant un espace où les personnes pouvaient trouver soutien, visibilité, et, si nécessaire, anonymat. Ce mouvement a donc initié des vagues de témoignages et a aidé à transformer cette série de récits individuels en une mobilisation et une prise de

1 PARTIE 1 : EMERGENCE DES VIOLENCES SEXISTES ET SEXUELLES
COMME QUESTION POLITIQUE ET SCIENTIFIQUE

Hashtag	Date d'apparition	Thème
#Metoo	Octobre 2017	Témoignages d'agressions sexuelles et de harcèlement
#BalanceTonPorc	Octobre 2017	Encouragement à dénoncer publiquement les agresseurs en France
#MetooInceste	Janvier 2021	Témoignages d'inceste
#MetooGay	Janvier 2021	Témoignages des violences sexuelles dans la communauté gay
#Iwas	Juin 2021	Témoignages des VSS chez les mineurs - Encouragement à donner son âge

TABLE 1 – Résumé des dates d'apparition et des thèmes des hashtags étudiés

conscience collective. Cela illustre le concept de *hashtag feminism* mentionné par [Clark-Parsons \(2021\)](#), et souligne l'importance des hashtags pour organiser et donner de la visibilité aux causes féministes sur les réseaux sociaux. En s'appuyant particulièrement sur la répétition et la visibilité des témoignages, ce mouvement a donc créé un récit unifié mettant en lumière des problématiques communes.

L'émergence du mouvement #Metoo s'explique aussi très bien par l'articulation entre les plateformes numériques et les médias traditionnels, et comment cette synergie a contribué à alimenter une nouvelle vague de féminisme. [Da Silva \(2022\)](#) a souligné la puissance des plateformes numériques en termes de flexibilité et de perméabilité, permettant une fluidité des témoignages et une articulation entre les espaces en ligne et hors ligne. Les médias traditionnels ont joué un rôle de catalyseur en donnant de l'écho aux affaires de grande ampleur, tandis que les médias numériques ont fourni une plateforme pour (presque) toutes les voix. Cette combinaison a contribué à donner de la force aux récits de toutes et tous et a accru la mobilisation collective. Par ailleurs, [Akrich and Méadel \(2007\)](#) ont identifié des niveaux d'action collective qui trouvent un écho dans le mouvement #Metoo.

Le premier niveau concerne les actions individuelles, comme les témoignages en ligne. Le deuxième niveau voit l'agrégation de ces témoignages en une voix collective qui transcende les plateformes numériques et trouve écho dans les médias traditionnels. Enfin, le troisième niveau consiste en la structuration de ces voix en organisations plus formelles et en actions concrètes, souvent hors ligne, telles que des manifestations ou des actions en justice. [Jouët et al. \(2017\)](#) examinent comment le féminisme en ligne utilise ces niveaux de manière novatrice, combinant tactiques en ligne et hors ligne. Les outils numériques, tels que les médias sociaux, les e-mails et les pétitions en ligne, ont été utilisés de concert avec des manifestations et des plaidoyers hors ligne pour créer un militantisme féministe renouvelé. Ce mélange de tactiques est emblématique de ce qui est parfois appelé la quatrième vague du féminisme, qui est marquée par l'utilisation de la technologie et des médias sociaux pour contester les inégalités de genre et faire progresser l'égalité ([Bertrand, 2018](#)). En résumé, l'émergence de #Metoo a été fortement influencée par l'interaction entre les médias numériques et traditionnels, et a démontré la puissance de l'action collective à plusieurs niveaux ([Akrich and Méadel, 2007](#)). Cela s'inscrit dans une tendance plus large de féminisme en ligne qui combine habilement les actions en ligne et hors ligne, pouvant symboliser une nouvelle vague du mouvement féministe ([Bertrand, 2018](#); [Jouët et al., 2017](#)).

1.2.2 Le corpus numérique de #Metoo ou la constitution d'archives inédites

L'importance des archives dans le féminisme est mise en évidence dans différents travaux scientifiques. [Charpenel \(2014\)](#) dans sa thèse « *Le privé est politique!* » *Sociologie des mémoires féministes en France*, explore comment les militantes

féministes en France évoquent le passé, en s'appuyant sur la sociologie de la mémoire de Maurice Halbwachs et la sociologie de l'action collective. Elle constate que malgré la diversité des voix au sein du mouvement féministe, il existe un consensus sur la nécessité de visibiliser les femmes dans l'histoire, ce qu'elle qualifie de « devoir de mémoire féministe ». Elle met également l'accent sur l'importance des récits personnels dans la formation de la mémoire collective, ce qui trouve un écho dans les témoignages numériques sur Twitter et autres plateformes de médias sociaux. Christine Bard, dans l'ouvrage *Les féministes et leurs archives* (Bard et al., 2023), et en tant que présidente de "Archives du féminisme"¹, met en avant l'urgence de préserver les sources de l'histoire des féminismes. Elle affirme que les archives sont essentielles pour coordonner les efforts visant à préserver la mémoire des associations et militantes féministes. Fraisse (2008) et Henneron (2005) rejoignent cette idée en soulignant que le féminisme est souvent perçu comme une action collective sans mémoire des luttes passées et affirment que la question de la transmission et de la mémoire, notamment via des traces écrites, est essentielle. Cette amnésie apparente souligne l'importance de conserver et d'analyser des archives solides.

*[L]es féministes s'approprient des moyens d'expression dont elles étaient exclues.
Ces écrits forment la mémoire de ce mouvement.*

Liane Henneron, 2005

Cette idée d'appropriation de l'espace et des moyens semble particulièrement appropriée dans l'étude de l'émergence de #MeToo et fait aussi écho au travail de Perrot (2020) qui montre l'exclusion historique des femmes et de leurs expériences

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

dans les archives. Le corpus numérique produit grâce à l'utilisation de Twitter par les victimes semble dès lors non seulement être une ressource parmi d'autres, mais bien une source inédite et importante pour la mémoire des victimes de violences sexuelles et des mouvements féministes. Ces traces numériques, notamment sur les réseaux sociaux, constituent d'ailleurs une source précieuse de données aussi en raison de leur volume et de leur richesse. Par ailleurs, bien que le principal objet d'étude de ce mémoire soit les témoignages de violences, l'ensemble des tweets associés au mouvement *#MeToo* constitue des traces féministes diversifiées, incluant des opinions, des messages de soutien, et autres réactions. Ces tweets, dans leur ensemble, participent à la construction et à la préservation de la mémoire féministe, en créant un corpus numérique riche qui reflète les différentes facettes et dimensions de la cause féministe. En outre, [Halbwachs \(1994\)](#) souligne l'utilité des témoignages et décrit comment les souvenirs, même personnels, sont liés à un ensemble de notions qui sont partagées collectivement. Evidemment, le concept de *mémoire collective* éclaire particulièrement le sujet de ce mémoire, interrogeant à la fois le corpus numérique composé de chaque tweet produit individuellement mais s'insérant dans un mouvement collectif plus vaste. [Pollak and Heinich \(1986\)](#) soulignent de plus la rareté des témoignages produits spontanément sans sollicitation officielle, valorisant ainsi les marqueurs de spontanéité et d'oralité dans les témoignages. [Martin \(1996\)](#) déplore, dans une approche historique, la déformation des mots des victimes pour qu'ils correspondent aux normes judiciaires et aux points de vue de l'intermédiaire, ce qui efface la complexité des témoignages, notamment les liens interpersonnels. Ces travaux insistent donc sur le caractère précieux de témoignages bruts, sans intermédiaires et sauvegardés tels qu'ils ont été produits. Enfin, un exemple emblématique de la valeur des archives person-

nelles dans le féminisme est représenté par les lettres adressées à Simone de Beauvoir, qui sont aujourd'hui conservées à la Bibliothèque nationale de France. [Rouch \(2017\)](#) étudie ces lettres, les décrit comme des « Mémoires féminines « ordinaires » » et souligne le caractère précieux de ces documents. Cette conservation de voix *ordinaires* dans les lettres adressées à Simone de Beauvoir peut être comparée aux multiples voix *ordinaires* qui s'expriment sur Twitter, soulignant l'importance de préserver et de valoriser ce corpus numérique.

Partie 2 : Méthodologie de collecte et d'analyse des données

Chapitre 3 : Les données de l'étude

2.1.1 Comprendre la structure de Twitter et de ses contenus

Le réseau social Twitter permet de mettre en relation ses utilisateurs par le contenu qu'ils produisent et une structure algorithmique de mise en avant ([Cardon, 2018](#)), elle-même influencée par l'interaction volontaire des utilisateurs avec certains contenus. Les usagers de la plateforme possèdent un ou plusieurs comptes chacun associé à un identifiant. Grâce à cela, ils peuvent décider de publier du contenu (texte, emojis, GIF, hyperliens, etc) où une publication correspond à un *tweet* qui est limité à 280 caractères depuis novembre 2017 (l'ancienne limite était de 140 caractères). Les usagers peuvent également s'abonner aux comptes d'autres usagers afin d'avoir un accès facilité à leurs contenus. Cela crée des relations potentiellement unilatérales. Une personne peut suivre un compte sans que l'inverse soit vrai. Cependant, il n'est pas nécessaire d'être abonné à un compte pour en voir son contenu. Cela permet à certains contenus de se propager faci-

lement sur la plateforme et atteindre des usagers non abonnés. Les utilisateurs peuvent aussi faire des recherches par mots clefs ou pseudos et identifiants par lesquels les usagers se reconnaissent. Le pseudo est modifiable quand l'identifiant ne l'est pas. Les mots clefs peuvent être mis en avant par l'utilisation de *hashtag*. Ainsi un mot peut être précédé d'un # pour devenir un mot clef désignant soit un mot important du contenu, soit un mot souvent repris et affilié à un mouvement. Evidemment, #Metoo est bien sûr un exemple. De plus, les hashtags sont cliquables ce qui facilite l'accès à tous les tweets contenant ce hashtag. Finalement, les usagers peuvent interagir grâce à différents outils. Que les personnes se suivent ou non, elles peuvent aimer (*liker*) un contenu d'autrui ou le partager (*retweet*) sur son propre compte. L'interaction peut aussi être discursive puisque en plus de repartager un contenu, l'utilisateur peut y inclure un contenu personnel. On parle de citation dans ce cas. Sinon, les utilisateurs peuvent interagir directement en répondant simplement au contenu posté. Cet outil s'apparente aux *commentaires* des autres plateformes.

2.1.2 L'utilisation de cette structure dans le mouvement #Metoo

La taille limitée des publications sur Twitter contraint certes le nombre d'éléments qui peuvent être précisés dans un récit. Mais ce cadre autorise aussi une concision du discours, voire une énumération sans transition d'éléments plus ou moins violents. En effet, puisque #Metoo est vite devenu visible et intelligible en France et notamment sur Twitter, les utilisateurs de la plateforme et du hashtag n'ont pas à préciser son bagage sémantique (Husson, 2016). Lorsqu'il s'agit d'un témoignage et non d'un tweet d'opinion, le hashtag transmet à lui seul la violence qu'il est difficile d'énoncer. Dès lors, les narratrices et narrateurs peuvent plus facilement

détailler ou s'exprimer sur cette violence, avec les normes habituels de Twitter (Paveau, 2013) : phrases nominales, mots clefs, autres hashtags (résumant une émotion ou un contexte), emojis, liens url, etc.

Trovato (2022) étudie les discours du mouvement #MeToo sur Twitter et identifie deux tendances dans les témoignages : l'évitement du dire et l'exposition détaillée. L'évitement comprend l'utilisation du hashtag comme porte-parole, la non-qualification des violences, l'usage d'euphémismes, et l'anonymisation du narrateur, un aspect qui n'est pas explicitement mentionné par Trovato mais que Cavalin et al. (2022) note comme courant dans #MeToo pour permettre l'expression tout en se protégeant. En outre, le hashtag #Balancetonporc encouragerait spécifiquement la dénonciation des agresseurs, et Lopez et al. (2019) observe que cela conduirait à des témoignages français plus accusateurs, violents, émotionnels et précis comparé aux témoignages américains. Il sera donc intéressant de s'intéresser à cette question et d'élargir l'étude à d'autres hashtags.

2.1.3 Deux formes de collecte : la base INA et l'API Twitter

Dans le cadre de ce mémoire, une base de données de tweets riche et suffisamment exhaustive a été obtenue en fusionnant deux collections de tweets distinctes. La première collection provient d'une base de données de l'Institut national de l'audiovisuel (INA), qui comprend des tweets collectés quotidiennement entre 2017 et le 7 mars 2022. Cette base est extrêmement précieuse car elle contient des tweets qui ne sont plus accessibles aujourd'hui en raison de suppressions de comptes ou de tweets. En effet, le caractère journalier de la collecte permet une restitution presque parfaite du contenu au jour le jour sur la période étudiée.

La deuxième collection a été obtenue via l'API de Twitter en utilisant un

compte académique. Cet outil permettait, jusqu'en février 2023, aux équipes de recherche en faisant la demande d'accéder à un permis de collecte de tweets. Cette permission autorisait la récupération d'un volume limité mais suffisant de contenu selon des filtres fixés par l'équipe de recherche. Cet outil a permis de récupérer des tweets entre 2017 et décembre 2022. Cette collecte correspond à l'état des lieux de Twitter à la date de la collecte. Ainsi, si des tweets ou comptes ont été supprimés entre temps, ils ne sont pas présents dans cette seconde base.

Finalement, la base de données fusionnée contient 335 612 tweets uniques (sans doublons entre les bases) en langue française et sans retweets. Tous les tweets collectés mentionnent les hashtags suivants : #Metoo, #BalanceTonPorc, #Iwas, #MetooInceste et #MetooGay (voir Table 3). Parmi ces tweets, 61 609 ont été potentiellement supprimés (soit les tweets eux-mêmes, soit les comptes des utilisateurs, par ces derniers ou par la plateforme). Ces tweets sont en effet présents dans la collecte en direct de l'INA mais ne se trouvaient pas dans la collecte via l'API fin décembre 2022. La table 2 rappelle les effectifs des bases. Par ailleurs, comme mentionné dans la sous-partie 2.1.1, les tweets publiés génèrent un certain nombre de métadonnées (données secondaires fournissant de l'information supplémentaire sur la donnée initiale, ici le contenu du tweet). Parmi ces métadonnées, on dispose de deux catégories d'information : celles sur le tweet et celle sur le compte d'où provient le tweet. Dans la première catégorie, on peut citer le nombre de likes, de retweet ou la date de publication. Concernant le compte, les informations suivantes sont disponibles : la date de création du compte, le nombre de comptes abonnés, le nombre d'abonnements, le nombre de tweets publiés, le pseudo et l'identifiant. D'autres métadonnées sont disponibles via les collectes issues de Twitter mais n'ont pas été récupérées ou utilisées dans ce travail.

	API Twitter	INA	Total
Nombre total (sans RT)	273 913	146 147	420 060
Nombre de doublons avec l'autre base	84 439	84 439	-
Nombre de tweets uniques dans cette base	189 474	61 699	335 612

TABLE 2 – Description des bases de données de tweets

En outre, il convient de souligner l'intérêt de cette base de données, en particulier en raison des circonstances de sa collecte. En février 2023, la nouvelle direction de Twitter par Elon Musk a mis en œuvre des règles de collecte de données plus strictes et a modifié les règles d'utilisation de la plateforme. Ces nouvelles règles limitent significativement l'accès aux données. De plus, les nouvelles règles imposées par Elon Musk pourraient aussi potentiellement modifier la manière dont les utilisateurs interagissent avec la plateforme, que ce soit par la suppression de comptes, une prise de distance avec la plateforme ou des modifications de l'algorithme de recommandation de Twitter (Le Querrec, 2023). Ces évolutions pourraient avoir des conséquences importantes sur les types de contenu partagés et les comportements d'interaction observés sur la plateforme. Ainsi, une collecte antérieure à ce changement évite de se confronter à une modification potentiellement significative du contenu et des pratiques sur la plateforme.

Un autre aspect essentiel qui renforce la valeur de la base de données étudiée est le recul de cinq ans depuis le lancement du mouvement #MeToo. Cette période de cinq ans permet d'avoir accès à une masse de données substantielle et offre un panorama intéressant des différentes dynamiques du mouvement. Avec une telle quantité de données, il est possible d'observer des tendances, des évolutions et des schémas récurrents, ce qui constitue une ressource inestimable pour ce travail de mémoire.

Hashtag	Date d'apparition	Thème
#Metoo	Octobre 2017	Témoignages d'agressions sexuelles et de harcèlement
#BalanceTonPorc	Octobre 2017	Encouragement à dénoncer publiquement les agresseurs en France
#MetooInceste	Janvier 2021	Témoignages d'inceste
#MetooGay	Janvier 2021	Témoignages des violences sexuelles dans la communauté gay
#Iwas	Juin 2021	Témoignages des VSS chez les mineurs - Encouragement à donner son âge

TABLE 3 – Résumé des dates d'apparition et des thèmes des hashtags étudiés

Chapitre 4 : La méthodologie

L'analyse des données textuelles présente des défis distincts par rapport à celle des données numériques. Alors que les données numériques, telles que les valeurs quantitatives ou les statistiques, peuvent être traitées à l'aide d'opérations mathématiques simples, les données textuelles requièrent une approche plus complexe. En effet, le texte est un support riche, non structuré, qui véhicule des significations contextuelles, des nuances, et des relations implicites entre les mots. Cette nature implique que les outils pour traiter les données textuelles doivent être capables d'analyser la sémantique, la syntaxe, et parfois même le pragmatique du langage.

En conséquence, des méthodes spécifiques issues du traitement automatique du langage naturel (TAL/NLP) sont nécessaires pour structurer et interpréter le contenu textuel. Ces méthodes permettent d'identifier des tendances cachées, de classer les textes selon leur contenu, et de découvrir des modèles à partir de grands corpus de données. Dans le cadre de cette étude, plusieurs approches ont été mobilisées pour analyser à la fois le texte des tweets et leurs métadonnées, afin

d'extraire un maximum d'informations sur les témoignages relatifs aux violences sexuelles et sexistes partagés dans le mouvement #MeToo. Quelques travaux ont étudié le contenu textuel des tweets de #Metoo via une comparaison internationale (Lopez et al., 2019), l'étude du contenu anglophone (Mendes et al., 2018) ou l'étude du contenu francophone (Méadel (2022); Trovato (2022)). Cependant, la courte période parfois étudiée ou la circonscription autour de quelques tweets limitent les conclusions de ces travaux. Aussi, la confusion est parfois faite entre le nombre de tweets mentionnant les hashtags du mouvement et la part effective de tweets relatant un témoignage, ce qui rend vastidieux le travail de lecture et d'analyse de ces données très diverses.

2.2.1 La classification automatique de données textuelles

Identifier les différents contenus relayés

Dans ce travail de mémoire, une part cruciale du travail a consisté à identifier et classer les tweets contenant des témoignages de violences sexuelles et sexistes (VSS), parmi d'autres types de contenus comme des opinions ou des articles de médias. En effet, il semble essentiel de comprendre de quoi est composé exactement le contenu associé aux hashtags du mouvement #Metoo en France. Est-ce, comme la presse l'a souvent décrit (exemple dans l'encadré 2.2.1), un mouvement uniquement destiné à la dénonciation ?

Cela fait six mois que le hashtag #BalanceTonPorc, émule français du planétaire #MeToo, est apparu sur les réseaux sociaux à la suite de l'affaire Harvey Weinstein. Depuis, près de 850 000 messages ont été postés sur les différentes plateformes, dénonçant de façon quasi-cathartique toutes les violences faites aux femmes.

Pierre Bafoil dans *Les Inrockuptibles*, 2018

Pour identifier automatiquement les différents types de contenu, j'ai recours à des modèles, dits, de classification. L'objectif est de choisir le nombre de classes à identifier, de donner des exemples du contenu associé à chacune des classes, puis d'apprendre au modèle à la faire automatiquement.

Le modèle utilisé pour cette classification est CamemBERT (Martin et al., 2019), une variante de BERT spécifiquement entraînée sur un large corpus de textes en français. BERT (Bidirectional Encoder Representations from Transformers) est une avancée majeure dans le domaine du traitement du langage naturel (Liu et al., 2019)(Devlin et al., 2018). Au cœur de BERT se trouve l'architecture dite des *Transformers* (Vaswani et al., 2017). Ces derniers ont révolutionné le domaine de l'apprentissage automatique en introduisant une approche novatrice pour modéliser très finement les structures qui lient entre eux les mots d'un texte. Cette capacité à comprendre le contexte global d'un mot permet d'améliorer significativement la qualité des représentations sémantiques produites par le modèle, et donc ses performances sur diverses tâches de traitement du langage naturel. CamemBERT est une adaptation de BERT pour la langue française. Il a été entraîné sur un vaste corpus de textes en français, incluant Wikipédia et des textes issus du web. Cet entraînement lui a permis de saisir la structure spécifique de la langue française et de produire des représentations sémantiques adaptées.

Dans notre cas, l'avantage a été tiré de l'apprentissage préalable de CamemBERT en ajoutant à sa structure des couches supplémentaires spécifiquement entraînées pour notre tâche, à savoir la classification des tweets en quatre catégories, et sur les données textuelles spécifique à Twitter. Ainsi, le modèle que j'ai entraîné

bénéficie à la fois des représentations sémantiques de haute qualité inhérentes à CamemBERT, et d'un apprentissage spécifique effectué à partir des données que j'ai fournies et de la tâche que j'ai entraînée. Le but étant ici d'identifier parmi les plus de 300 000 tweets à disposition, les 4 classes suivantes : témoignages, opinions, médias, autres.

Tweet 3

Exemple de contenu médiatique : *Harcèlement sexuel : avec le #balancetonporc, des centaines de femmes témoignent <https://t.co/gUHJLbc2A3> via @franceinfo*

Tweet 4

Exemple de contenu d'opinion : *Les #Iwas c'est choquant...force aux femmes qui ont du subir ça ..*

Tweet 5

Exemple de contenu d'opinion : *Entre #metoo, l'appropriation culturelle, et les études post-coloniales, « Un Indien dans la ville » n'aurait jamais vu le jour en 2018 #UnIndienDansLaVille*

Tweet 6

Exemple de tweet hors contexte : *"#joyeusespaques #premieravril #jeremstargates #balancetonporc #sondage vous faites quoi en premiers quand vous avez un kider surprise(follow,je fais des sondages)"*

Concrètement, l'utilisation de ces modèles pour la classification repose sur plusieurs étapes clés (Figure1) : l'annotation manuelle d'une partie des données, l'entraînement du modèle, et les phases de test et de validation. Une fois que

le modèle produit les meilleures performances possibles, il est possible d'annoter automatiquement tout le reste du corpus.

1. Sélection d'un échantillon représentatif de tweets :

Un sous-échantillon de tweets a été extrait de la base complète pour être annoté manuellement. Cet échantillon inclut une diversité de contenus (témoignages, opinions, articles de presse, commentaires non pertinents, etc.), afin de garantir que le modèle puisse généraliser correctement à l'ensemble des données. Pour ce faire, j'ai tiré aléatoirement des tweets par lots de 100. Puis j'ai répété l'opération 20 fois. Le tirage aléatoire permet d'assurer la représentativité de l'échantillon.

2. Annotation manuelle des tweets :

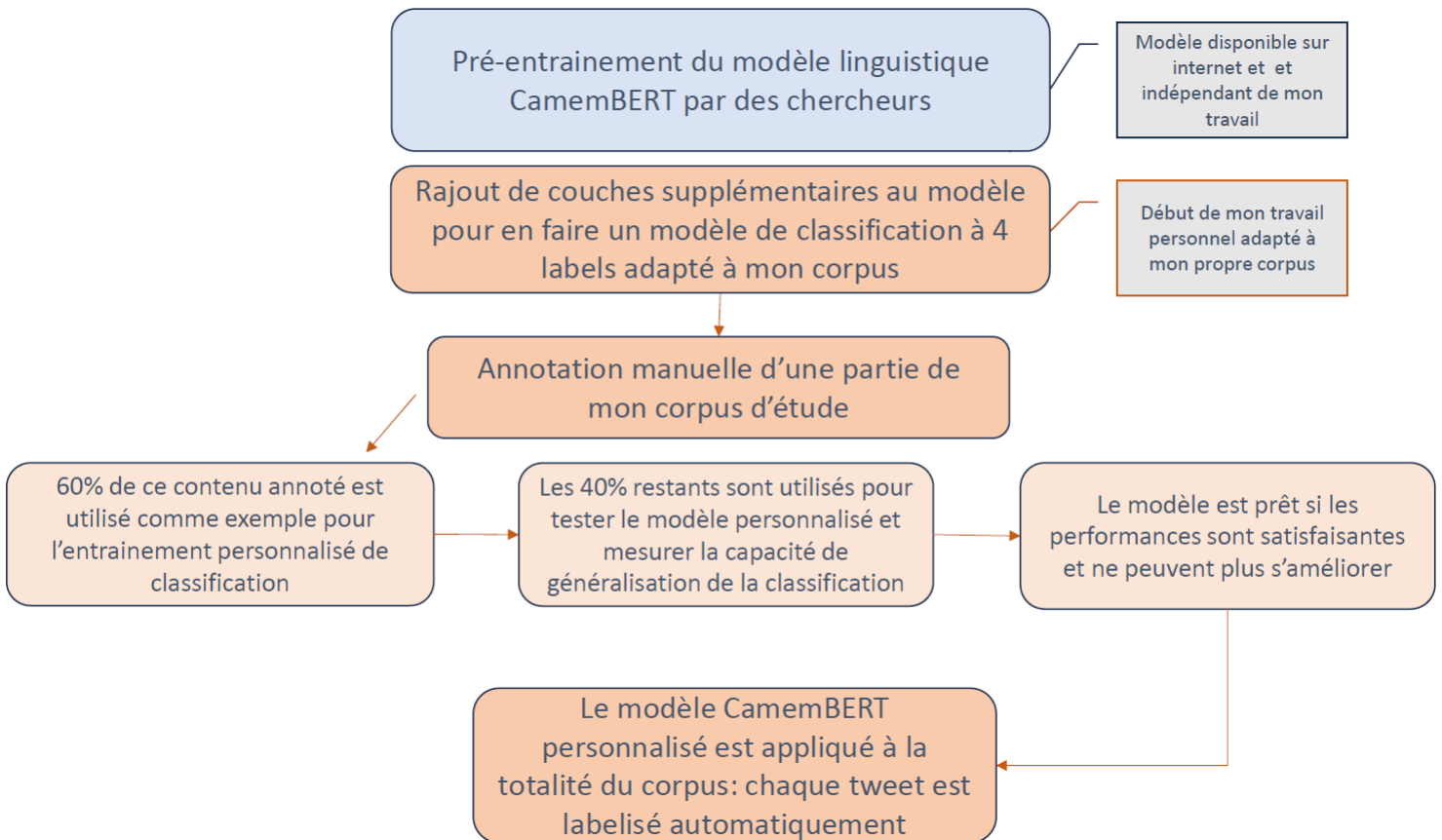
Chaque tweet de l'échantillon a été classé manuellement dans une des catégories suivantes : **témoignage** (contenu où une victime rapporte une violence), **opinion** (où un utilisateur donne son avis sur le mouvement, du plus sérieux au plus ironique), **média** (tweet renvoyant à un article de presse, une vidéo Youtube, un podcast, ...), ou **autre** (tweet hors sujet ou spam). Après un travail d'annotation manuelle de plus de 6 heures, l'échantillon présente la distribution suivante : "opinion" (1331 tweets), "médias" (412 tweets), "témoignage" (182 tweets), et "autre" (81 tweets). Le travail fastidieux que cela représente explique aussi la nécessité d'automatiser la tâche.

3. Entraînement du modèle :

Une fois une partie des données annotées, elles sont utilisées pour entraîner le modèle CamemBERT. Le modèle de base est déjà pré-entraîné sur de vastes corpus de langue française mais il doit ensuite s'entraîner sur le corpus

spécifique de l'étude pour qu'il puisse classer les tweets selon les catégories d'intérêt. Pour ce faire, l'échantillon annoté est divisé en deux ensembles : l'ensemble **d'entraînement** (ici 60 % des données, de l'échantillon annoté) et l'ensemble **de test** (40 %). L'ensemble d'entraînement est utilisé pour ajuster les paramètres du modèle, tandis que l'ensemble de test est utilisé pour évaluer ses performances sur des données qu'il n'a pas vues pendant l'entraînement. Lorsque le modèle obtient des performances satisfaisantes (ie, il prédit suffisamment de fois les bonnes classes de contenu), il peut-être généralisé à tout le reste du corpus de tweets.

FIGURE 1 – Structure schématique de la méthodologie utilisée pour la classification automatique



Identifier les contextes de violence décrits parmi les témoignages

En plus de la classification des types de tweets, une seconde phase d'analyse a été menée pour identifier, parmi les témoignages, le contexte des violences évoquées. Cette analyse vise à mieux comprendre dans quels environnements et à quelles circonstances ces violences ont eu lieu (par exemple, au sein de la famille, dans le cadre professionnel, etc.). En effet, lors de ma lecture des tweets lors des précédentes phases d'annotation, j'ai notamment remarqué la récurrence de la mention du contexte du témoignage, plus systématiquement précisé que la mention du type de violence (agression sexuelle, injure, ...) par exemple.

Pour cette tâche, j'ai une nouvelle fois utilisé un modèle basé sur CamemBERT, entraîné cette fois-ci à repérer des informations spécifiques sur le contexte des violences dans les tweets précédemment classés comme témoignages. Les tweets de témoignages sont cette fois analysés pour identifier les mentions explicites ou implicites de contexte (ex. : « dans ma famille », « au travail », « par mon patron », etc.), en combinant l'annotation d'un échantillon avec un modèle de classification automatique utilisant CamemBERT. Cependant, j'ai ajouté une dimension d'automatisation à cette étape : une annotation automatique basée sur des expressions et vocabulaires sans ambiguïté, tels que "à l'école", "au travail", "mon ex-mari/copain" ou "c'était mon père". Grâce à la nature concise des témoignages que j'avais déjà isolés, cette méthode d'annotation automatisée s'est avérée extrêmement efficace. Or, augmenter le volume de la base d'entraînement est extrêmement intéressant pour produire un modèle plus performant.

- 0 : **Autre** : Pour tous les témoignages qui n'entrent pas clairement dans l'une des catégories pré-définies ou qui n'évoquent pas les circonstances de l'évènement.

- 1 : **Connaissances** : Cela englobe les agressions perpétrées par des individus connus de la victime, sans pour autant avoir un lien familial, professionnel ou amoureux. Exemples : voisins, colocataires, amis,...
- 2 : **Famille** : Cette catégorie englobe les agressions commises par des membres de la famille, qu'ils soient proches (comme les parents ou frères et sœurs) ou plus éloignés.
- 3 : **Professionnel** : Cette catégorie inclut toutes les agressions survenues dans un cadre professionnel, que ce soit entre collègues, avec un supérieur hiérarchique ou dans toute autre relation de travail.
- 4 : **École, études supérieures** : Pour toutes les agressions survenues dans un environnement scolaire ou universitaire, que ce soit entre élèves, étudiants ou avec des membres du personnel éducatif.
- 5 : **Espace public** : Cette catégorie englobe les témoignages relatant des agressions survenues dans des lieux publics tels que la rue, les transports en commun ou d'autres lieux accessibles au public.
- 6 : **Couple** : Les témoignages dans cette catégorie se rapportent à des agressions perpétrées par un partenaire amoureux ou conjugal, qu'il s'agisse de relations passées ou actuelles.

La constitution de cette classification a été établie à partir de la lecture des témoignages de la base et de rapports et études spécialisés sur les VSS ([Jaspard \(2007\)](#), [Bajos et al. \(2008\)](#), [Salmona \(2021\)](#)). Ces catégories offrent une vision précise des contextes évoqués par les victimes sur Twitter, tout en permettant des parallèles pertinents avec les données statistiques et les rapports déjà exis-

tants dans ce domaine ([Brown et al., 2020](#)).

2.2.2 D'autres méthodes pour l'analyse des données textuelles

En complément de la classification automatique par CamemBERT, d'autres méthodes plus traditionnelles d'analyse textuelle ont été utilisées pour explorer les tweets. Ces techniques permettent d'extraire des informations complémentaires sans travail préalable d'annotation, notamment sur les termes récurrents.

La première méthode employée est le **TF-IDF (Term Frequency-Inverse Document Frequency)**, une technique statistique utilisée pour évaluer l'importance d'un mot dans un corpus de textes (voir Méthode dans l'encadré [2.2.2](#)). Plus précisément, cela permet de mettre en évidence les termes spécifiques à un texte relativement aux autres textes (ie. des termes très présents dans un texte mais peu présents en général). Cette méthode permet de dégager facilement les thèmes spécifiques de différents documents/textes/...

Méthode Term Frequency-Inverse Document Frequency (TF-IDF)

Le TF-IDF est une mesure statistique utilisée pour évaluer l'importance d'un mot dans un document qui fait partie d'un corpus. Elle est le produit de deux statistiques, la fréquence du terme (TF) et la fréquence inverse du document (IDF).

$$\text{TF}(t, d) = \frac{\text{Nombre de fois que le terme } t \text{ apparaît dans le doc } d}{\text{Nombre total de termes dans le document}}$$
$$\text{IDF}(t, D) = \log \left(\frac{\text{Nombre total de documents dans le corpus } D}{\text{Nombre de documents où le terme } t \text{ apparaît}} \right)$$
$$\text{TF-IDF}(t, d, D) = \text{TF}(t, d) \times \text{IDF}(t, D)$$

Les **expressions régulières (regex)** ont aussi été utilisées, permettant d'effectuer une recherche systématique dans un corpus textuel. Ces expressions régulières sont des séquences de caractères qui définissent des motifs à rechercher dans du texte. Elles permettent de détecter de manière automatisée des occurrences spécifiques de mots, chiffres ou symboles dans de larges corpus textuels. Dans le cadre de cette étude, elles ont été utilisées pour identifier des mentions précises de l'âge des victimes au moment des violences, dans les témoignages postés sur Twitter. Les outils associés aux expressions régulières permettent non seulement d'identifier l'occurrence de ces séquences, mais aussi d'en extraire des caractères spécifiques (ici des nombres entre 0 et 99). Les nombres écrits en toutes lettres ("sept", "vingt-quatre", ...) ont été traduits en format numérique au préalable. Ainsi, j'ai utilisé les expressions régulières afin de rechercher des séquences telles que "j'avais ... ans" ou "à l'âge de ..." et d'en extraire l'âge mentionné.

2.2.3 L'analyse des métadonnées

Finalement, l'analyse des métadonnées associées aux tweets constitue une autre composante essentielle de la méthodologie de ce travail. Les métadonnées fournissent des informations contextuelles précieuses qui permettent d'étudier non seulement le contenu des tweets, mais aussi leurs dynamiques de diffusion et les pratiques des utilisateurs.

1. L'évolution temporelle des témoignages est analysée à partir des dates de publication des tweets. Cela permet de mettre en évidence les moments de forte activité autour de certains hashtags et de comprendre comment les contenus se sont succédé au fil du temps.
2. L'analyse des tweets populaires a été réalisée grâce aux likes et retweets. Ces indicateurs sont utilisés pour identifier les contenus ayant une forte résonance dans #Metoo sur Twitter. Cette analyse permet d'identifier les contenus et les comptes les plus populaires ou viraux.
3. Enfin, les informations sur les comptes (par exemple, le pseudo, le nombre d'abonnés ou le nombre de tweets publiés) ont été utilisées pour étudier les pratiques de discrétion des utilisateurs.

Chapitre 5 : Les mesures de la qualité des méthodes

La validation des méthodes utilisées est essentielle, notamment lors de l'utilisation de techniques de classification automatique. Ces méthodes sont au cœur des résultats de ce travail de mémoire et doivent donc être fiables. Étant donné la complexité des modèles et le nombre très important de données à traiter, une vérification manuelle complète serait inefficace. C'est pourquoi des mesures de

performance sont utilisées pour évaluer la qualité des classification. Ces indicateurs permettent de s'assurer que les modèles produisent des résultats fiables et généralisables.

Pour rappel, deux modèles de classification automatique ont été appliqués au corpus de tweets. Il est donc essentiel d'évaluer :

1. La fiabilité des classifications par type de contenu (témoignages, opinions, médias, etc.).
2. La pertinence des prédictions pour identifier le contexte des violences dans les témoignages (connaissance, couple, famille, espace public, scolarité, monde professionnel, autre) .

Les mesures utilisées sont les suivantes :

$$\mathbf{Accuracy} = \frac{\text{Nombre de prédictions correctes}}{\text{Nombre total de prédictions}}$$

$$\mathbf{Precision} = \frac{\text{Vrais positifs}}{\text{Vrais positifs} + \text{Faux positifs}}$$

$$\mathbf{Recall} = \frac{\text{Vrais positifs}}{\text{Vrais positifs} + \text{Faux négatifs}}$$

$$\mathbf{F1\ Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Résultats du premier modèle

Le premier modèle a été évalué avec les mesures suivantes :

Métrique	Valeur
Accuracy	0.8717
Precision	0.8642
Recall	0.8717
F1 Score	0.8649

TABLE 4 – Résumé des mesures de performance de l'identification des contenus

Ces résultats montrent que le modèle est performant, avec une accuracy de 87,17 %, ce qui indique que le modèle a correctement classé la majorité des tweets. La précision et le rappel sont également élevés, traduisant un bon équilibre entre les prédictions correctes et la détection des catégories pertinentes.

La matrice de confusion et le rapport de classification (Annexe 5) fournissent une image plus détaillée des erreurs commises par le modèle, ils permettent d'avoir le détail des mesures pour chaque catégorie. Ils montrent par exemple si le modèle a plus de mal à identifier les tweets classifiés comme *Autre*, il est particulièrement efficace pour identifier les autres catégories. Par exemple, un *rappel* de 0.92 pour les témoignages signifie que parmi tous les tweets réellement classés comme témoignages, le modèle en identifie correctement 92 %. Cela montre que le modèle est particulièrement performant pour ne pas manquer de témoignages (peu de faux négatifs).

Résultats du second modèle

Le second modèle a été développé pour identifier les contextes dans lesquels les violences ont eu lieu (familial, professionnel, scolaire, etc.) parmi les témoignages

de violences. Initialement, ce modèle s'appuyait, comme le premier modèle, exclusivement sur une annotation manuelle d'un échantillon de tweets, ce qui a permis de créer une première base pour l'entraînement. Cependant, le nombre important de catégories à classer dans ce modèle (7, contre 4 dans le premier) nécessite une base d'entraînement plus importante. En effet, dans cette phase initiale, le modèle a atteint une accuracy de 76 %. Cela signifie que, malgré une base d'entraînement annotée avec soin, le modèle avait encore des difficultés à identifier les contextes décrits parmi les 7 possibles. Pour surmonter cette limitation, une phase d'annotation automatique a été introduite à partir des champs lexicaux des différents contextes. Cette annotation automatique a significativement enrichi la base de données d'entraînement, permettant au modèle de bénéficier d'un échantillon plus large pour s'entraîner.

Voici les mesures de performance après l'ajout de l'annotation automatique :

Métrique	Valeur
Accuracy	0.9476
Precision	0.8972
Recall	0.8922
F1 Score	0.8939

TABLE 5 – Résumé des mesures de performance de la classification automatique des contextes d'agression

L'effet de cette extension de la base annotée a été particulièrement marquant : grâce à l'augmentation du nombre de tweets annotés (à la fois manuellement et automatiquement), le modèle a atteint une accuracy de 95 %, soit une amélioration

de près de 20 points. Ce bond montre bien que l'apprentissage sur un échantillon de données plus grand permet au modèle de mieux identifier les contextes spécifiques des violences rapportées dans les témoignages, tout en réduisant les erreurs de classification. En outre, la matrice de confusion et le rapport de classification (Annexe 5) montre que la fiabilité du modèle est stable sur toutes les catégories.

En complément des mesures de performance classiques telles que l'accuracy, le TF-IDF (voir Encadré 2.2.2) a été utilisé comme un indicateur de pertinence lexicale, permettant d'évaluer la qualité des résultats de manière plus intuitive et compréhensible. Le TF-IDF mesure l'importance relative d'un terme dans un corpus de texte donné. En appliquant cette méthode aux tweets classifiés comme témoignages, il devient possible de vérifier quels termes apparaissent fréquemment dans chaque contexte de violence identifié. Cela permet d'évaluer la cohérence des classifications en identifiant les mots qui caractérisent spécifiquement chaque catégorie de contexte.

Les résultats TF-IDF (détails en Annexe 5) montrent une bonne distinction des contextes d'agression en fonction des mots les plus représentatifs :

- **Connaissance** : Les termes récurrents comme "ami", "voisin", "soirée", "pote" montrent que ces violences surviennent principalement dans des cercles sociaux proches, du moins avec un certain degré de connaissance.
- **Inceste** : Les mots "père", "mère", "famille", "oncle" soulignent la forte présence des relations intra-familiales, validant la précision du modèle pour détecter ce contexte spécifique de violence.
- **Professionnel** : Les termes "collègue", "chef", "patron" mettent en lumière les dynamiques de pouvoir au travail, avec une focalisation sur les relations

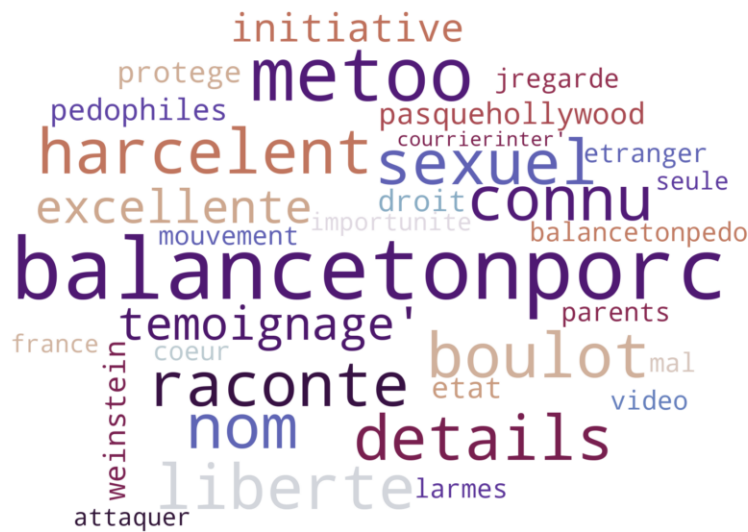
hiérarchiques et les interactions professionnelles.

- **Études** : Les mots "prof", "collège", "cours" montrent un lien direct avec le milieu scolaire.
- **Espace public** : Les termes "rue", "métro", "bus" pointent vers les violences dans des lieux de transport ou d'interaction publique.
- **Couple** : Les mots "copain", "mari", "ex" soulignent le lien conjugal présent ou passé dans les violences déclarées.

Ces résultats confirment la capacité du modèle à capturer les spécificités lexicales de chaque contexte de violence, offrant une validation supplémentaire de la pertinence des résultats.

Partie 3 : Nature, contenu et dynamique des témoignages sur Twitter

FIGURE 2 – Nuage de mots produits automatiquement à partir des mots les plus présents dans les tweets étudiés



Chapitre 6 : Temporalité et persistance de #MeToo sur Twitter

3.1.1 Chronologie et volumétrie

Le mouvement #MeToo a retenti profondément au niveau international, trouvant une résonance immédiate en France avec le hashtag #BalanceTonPorc, qui a émergé simultanément. Cette double mobilisation, locale et internationale, a suscité une réponse immédiate des victimes, qui ont rapidement relayé leurs témoignages en ligne via ces hashtags. Au-delà des scandales initiaux, les premières

semaines ont révélé une réappropriation intense du mouvement par les internautes, marquant un tournant dans la prise de parole numérique autour des violences de genre. Dès lors, il est intéressant de revenir sur la part réelle qu'ont pris les témoignages dans ce fort engouement numérique et quelles suites ont pu être observées dans le contenu partagé les mois et les années suivantes.

Tout d'abord, en observant la Figure 3, le premier constat à souligner est la persistance du mouvement sur Twitter. Ainsi bien que ce mouvement ait pris des dimensions politiques, sociales et médiatiques au-delà du support numérique, ce dernier demeure un support privilégié pour la parole des victimes et pour les commentateurs et commentatrices du mouvement (voir la catégorie *opinions*). De plus, un second constat concerne le moment même du lancement du mouvement. Puisque au lancement même des hashtags #Metoo et #Balancetonporc (respectivement le 15 et 13 octobre 2017), ce sont près de 15 000 témoignages de victimes qui ont été partagés entre mi-octobre et fin octobre 2017. Ce pic qui ne sera jamais plus atteint ni même approché durant la continuité du mouvement démontre bien une attente latente de la part de certaines victimes de témoigner des violences qu'elles ont vécues.

En termes de volume, on compte plus de 335 000 tweets répartis sur 5 années dont près de 58 000, tous contenus confondus, ont été publiés en l'espace de deux semaines (entre le 13 octobre 2017 et le 31 octobre 2017). Au niveau de la répartition du contenu (Figure 4), on observe une large domination du contenu d'opinion. Cette observation n'est pas surprenante puisque cela regroupe des contenus communs sur Twitter, qu'il s'agisse d'une réaction à une affaire médiatique, un tweet de soutien aux victimes ou une critique envers la démarche de celles-ci. Si une personne peut facilement publier plusieurs tweets de soutien ou de critique,

FIGURE 3 – Nombre de tweets mensuels par contenu

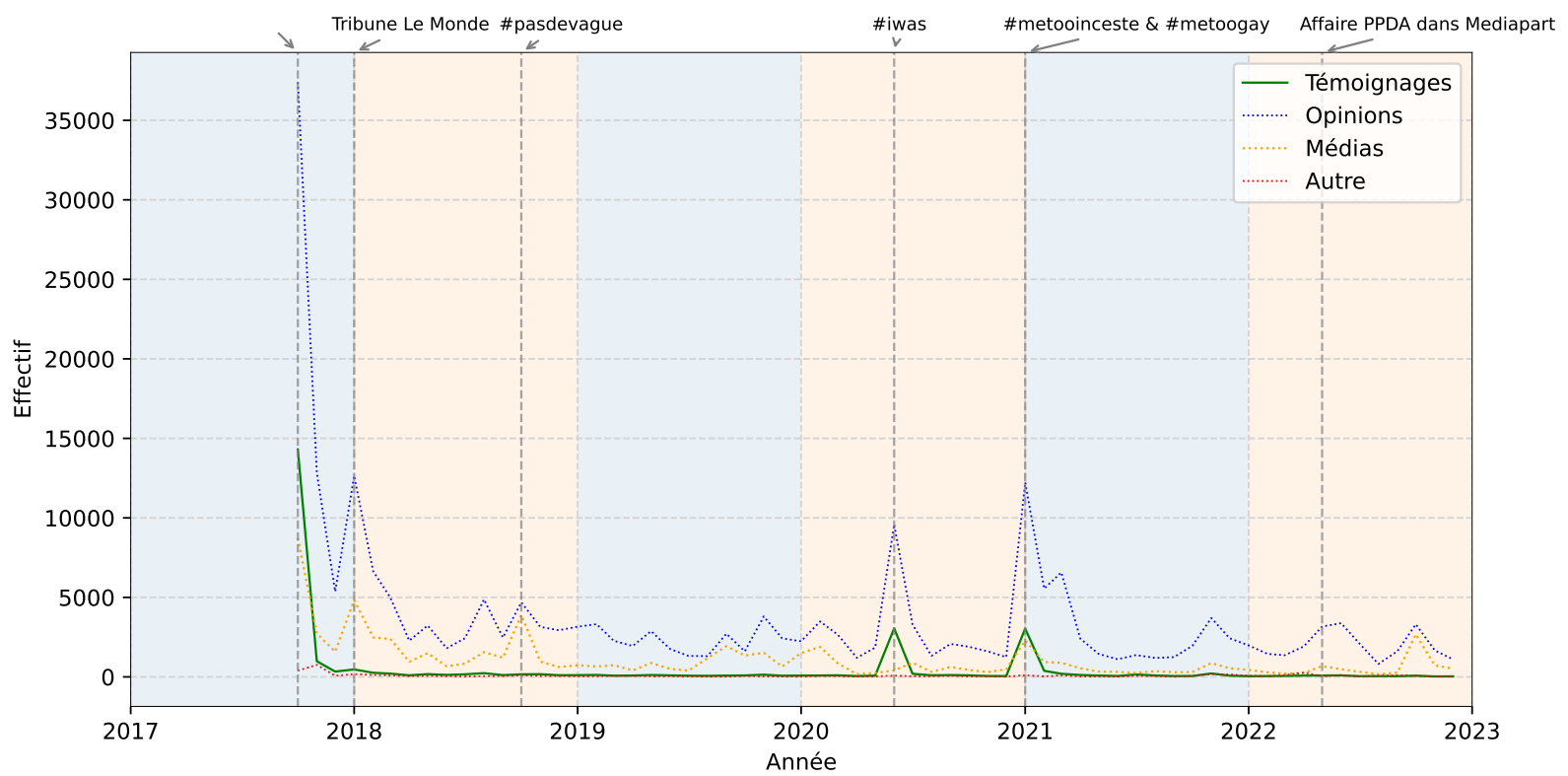
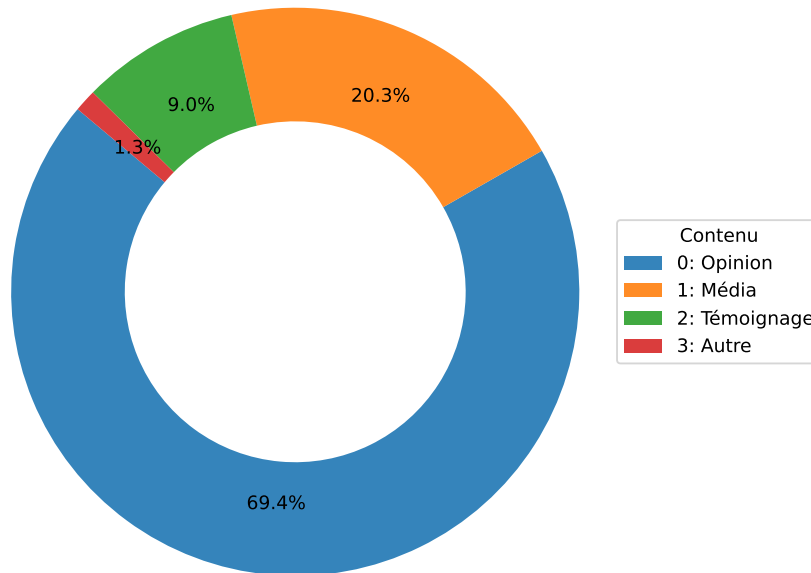


FIGURE 4 – Répartition des données par contenu



elle va ne potentiellement publier qu'un ou deux tweets pour témoigner d'une violence vécue. Il est donc naturel de recenser une grande part de contenus d'opinion. La présence de contenu médiatique (articles de presse, de podcast, d'émission radio, ...) s'explique aussi très bien car une grande majorité des médias numériques, radiophoniques, ou papier relaient quotidiennement tous leurs titres sur Twitter. Les témoignages trouvent donc une place plus petite mais non négligeable face à la profusion d'opinions et de contenus médiatiques. En effet, le modèle de classification a permis d'identifier près de 29 000 témoignages publiés entre octobre 2017 et décembre 2022 sur Twitter.

Parmi ces témoignages, près de la moitié ont été publiés dès les deux premières semaines du mouvement, soit en octobre 2017. Ensuite, seuls deux moments sont associés à de nouveaux pics de témoignages dans le graphique : le lancement de

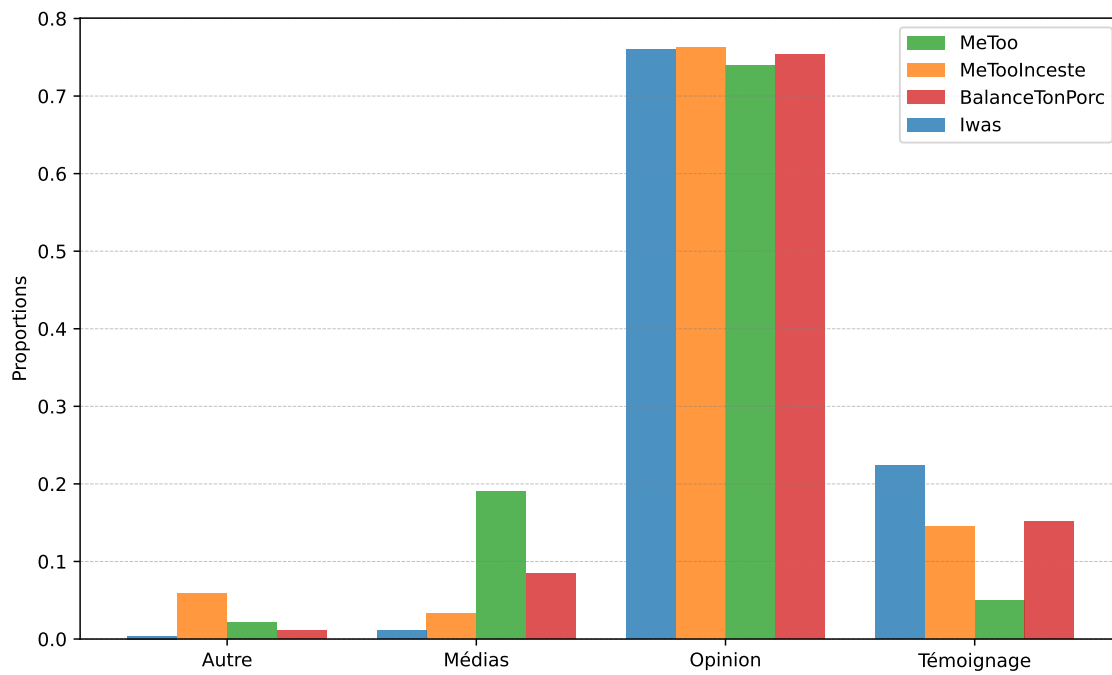
#Iwas en 2020 et le lancement de #MetooInceste et #MetooGay en janvier 2021. En réalité, cela ne signifie pas que ce sont les seuls élans de témoignages sur Twitter durant les 5 années étudiées, mais plutôt que le lancement d'un nouveau hashtag initie généralement une vague de témoignages très concentrés. En effet, si d'autres hashtags avaient été ajoutés à l'étude, la même concentration de témoignages aurait pu être observée. Il s'agit toutefois ici de hashtags très influents qui ont permis la publication de plusieurs milliers de nouveaux témoignages à leurs lancements. Par ailleurs, bien que les jours suivants les lancements de hashtags aient permis de recenser près de 20 000 témoignages, 9 000 autres ont été publiés tout au long des 5 années étudiées, en dehors de tout moment médiatique ou politique. Cette chronologie démontre le besoin persistant pour certaines victimes d'utiliser cette plateforme numérique pour témoigner.

3.1.2 Les types de contenu partagés et leurs évolutions

Nous l'avons vu, une partie du contenu numérique produit autour de #Metoo sur Twitter répond à certaines dynamiques temporelles, avec certains moments plus forts que d'autres, malgré un flux continu. Les lancements de nouveaux hashtags, souvent associés à des moments politiques ou sociaux spécifiques, rythment en partie la publication de contenu mais diffèrent dans leur message et dans leurs effets.

Premièrement, en comparant les hashtags #Metoo et #Balancetonporc, tous deux lancés simultanément en octobre 2017, on observe des différences marquantes dans le contenu produit (Figure 5). Si les tweets d'opinion restent proportionnellement stables, les parts de tweets médiatiques et de témoignages diffèrent nettement. On constate une proportion bien plus élevée de contenus médiatiques pour

FIGURE 5 – Répartition des données par contenu et par hashtag



#Metoo, au détriment des témoignages. En effet, au fil des mois et des années, *Metoo* s'est ancré dans le langage courant pour désigner des dénonciations de violences sexuelles, en particulier dans la sphère publique, et ce, bien au-delà de son contexte numérique. Progressivement, le terme a été repris par les médias, non seulement pour commenter le mouvement en ligne, mais aussi pour évoquer plus largement les cas de violences sexuelles à partir de 2017.

Ainsi, l'expression *metoo* est souvent employée pour décrire toute révélation postérieure à 2017 concernant une personnalité publique en France, que la victime revendique ou non l'utilisation de ce terme. *Metoo* devient alors non seulement une date ("depuis #Metoo"), mais aussi un nom ou adjectif désignant une révélation ("une affaire #metoo", "un nouveau #metoo", ...). Cet ancrage dans le vocabulaire médiatique, associé à son lien fort avec les personnalités publiques françaises, explique cette sur-représentation de contenus médiatiques associés à ce hashtag en comparaison aux autres. Cette évolution a également pu restreindre son usage par des victimes souhaitant témoigner, mais ne se reconnaissant pas dans un terme perçu comme davantage lié aux sphères médiatiques et politiques qu'aux communautés numériques.

FIGURE 6 – Exemple d'utilisation du terme #Metoo par un média en 2022



À l'inverse, le hashtag #Balancetonporc a suscité une utilisation massive, avec

près de 55 000 mentions dans les premiers jours du mouvement, soit bien plus que les 10 000 mentions de #Metoo sur la même période (Annexe 13). Pourtant, #Balancetonporc n'a pas acquis le même ancrage sémantique que #Metoo. Il est néanmoins associé à une proportion de témoignages trois fois plus élevée. La démarche initiale de ce hashtag, centrée sur le partage direct d'expériences, a ainsi été bien suivie, bien que sa popularité n'ait pas perduré dans le temps.

Pour autant, nous l'avons vu précédemment, les témoignages ne se sont pas essouffés après octobre 2017 et ont même été largement relancés à plusieurs moments, lors d'autres temps politiques, médiatiques, sociaux ou numériques. Les cas de #Metooinceste et #Iwas incarnent très bien ces temps forts. En effet, au-delà de réactiver la prise de parole numérique, ces initiatives ont permis l'émergence de nouveaux types de récits centrés sur des expériences de violences spécifiques.

Les données permettent de montrer une variation temporelle marquée selon les contextes d'agression dénoncés (Annexe 5). Les premiers mois, dominés par #Metoo et #Balancetonporc, sont caractérisés par un pic de témoignages dans les espaces publics et professionnels, en résonance avec les récits d'Alyssa Milano et de Sandra Muller, initiatrices de ces hashtags en 2017 et dénonçant des violences dans des relations professionnelles respectivement dans l'industrie cinématographique et médiatique. Ce volume initial de témoignages tend à s'essouffler par la suite, ce qui peut s'expliquer par l'essouffement du hashtag #Balancetonporc en France. Au total ce sont plus de 5 000 témoignages de violences dans l'espace public et 4 000 dans la sphère professionnelle qui ont été identifiés (Figure 7).

À partir de 2020, une évolution significative apparaît avec une augmentation de témoignages liés à l'enfance. Ce changement est particulièrement marqué avec le hashtag #Iwas, qui appelle à mettre en avant l'âge au moment des violences afin

d’alerter sur la pédocriminalité. Quelques mois plus tard avec l’arrivée #Metooinceste, l’attention sera recentrée sur les violences intra-familiales telles que l’inceste. Ces deux hashtags permettent un élargissement du mouvement vers des violences plus intimes, rarement évoquées dans les premières vagues de témoignages. Cependant, malgré un moment médiatique fort autour de la question de l’inceste et donc du cadre familial et la présence de la question des violences conjugales dans l’agenda politique (par exemple avec le Grenelle des violences conjugales fin 2019), les témoignages de violences conjugales n’apparaissent que rarement dans les récits étudiés. En effet, alors que l’on compte plus de 4 000 témoignages sur les violences incestueuses, environ 800 seulement portent sur les violences conjugales. Ces chiffres suggèrent que, malgré l’attention médiatique et politique, les témoignages numériques ne suffisent pas à répondre aux besoins de nombreuses victimes. Celles-ci, souvent en situation de danger immédiat, sous emprise psychologique ou économique de longue durée, ou en recherche d’un accompagnement social durable, sont très probablement moins représentées dans ces récits en ligne.

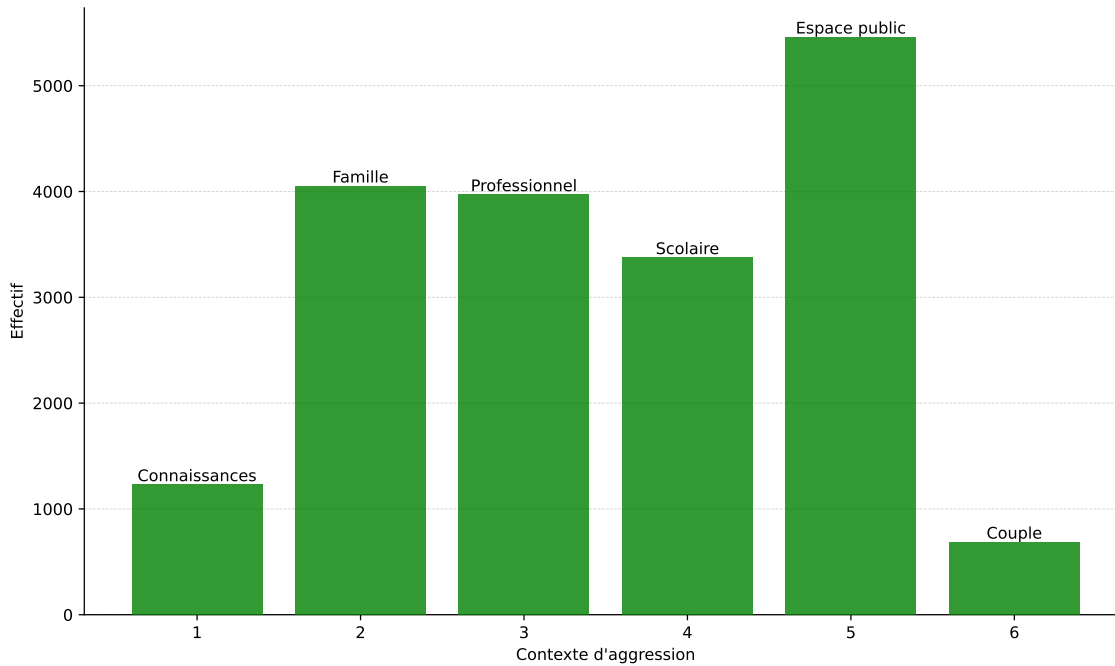
Finalement, les récits relatifs aux cadres scolaires et aux relations de connaissance se révèlent, quant à eux, assez stables sur la période étudiée.

Chapitre 7 : Les spécificités des récits numériques étudiés

3.2.1 Des formes de témoignages propres à la plateforme

Les témoignages publiés sur Twitter présentent des caractéristiques distinctives qui les différencient des récits exprimés dans d’autres espaces de déclaration. La plateforme impose une limitation de caractères par tweet, ce qui amène naturellement les utilisateurs et utilisatrices à faire des choix de mots précis et à

FIGURE 7 – Répartition des données par contenu et par hashtag



structurer leur récit de façon succincte. Cependant, il est à noter que la plupart des victimes n'utilisent pas la totalité des caractères autorisés par Twitter. Ce choix volontaire de concision dépasse donc la simple contrainte technique : il reflète une tendance à se concentrer sur les éléments les plus marquants de leur expérience.

En ce sens, les témoignages observés se déclinent principalement en trois formes distinctes :

1. **L'énonciation factuelle d'un événement unique** : Des faits isolés, relatés avec peu de mots pour exprimer un seul moment marquant. Ici, ce sont généralement des éléments descriptifs qui sont décrits par les victimes : leur âge au moment des violences, le contexte et parfois le type de violence.
2. **La dénonciation d'une période ou d'un contexte de violences vécues** :

Certaines victimes choisissent de dénoncer une phase étendue de violences, ou de décrire un environnement (familial, professionnel, scolaire) dans lequel les violences se sont inscrites de manière répétitive.

3. **La dénonciation d'un système :** Au-delà des faits vécus, certains témoignages cherchent à témoigner des systèmes de *silenciation* (Dotson, 2011), les difficultés rencontrées dans le parcours de dénonciation (difficulté à porter plainte, obstacles rencontrés dans les institutions ou parmi les proches). Ces témoignages dénoncent donc non seulement un ou des faits mais plus largement les difficultés dans leur parcours.

Tweet 7

"#Iwas 11 ans à ma première main au fesses dans les transports."

Tweet 8

#metooinceste Mon père nous a violenté y compris sexuellement toute notre enfance et adolescence mes 2 frères et moi. Ma mère ne nous a pas protégé. J'ai eu peur de lui jusqu'à mes 40. J'en ai 53 aujourd'hui. C'est la première fois que j'ose une parole publique.

Tweet 9

"j'en avais parlé à ma mère qui ne m'a pas cru."

Tweet 10

"#balancetonporc portée plainte, le policier a dit que c'était de ma faute."

Les témoignages présentent parfois plusieurs de ces formes réunies.

Tweet 11

j'avais 8 ans et on m'as dis d'arrêter de raconter des bêtises

j'avais 12 ans et on m'as dis de ne pas faire de scandale

j'avais 15 ans et on m'as traité de menteuse #MeToo

On observe également que le vocabulaire juridique est peu mobilisé dans ces témoignages. Les victimes décrivent parfois des faits que l'on pourrait clairement assimiler à une infraction (agression sexuelle, harcèlement), mais elles préfèrent des descriptions factuelles aux termes juridiques. Parfois même, la caractérisation de la violence n'est pas envisagée et des mots plus généraux tels que "violence", "abus", "expériences" ou "histoires" sont utilisés.

De plus, le ton des témoignages est généralement neutre. Le choix des mots et la concision confèrent une sobriété qui se traduit par un langage précis. Enfin, la dénonciation nominative est extrêmement rare. Parmi les lectures effectuées (plus de 8 000 tweets), moins de 15 cas dénoncent une personne identifiable. Dans certains de ces cas, le nom est remplacé par une fonction ou un titre identifiant (maire d'une ville par exemple). Cette rareté montre que, malgré le caractère public de la plateforme, la majorité des victimes choisissent de préserver l'anonymat de leurs agresseurs, et ce malgré les critiques du "tribunal médiatique" souvent associées à Twitter. La part relative des témoignages dans le mouvement et la forme de ces témoignages vont à l'encontre de l'article de [Lopez et al. \(2019\)](#), qui avance que l'approche française du mouvement est particulièrement *aggressive et accusatrice*.

Par ailleurs, ces différentes façons de témoigner révèlent deux approches principales dans la construction du récit sur Twitter : une approche factuelle et concise d'un côté, et une approche introspective et contextuelle de l'autre. La première

forme, que l'on pourrait qualifier de récit factuel, consiste en une énonciation directe et concise des éléments de l'agression. Ce choix, souvent minimaliste, semble répondre à une volonté de présenter les faits de manière brute. Ce format très succinct peut inclure des informations comme l'âge, le lieu, et parfois le type de violence. La seconde approche néglige parfois la description des faits pour exprimer une expérience plus introspective, souvent plus personnelle et émotionnelle. Dans ces (courts) récits, la victime ne se contente pas de rapporter l'agression ; elle aborde également son rapport à l'événement, ses répercussions émotionnelles et les réactions sociales ou institutionnelles auxquelles elle a fait face. Ces deux approches montrent bien la diversité des stratégies de témoignage sur Twitter, avec des victimes qui choisissent de témoigner d'un fait (ou plusieurs) et d'autres d'un parcours. Cette variété dans les modes de récits est particulièrement intéressante, car elle montre comment chaque victime adapte son témoignage à ses besoins et à son rapport personnel à la parole.

3.2.2 Des pratiques de publication propres aux témoignages

Au-delà du choix des mots et de la forme narrative, les victimes disposent sur Twitter de divers moyens pour ajuster leur récit. Alors que les choix sémantiques et syntaxiques permettent de transmettre leur expérience de manière plus personnelle ou distanciée, d'autres outils techniques disponibles modulent également la manière dont ces personnes partagent leurs témoignages. Ces outils numériques élargissent le spectre des pratiques de publication, en donnant aux victimes un contrôle accru sur la visibilité, la durée, et l'accessibilité de leurs récits et de leur identité. Les outils numériques disponibles sur Twitter offrent en effet des options pour moduler l'exposition et la pérennité de leurs discours. Dans le contexte

du mouvement #MeToo, ces instruments, tels que l’anonymat, l’éphémérité et les options de suppression, peuvent être saisis par les utilisateurs ou utilisatrices pour partager des récits isolés ou, au contraire, pour rechercher un soutien ou des interactions numériques.

Suppression volontaire ou involontaire des tweets

Un constat important dans l’analyse des données est l’éphémérité de certains tweets, dont une proportion significative semble avoir été supprimée. En comparant les bases de données collectées, on constate que 62 000 tweets présents dans la base de l’INA (issue d’une collecte au jour le jour) sont absents de la collecte de fin décembre 2022. Cette différence pourrait suggérer qu’une partie de ces tweets a été supprimée de la plateforme, entre sa publication et la collecte de décembre 2022. Par ailleurs, un nombre important de ces tweets sont des témoignages et des tweets d’opinion. La surreprésentation de témoignages dans cet échantillon potentiellement supprimé (13 % contre 7,45 % dans la collecte de l’API¹) suggère qu’une part de ces tweets pourrait avoir été volontairement retirée. Ce phénomène pourrait être dû à une volonté de retrait de la parole après un moment de partage, que ce soit par choix personnel, pour éviter des répercussions sociales, en réaction à des commentaires hostiles, ou pour éviter de la violence supplémentaire en ligne ou hors ligne.

Outre la suppression volontaire, une part de ces disparitions pourrait également résulter de la modération exercée par la plateforme. La recherche de [Grison et al.](#)

1. Je compare ici un sous-échantillon de 62 000 tweets issus de la base de données INA avec l’ensemble de la collecte API. Ces 62 000 tweets sont potentiellement supprimés, mais ne représentent pas l’intégralité de l’échantillon de l’INA illustré dans la Figure 8. C’est pour cette raison que la proportion de 13 % dans ce sous-échantillon diffère des 11,35 % observés dans l’échantillon complet.

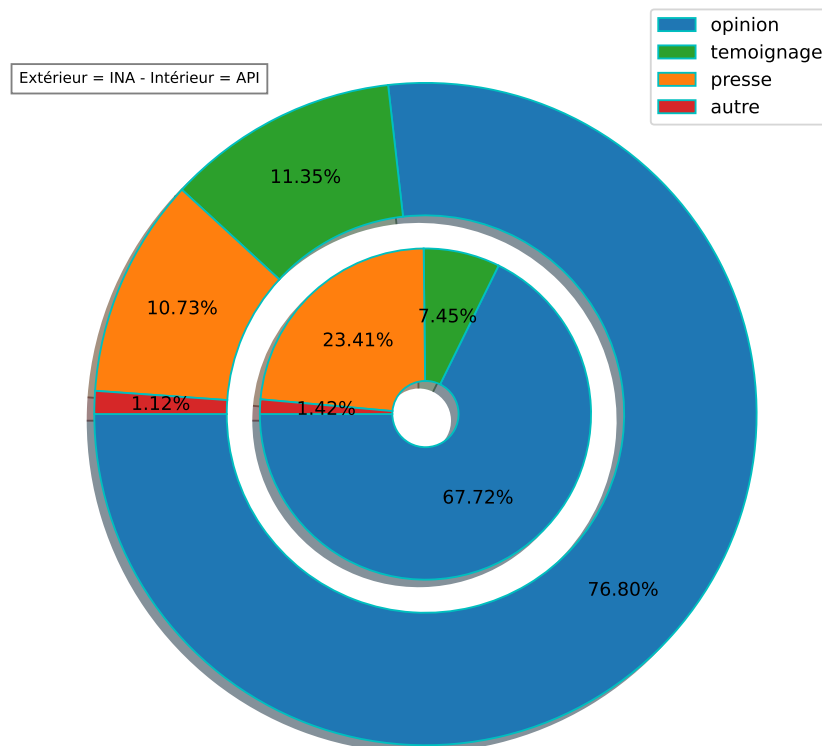
(2023) montre que les dispositifs de modération peuvent affecter la visibilité de certains groupes, en particulier les communautés LGBT et du travail du sexe. Cette étude révèle que certains mots-clés, en raison de leur association à une culture sexuelle ou de genre, peuvent déclencher une modération automatisée, même lorsque leur usage n'est pas injurieux. Il est dès lors possible que des témoignages du mouvement, en utilisant un vocabulaire inhérent à la description de faits de violences ou de discrimination, se retrouvent malencontreusement dans le filet de la modération. En conséquence, certaines voix peuvent être invisibilisées malgré elles, mettant en lumière le besoin de stratégies de modération mieux calibrées pour éviter de censurer des témoignages légitimes. Cette censure involontaire peut être une forme de violence silencieuse qui empêche l'expression de violences et prive des victimes de cet espace numérique.

Stratégies de discrétion et d'anonymat

Les victimes qui choisissent de partager leur expérience de violence sur Twitter semblent adopter des stratégies de discrétion et d'anonymat plus fréquemment que d'autres utilisateurs et utilisatrices. En plus de pouvoir supprimer leurs tweets à tout moment, les utilisateurs disposent d'autres options dès la création de leur compte. Ainsi, la publication de témoignages sous des pseudonymes, l'utilisation de comptes récents ou peu suivis, et la diffusion unique de leurs récits sont des stratégies couramment utilisées pour gérer l'exposition personnelle. Ces choix témoignent de la précaution prise par les victimes pour éviter d'être identifiées ou associées directement à leur récit.

La Table 6 présente une analyse des utilisateurs dans l'ensemble de la base de tweets, en comparant les comportements des comptes ayant publié des témoignages

FIGURE 8 – Répartition des tweets par contenu selon le mode de collecte



Lecture : Dans le disque intérieur est représentée la répartition des tweets par contenu de la collecte issue de l'API Twitter, ie. la collecte a posteriori effectuée début 2023. Le disque extérieur représente cette répartition pour la collecte journalière faite par l'INA.

de violences à ceux ayant partagé d'autres types de contenus. Les données montrent une nette tendance à l'anonymat parmi les comptes qui publient des témoignages :

- 13,07 % des témoignages proviennent de comptes ayant moins de 5 abonnés, contre seulement 6,92 % pour les autres types de tweets. Cela indique que les victimes de violences sont plus susceptibles de publier leur expérience depuis des comptes peu suivis, minimisant ainsi leur visibilité au sein du réseau.
- 3,40 % des témoignages proviennent de comptes ayant publié un seul tweet, contre seulement 0,66 % pour les autres types de tweets. Ce phénomène souligne l'usage de comptes créés uniquement pour la diffusion d'un témoignage ponctuel.
- 6,25 % des témoignages proviennent de comptes créés dans la semaine précédant la publication, contre 2,72 % pour les autres tweets. Cette donnée révèle une préférence pour les nouveaux comptes, suggérant de nouveau que certaines victimes ont pu créer un profil temporaire uniquement pour témoigner.

Base complète d'utilisateurs	Témoignages uniquement	Sans témoignages	Total
- de 5 abonnés	13.07%	6.92%	7.73%
1 seul tweet	3.40%	0.66%	1.02%
- de 5 abonnés & 1 tweet	3.29%	0.62%	0.97%
Compte créé \leq 1 semaine	6.25%	2.72%	3.18%
Effectif	15 430	101 974	117 404

TABLE 6 – Analyse des utilisateurs dans la base de tous les tweets

3 PARTIE 3 : NATURE, CONTENU ET DYNAMIQUE DES TÉMOIGNAGES SUR TWITTER

Base d'utilisateurs avec des tweets supprimés	Témoignages uniquement	Sans témoignages	Total
- de 5 abonnés	15.39%	10.83%	11.55%
1 seul tweet	4.66%	1.34%	1.87%
- de 5 abonnés & 1 tweet	4.49%	1.24%	1.75%
Compte créé \leq 1 semaine	7.05%	4.66%	5.04%
Effectif	4 768	25 514	30 282

TABLE 7 – Analyse des utilisateurs dans la base de tweets potentiellement supprimés

Les mêmes tendances se retrouvent dans la Table 7, qui analyse cette fois les utilisateurs de la base de tweets potentiellement supprimés. Dans cette base de données :

- 15,39 % des témoignages proviennent de comptes avec moins de 5 abonnés, et 4,66 % de comptes ayant émis un seul tweet. Ces proportions sont nettement plus élevées que dans la base complète, indiquant que les tweets de témoignages issus de comptes discrets sont également plus susceptibles d'être supprimés.
- 7,05 % des témoignages de cette base proviennent de comptes créés dans la semaine précédant la publication, contre 4,66 % pour les autres tweets supprimés. Ce chiffre semble confirmer l'hypothèse selon laquelle la création de comptes temporaires est une stratégie privilégiée par les victimes de violences pour témoigner en limitant l'exposition prolongée de leur récit.

Ces observations montrent que les pratiques d'anonymat et de discrétion peuvent se renforcer parmi certaines victimes qui semblent vouloir protéger leur exposition sur la plateforme et qui retirent parfois leur témoignage. Ces pratiques permettent

aux victimes de contrôler leur visibilité sur la plateforme, dans un contexte où elles peuvent s'exposer à des formes de violence numérique.

Violences numériques et anonymat

La volonté de préserver l'anonymat peut également être expliquée par la violence numérique que peuvent subir les victimes lorsqu'elles témoignent publiquement sur des réseaux sociaux. En partageant leurs expériences de violences, notamment dans le cadre du mouvement #Metoo, les victimes s'exposent à des réactions hostiles. Harcèlement, insultes et tentatives de discrédit sont fréquents, et peuvent exercer une pression psychologique importante sur les personnes concernées. Cette dynamique peut décourager l'expression publique et pousser les victimes à adopter des stratégies de protection, comme l'usage de pseudonymes ou la création de comptes temporaires.

[Mendes et al. \(2018\)](#) analysent cette ambivalence et soulignent que, bien que les plateformes numériques comme Twitter offrent un espace essentiel pour briser le silence autour des violences sexuelles, elles sont également le théâtre de *pièges* qui menacent les victimes d'une violence secondaire. En effet, le simple fait de partager un témoignage en ligne expose les victimes à des agressions numériques, des réponses de dénigrement ou des attaques en ligne, exacerbant souvent la vulnérabilité initiale. Cette violence numérique peut donc expliquer en partie le choix de l'anonymat et des comptes peu visibles ou temporaires comme stratégie d'anticipation. En s'exprimant à travers un compte jetable, les victimes limitent les risques de représailles directes et évitent de rendre leur récit accessible. La protection de leur identité et de leur témoignage est donc une mesure de sécurité face aux menaces inhérentes à l'espace public numérique.

Partie 4 : Analyse comparative de différents espaces de déclaration

Les témoignages de violences publiés sur Twitter diffèrent bien sûr des données issues des données nationales. Contrairement à ces dernières, où la collecte d'informations est dirigée par un service statistique, structurée autour de questionnaires précis ou émanant directement des plaintes et signalements, les déclarations sur Twitter sont spontanées, non sollicitées, et ne suivent aucun cadre formel de recensement. Comme nous l'avons vu précédemment, cette liberté de forme confère aux récits en ligne une dimension subjective, où l'intention des témoignages diffère selon les parcours et besoins des victimes. Les statistiques nationales comportent pourtant elles aussi des biais (voir sous-partie 1.1.2), notamment en raison des mécanismes de sous-déclaration, de sous-estimation ou d'amnésie traumatique. Un nombre important de victimes ne signalent pas les faits comme l'illustrent les données de l'enquête Virage (Brown et al. (2020)) et les rapports du ministère de l'Intérieur (données du SSMSI). Ainsi, comparer les biais et les dynamiques de déclaration de différents espaces est pertinent pour élargir notre compréhension, non seulement de la prévalence réelle des violences, mais surtout des mécanismes par lesquels les victimes choisissent, ou non, de rendre compte de leur vécu. Cette comparaison nous aide à mieux cerner les obstacles à la parole et les spécificités des récits dans chaque espace, permettant de mieux appréhender les dynamiques de déclaration et de non-déclaration, notamment en comparant des supports en ligne et hors ligne.

Chapitre 8 : Comparaison entre déclarations numériques et statistiques nationales

4.1.1 Prévalence des violences anciennes dans les témoignages numériques : une confirmation des tendances observées depuis #MeToo

Des récits de faits anciens, au-delà de l'instantanéité de Twitter

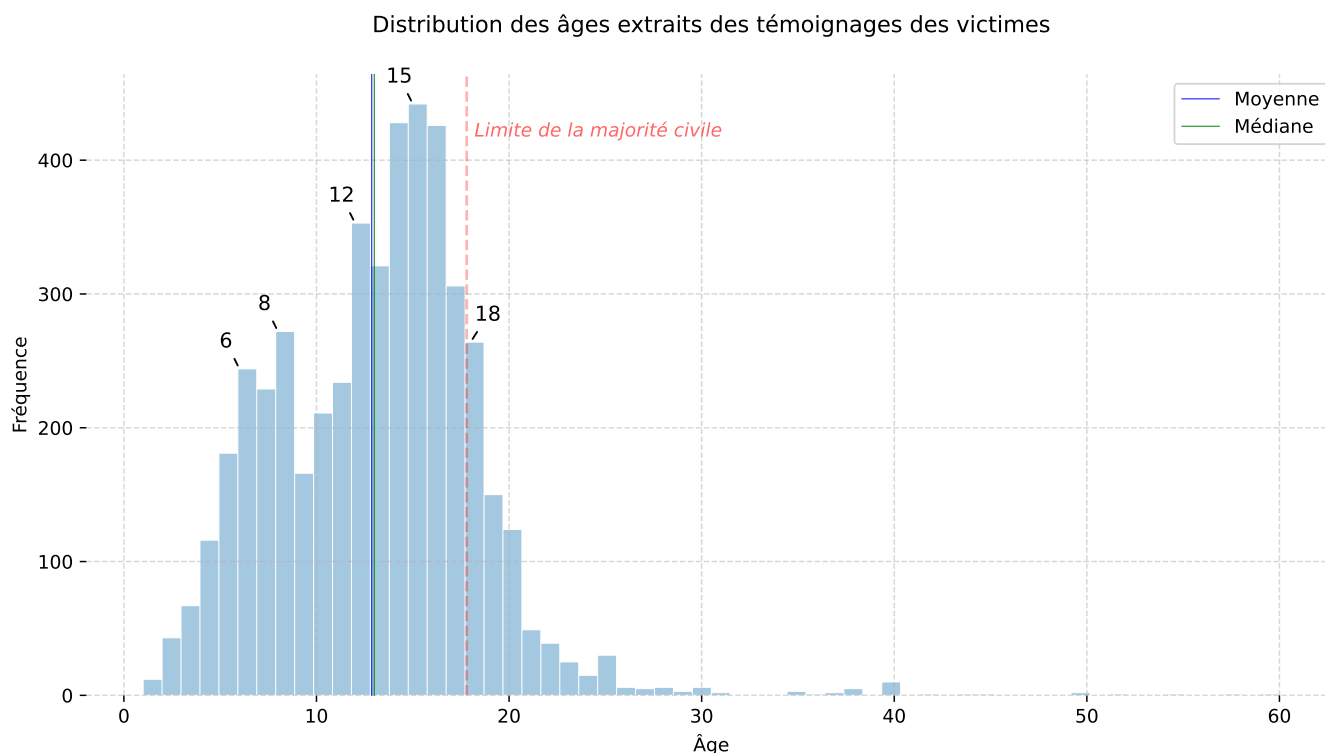
L'analyse des témoignages sur Twitter révèle que les utilisateurs et utilisatrices relatent fréquemment des événements marquants survenus plusieurs années auparavant, plutôt que des violences *quotidiennes* ou récentes. Dans le contexte de Twitter où les publications sont associées à l'instantanéité et l'actualité, ce constat est particulièrement notable.

Ce constat appuie les différents résultats concernant les délais de prises de parole des victimes, que ce soit aux proches ou aux institutions. Dans de nombreux cas de violences faites aux enfants, ceux-ci les verbalisent au minimum 12 ans après les faits, 14 ans en cas de viols¹. Pour autant, si cette verbalisation est souvent tardive, les signalements de ces faits anciens ont progressé. Le ministère de l'Intérieur a constaté que depuis l'émergence de #MeToo, les déclarations de violences anciennes ont fortement augmenté. Selon un rapport publié en 2021, "le nombre de victimes ayant déclaré des faits commis plus de cinq ans auparavant a plus que doublé entre 2016 et 2020"². Twitter semble ainsi répondre la même dynamique de verbalisation différée qui est déjà constaté chez les jeunes victimes et parmi les signalements.

1. Enquête IPSOS pour l'Association Mémoire Traumatique et Victimologie, *Violences sexuelles de l'enfance*, 2019

2. SSMSI, « En 2020, le nombre de plaintes enregistrées pour des faits anciens de violences sexuelles continue à augmenter - Les délais d'enregistrement des faits de délinquance », Éclairage n° 2, in *Insécurité et délinquance en 2020 – Bilan statistique*, avril 2021c.

FIGURE 9 – Répartition des âges déclarés dans les témoignages



Distribution des âges dans les témoignages : une forte proportion de mineurs

L'étude des âges mentionnés dans les témoignages appuie également ce constat de déclaration différée. Comme le montre la Figure 9, la distribution des âges suit une distribution proche d'un mélange de gaussiennes, avec un pic majeur à 15 ans et un autre, plus modeste, entre 6 et 8 ans. La majorité des victimes déclarant leur âge au moment des violences dans les tweets les situent donc bien avant 15 ans, avec quelques âges extrêmes (de 1 à 60 ans) rappelant que la vulnérabilité aux violences n'a pas de limite d'âge.

Tweet 12

"J'avais 1 an. Un adulte voisin. Répétées. [...] Dépôt plainte, classé sans suite par procureur. Il est mort. [...]"

Tweet 13

"Un associé medecin : ne veut pas partir de chez moi [...] j'avais 60 ans ! [...]"

Pour faciliter la comparaison avec les enquêtes nationales, les tranches d'âge utilisées dans le tableau 8 reprennent celles de l'enquête Virage. Bien que la catégorie "25 ans et plus" puisse sembler large, elle reflète un constat important : les déclarations de violences pour des âges adultes sont beaucoup plus rares, tant sur Twitter que dans l'enquête Virage (qui interrogeait des personnes de 18 à 69 ans). En effet, la répartition par tranches d'âge sur Twitter montre des écarts notables avec Virage :

- Les groupes d'âge 0-10 ans, 11-14 ans, et 15-17 ans sont fortement sur-représentés sur Twitter.
- Les tranches 18-24 ans et "25 ans et plus" sont sous-représentées, surtout cette dernière, quasiment absente de Twitter.

TABLE 8 – Tableau des répartitions des âges déclarés dans l’enquête Virage et dans les données du mémoire

Groupe d’âge	Moyenne des résultats Virage ¹	Observations Twitter ²
0-10 ans	26.2	32.1
11-14 ans	17.4	27.8
15-17 ans	14.9	24.4
18-24 ans	15.9	13.9
25 ans et +	25.4	1.8
Ne sait pas	0.2	-
Total	100	100

Ces résultats suggèrent donc que les déclarations sur Twitter tendent à se concentrer sur les violences subies durant l’enfance. Pour autant, les hashtags jouent un rôle crucial dans cette dynamique. Les hashtags comme #Iwas et #Metooinceste sont associés de manière significative à la mention d’âges précoces : plus de 70 % des tweets spécifiant un âge incluent ces hashtags. Le hashtag #Iwas, par sa construction même, invite à une déclaration mentionnant un jeune âge en incitant l’utilisateur à compléter la phrase, par exemple ”I was 8 [*when it happened*]”. Cette formulation met en exergue l’âge comme élément central du témoignage, et invitait à dénoncer l’ampleur de la pédocriminalité. De même, #Metooinceste, bien qu’il n’impose pas un format explicite, est associé aux violences intrafamiliales qui touchent principalement les jeunes enfants. Ces hashtags orientent donc la structure du récit, alors que la majorité des témoignages ne mentionnent pas

1. Debauche, Alice, et al. ”Présentation de l’enquête Virage et premiers résultats sur les violences sexuelles.” Ined, Document de travail 229 (2017).

2. Parmi les témoignages mentionnant un âge aux violences (n = 4 809)

d'âge. Par ailleurs, les résultats de Virage indiquent que 80,7% des femmes et 84,9% des hommes ayant été agressés en étant mineurs affirment que les auteurs de l'agression a profité de leur âge, plaçant cette information comme un élément clef du contexte des violences.

La démographie de Twitter, plus jeune que la population générale, pourrait en partie expliquer cette tendance à la mention d'âges précoces dans les récits, mais elle ne justifie pas la sous-représentation des âges supérieurs à 18 ans. Il est aussi important de noter qu'une majorité de témoignages (23 498 sur 28 307) ne mentionnent pas d'âge. Alors que nous avons déjà constaté la concision des témoignages numériques (voir 3.2), les victimes semblent opérer une sélection des informations les plus marquants de leur expérience. Les adultes, en particulier, semblent privilégier d'autres aspects de leur récit, comme le lieu de l'agression ou la nature de leur relation avec l'agresseur, pour refléter les éléments qu'ils jugent les plus significatifs de leur expérience.

4.1.2 Différences de contextes de violences : des écarts qui interrogent les dynamiques de déclaration

Les témoignages publiés sur Twitter depuis #Metoo diffèrent également de manière significative des données issues de l'enquête Virage en termes de contexte des violences rapportées. La Table 9 compare les pourcentages de témoignages selon différents contextes dans Virage et sur Twitter, et plusieurs écarts notables apparaissent :

- Les violences dans l'**espace public**, qui représentent 47,02 % des déclarations dans Virage, ne constituent que 29,08 % des témoignages sur Twitter.
- À l'inverse, les violences subies dans des **milieux professionnels et sco-**

laïres sont proportionnellement plus présentes sur Twitter, représentant respectivement 21,13 % et 17,97 % des témoignages, contre 10,83 % et 7,10 % dans Virage.

- Les **violences incestueuses** sont présentes en proportions proches dans les deux sources de déclarations.
- Les **violences conjugales**, enfin, sont sous-représentées sur Twitter (3,66 %) par rapport aux déclarations recueillies dans Virage (8,32 %).

Influence de la méthodologie et des dynamiques de plateforme

Ces écarts peuvent en partie s'expliquer par la différence de méthodologie entre les deux espaces de déclaration. Virage, en tant qu'enquête de victimation nationale (voir 1.1.3) interroge un échantillon de manière systématique, en couvrant tous les espaces (public, privé, professionnel, etc.) et en recueillant tous types de violences (verbales, psychologiques, physiques, sexuelles). Les témoignages sur Twitter, quant à eux, sont des déclarations spontanées et non encadrées, où les victimes sélectionnent librement les expériences qu'elles souhaitent partager. Ces différences méthodologiques rendent pertinente la comparaison entre les deux sources, car elles révèlent non seulement les contextes majoritairement évoqués par les victimes dans leurs récits publics spontanés, mais aussi la façon dont les dynamiques de la plateforme orientent implicitement les choix de déclaration, par des hashtags, des mouvements et une influence des affaires publiques et médiatisées. En effet, les déclarations sur Twitter ont fortement évolué au fil des mois et années entre 2017 et 2022 pour dénoncer des contextes différents (voir 3.1.2), d'abord une majorité de témoignages dénonçant des violences au travail et dans l'espace public puis une majorité de témoignages de violences pédocriminelles et incestueuses en 2021 (Annexe 5).

TABLE 9 – Comparaison des contextes déclarés des témoignages selon Virage et Twitter

Contexte	Virage (%)	Twitter (%)
Espaces publics	47.02	29.08
Famille et proches	26.68	28.14
Professionnel	10.83	21.13
Scolaire	7.10	17.97
Conjugal	8.32	3.66
Approximation du total	100	100

La notion de gravité perçue dans les déclarations

Les écarts observés entre les données de Virage et les témoignages publiés sur Twitter pourraient aussi s’expliquer par la notion de gravité perçue des violences. Selon les résultats de l’enquête Virage, toutes les violences ne sont pas perçues avec le même degré de gravité par les victimes, et cette perception va bien au-delà d’une simple hiérarchisation juridique. En effet, si la gravité d’un acte isolé peut être comprise de manière relativement universelle, d’autres facteurs influent fortement sur la perception de gravité et les conséquences sur la santé : la répétition d’actes violents, la durée d’exposition à la violence, le contexte dans lequel elle se déroule et la présence d’autres types de violences concomitantes. Ces éléments peuvent aggraver les conséquences sanitaires et sociales de la violence vécue, particulièrement dans des environnements clos et durables, comme le milieu conjugal ou familial.

Dans l’enquête Virage, les violences déclarées dans l’espace public sont, dans la majorité des cas, perçues comme moins graves par les victimes, notamment

si elles sont très rares ou ponctuelles. Cette perception de gravité peut expliquer pourquoi les violences en espace public sont moins souvent rapportées sur Twitter. Dans cet espace de déclaration spontanée, les victimes ont tendance à partager des expériences qu'elles estiment marquantes ou représentatives de leur vécu, alors que les violences jugées moins graves trouvent davantage leur place dans des enquêtes statistiques comme Virage, où tous les types de faits sont recensés de manière systématique. La relation entre gravité perçue, conséquences sur la santé et choix de déclaration pourrait donc jouer un rôle important. Un fait ancien qui a eu des répercussions profondes sur la vie et la santé de la victime – émotionnelle, physique ou sociale – est plus susceptible de rester marqué dans la mémoire et d'être révélé dans un témoignage public, même longtemps après les faits. À l'inverse, un événement récent, mais perçu comme moins grave, est moins probable d'être partagé sur une plateforme comme Twitter.

Cependant, il est essentiel de nuancer cette relation entre gravité et déclaration. Comme souligné dans la littérature (voir chapitre 1), des mécanismes sociaux et psychologiques contribuent à une forte *silenciation* des victimes, rendant les déclarations publiques rares et partiellement représentatives. Les récits disponibles sur Twitter ne représentent donc qu'une fraction des expériences vécues, en excluant les voix silencieuses ou silencieées. De plus, si la gravité des violences semble effectivement jouer un rôle dans la décision de témoigner sur les réseaux sociaux, cette gravité perçue est souvent associée à des atteintes graves à la santé et au bien-être.

Les violences conjugales, par exemple, sont typiquement associées à une exposition prolongée et répétée, menant à des conséquences sévères sur la santé des victimes. Cela pourrait expliquer leur sous-représentation dans les témoignages

Twitter. Les victimes de violences conjugales, soumises à une détérioration progressive de leur santé physique et mentale, nécessitent souvent un soutien sanitaire, juridique ou social avant de pouvoir témoigner publiquement. Twitter, avec ses récits différés dans le temps, reflète davantage les parcours de victimes ayant entamé une introspection ou un processus de verbalisation déjà entamé. En ce sens, les récits publiés sur Twitter semblent concerner des violences perçues comme graves et vécues dans des environnements violents ou marquants, mais émanent également de victimes ayant acquis un certain recul sur leur expérience.

Chapitre 9 : Comparaison entre les déclarations en ligne et hors ligne

4.2.1 De la porosité entre les espaces numériques, médiatiques et politiques : les cas de #BalanceTonPorc et #MeTooInceste

Les cas #BalanceTonPorc et #MeTooInceste montrent comment des témoignages de violences peuvent transiter entre le numérique, la justice, et les médias en France, mais ils se distinguent également par la direction de leur résonance. Dans le cas de #BalanceTonPorc, c'est un témoignage numérique qui a entraîné un fort retentissement médiatique et judiciaire, tandis que pour #MeTooInceste, c'est une révélation hors ligne – par la publication d'un livre – qui a engendré une vague de témoignages numériques. Ces deux cas illustrent les interactions complexes entre espaces en ligne et hors ligne dans la médiatisation et la gestion des récits de violences sexuelles.

#BalanceTonPorc : de la dénonciation numérique à la confrontation judiciaire

Le mouvement #BalanceTonPorc, initié par le tweet de Sandra Muller, journaliste, accusant un ancien patron de chaîne de télévision de harcèlement, a cristallisé certaines critiques récurrentes du mouvement plus global #MeToo en France, notamment celles associant le numérique à un "tribunal médiatique". Non seulement des tribunes sont signées dans les médias traditionnels pour s'opposer en partie au mouvement, mais très vite la croyance que tous les témoignages pointent du doigt et nomment des personnes sans preuve sur Twitter se répand. Les expressions de "tribunal populaire" ou "tribunal médiatique" sont employées et inquiètent. C'est pourquoi le cas de #Balancetonporc est intéressant. Après son tweet initial, Sandra Muller a été poursuivie pour diffamation précisant en raison de ses tweets, et bien que condamnée en première instance, elle a ensuite été disculpée en appel. Ce procès, dont la première décision a été très médiatisée, a amplifié les critiques envers le mouvement.

L'impact de cette affaire a eu un effet symbolique et dissuasif. Sandra Muller a été contrainte de supprimer son tweet à la suite du jugement, marquant la première implication de la justice dans les racines numériques du mouvement. Dans les témoignages étudiés dans ce mémoire, cette crainte de poursuites juridiques pourrait expliquer en partie la proportion de tweets potentiellement supprimés. Ce procès a aussi pu avoir des conséquences sur le choix d'anonymat ou sur le non recours à la dénonciation nominative sur Twitter. Cependant la condamnation a eu lieu en septembre 2019, soit deux ans après le lancement du mouvement et il n'y avait déjà quasiment aucune nomination.

#MeTooInceste : une dynamique hors ligne renforcée par le numérique

Le cas #Metooinceste incarne une dynamique inverse, où une révélation hors ligne – la publication du livre *La Familia Grande* de Camille Kouchner (Kouchner, 2021) – a déclenché une réaction numérique, puis politique. Ce récit littéraire a exposé une affaire de violences incestueuses impliquant une personnalité publique, suscitant une résonance médiatique et émotionnelle intense. En quelques jours, cette onde de choc a donné naissance au hashtag #Metooinceste, utilisé par de nombreuses victimes pour partager leur propre expérience de violences intrafamiliales (Figure 3). Ce mouvement en ligne, né d'un témoignage littéraire, a aussi eu une résonance politique. Emmanuel Macron s'est exprimé peu après, sur Twitter, afin d'annoncer que des mesures seront prises. En mars 2021, Adrien Taquet, Secrétaire d'État chargé de la protection de l'enfance, annonce la création de la Commission indépendante sur l'inceste et les violences sexuelles faites aux enfants (Ciivise), avec pour mission de recueillir les témoignages des victimes et de formuler des recommandations. La Ciivise a d'ailleurs adopté une démarche participative, en lançant des appels à témoignages à l'échelle nationale, une approche qui montre pourrait être vue comme une institutionnalisation des appels à témoignages numériques que constituent les créations successives de hashtags. En effet, de telles démarches participatives en terme de violences sexuelles, visant à recueillir des témoignages en dehors du cadre judiciaire afin de produire des recommandations ou des réparations, n'existaient pas avant #Metoo. Le cas de la CIASE (Commission indépendante sur les abus sexuels dans l'Église) commandée par la Conférence des évêques de France en 2018 en est un des premiers exemples. Ces cas se rapprochent du fonctionnement des commissions de vérité et de réconciliation (CVR) et seront réitérés par différentes instances pour comprendre l'ampleur et les causes des violences sexuelles dans différents secteurs

socio-culturels en France. Ainsi, cela prouve une nouvelle fois la difficulté actuelle pour la justice de traiter l'ampleur des violences sexuelles et permet de penser la légitimité de la parole des victimes en dehors du cadre judiciaire. Le cas de #Metooinceste démontre la porosité entre culture, numérique et politique, tout en amenant à réfléchir à des cadres d'accueil de la parole des victimes pour l'action publique.

4.2.2 Entre démocratisation et *silence des dominés*

Les espaces de témoignages de violences, tels que Twitter et les médias traditionnels, révèlent des différences marquées dans les profils de déclarants, les formes de récit et les intentions associées aux témoignages. Les hashtags #Metoo, #Metooinceste et #Balancetonporc illustrent par exemple le constat que les affaires initiales médiatisées émanent de milieux socioculturels favorisés, que ce soit par la victime, la personne accusée, ou les deux. Cette section examine ces distinctions en se concentrant sur les écarts socioculturels entre les espaces, les formes d'expression et les différentes intentions qui motivent les déclarations. L'analyse est limitée par la nature des données de Twitter, ne fournissant que peu d'informations sur le profil des utilisateurs. Nous avons choisi de n'inférer aucune conclusion sur le genre, l'opinion politique ou l'origine des profils, car ces méthodes comportent de nombreux biais et limites, et beaucoup de profils préfèrent probablement des pseudonymes ou des photos non conformes. Nous proposons néanmoins des pistes de réflexion pour réfléchir aux dynamiques socioculturelles des différents médias.

En analysant les 10 témoignages les plus retweetés (Annexe 5), une tendance intéressante se dégage : sur ces 10 témoignages très visibles, avec plusieurs cen-

4 PARTIE 4 : ANALYSE COMPARATIVE DE DIFFÉRENTS ESPACES DE DÉCLARATION

taines ou milliers de partages, seuls deux sont associés à des profils largement reconnus des médias, journaliste ou chroniqueuse. Cette observation montre que, bien que la visibilité médiatique joue un rôle, la majorité des témoignages les plus partagés émanent de personnes sans notoriété publique préalable. Cette observation contraste avec les médias traditionnels, où les témoignages sont souvent portés par des personnalités publiques. Sur Twitter, les formes de déclarations sont souvent non nominatives et reprennent rarement les caractérisations juridiques, en comparaison avec les dénonciations plus formelles et nominatives faisant généralement appel au vocabulaire juridique dans les médias traditionnels. La médiatisation peut être volontaire pour dénoncer ou mobiliser, ou involontaire lorsque l'accusation concerne une personnalité publique.

FIGURE 10 – Exemples d'affaires médiatisées

#MeToo dans le cinéma : l'actrice Adèle Haenel brise un nouveau tabou

L'actrice Adèle Haenel accuse le réalisateur Christophe Ruggia d'« atteintes » et de « harcèlement sexuel » lorsqu'elle était âgée de 12 à 15 ans. Son récit est conforté par de nombreux documents et témoignages. Mediapart retrace son long cheminement, de la « prise de parole impossible » au « silence devenu insupportable ». Le cinéaste conteste « catégoriquement » les faits.

Marine Turchi
3 novembre 2019 à 18h53

Troisième vague #MeToo dans le jeu vidéo : au Canada, une championne du monde prend la parole

Lorsqu'elle est arrivée chez Ubisoft à Montréal en 2009, Stéphanie Harvey dit avoir subi harcèlement ou sexisme "dès le jour 1".

20 juil. 2020

Libération

Récits de harcèlement et d'agressions sexuelles à Ubisoft : « Les jeux vidéo c'est fun, on peut tout faire, rien n'est grave »

«Libération» a recueilli une vingtaine de témoignages qui décrivent, au siège du numéro 3 mondial du secteur, un système toxique, dominé par...

1 juil. 2020

tv5monde

Scandale de harcèlement sexuel chez Ubisoft : une "culture d'entreprise" à revoir

L'industrie du jeu vidéo se retrouve à son tour prise dans la vague #MeToo. Plusieurs têtes sont tombées chez Ubisoft après des accusations...

4 août 2020

ELLE

Affaire Richard Berry : le bruit et la douleur

Alors que la parole des victimes se libère, l'affaire Richard Berry a mis les réseaux sociaux en surchauffe. Que faire de cette tempête médiatique ?

10 févr. 2021

20 Minutes

Richard Berry accusé d'inceste : Sa fille Coline entendue par les enquêteurs

La fille aînée de Ricard Berry a déposé plainte contre l'acteur le 25 janvier.

11 févr. 2021

Femme Actuelle

Richard Berry accusé d'inceste : sa fille Coline réaffirme sa position et enfonce le clou

Après avoir porté plainte contre Richard Berry, qu'elle accuse d'inceste, la fille du comédien, Coline Berry, a réaffirmé sa position contre...

27 juin 2021

À partir de notre précédente classification, nous avons extrait les mots-clés de

ces contenus médiatiques partagés sur Twitter à l'aide d'une approche TF-IDF (Term Frequency-Inverse Document Frequency) mensuelle (voir encadré 2.2.2). Cette analyse des mots-clés extraits des contenus médiatiques repartagés sur Twitter (articles de presse, podcast, émissions de télévision, ...) montre une nette ponctuation par des affaires de violences sexuelles et sexistes associées à une personnalité publique. En les observant de plus près, nous pouvons distinguer plusieurs catégories. Parfois, c'est le nom de la victime qui ressort, comme pour Adèle Hanel, Sandra Muller ou Peng Shuai. Dans la majorité des cas, cependant, c'est le nom de l'accusé qui est mis en avant, avec des figures telles que Woody Allen, Jeffrey Epstein, Gabriel Matzneff, Bernard Preynat, Roman Polanski ou Richard Berry. Tous ces noms sont apparus comme l'un des cinq "mots-clés" mensuels dans nos résultats, qui se présentent comme dans le Graphique 11. Une autre catégorie de mots-clés concerne le contexte ou l'environnement dans lequel les violences sont survenues. Ces mots-clés incluent des noms d'entreprises ou d'institutions (comme Ubisoft), des lieux (comme Copenhague ou le Danemark), ou des termes plus généraux évoquant le contexte des violences (comme "inceste" ou "gay"). Pour autant, chaque mois ou presque, un nom d'une personnalité publique française ou étrangère ressort dans les contenus médiatiques partagés sur Twitter.

FIGURE 11 – Exemples de résultats de l'extraction de mots-clefs par fréquence d'apparition dans les contenus médiatiques partagés sur Twitter

Médias - Mots importants pour 2022-09: co (TF-IDF: 0.7204) bayou (TF-IDF: 0.2117) quatennens (TF-IDF: 0.1565) cinq (TF-IDF: 0.1489) radicalisent (TF-IDF: 0.1417) ans (TF-IDF: 0.1376) immobilisme (TF-IDF: 0.1369) femmes (TF-IDF: 0.1215)	Médias - Mots importants pour 2021-07: co (TF-IDF: 0.6938) bragelonne (TF-IDF: 0.2521) écarté (TF-IDF: 0.1738) cosby (TF-IDF: 0.1624) édition (TF-IDF: 0.1308) fonctions (TF-IDF: 0.1291) sextape (TF-IDF: 0.1047) libéré (TF-IDF: 0.0966)
---	--

Par ailleurs, certains médias prennent le temps et les moyens d'une enquête depuis #MeToo ([Beaulieu, 2022](#); [Ruffio, 2022](#)). Cela permet potentiellement un meilleur contexte d'accueil de la parole, des investigations sérieuses, mais émane notamment d'une crainte latente de critique de "tribunal médiatique" ou de diffamation. Il est ainsi important de souligner que les médias traditionnels ne sont généralement pas seulement un support pour un témoignage, qui ne sont pas publiés sous forme de tribune libre, mais principalement sous forme d'enquête, imposant un fort intermédiaire entre la parole de la victime et la médiatisation, contrairement à la publication sur Twitter.

Ces dynamiques rappellent les constats rapportés par [Achin* and Naudier** \(2010\)](#) étudiant les mouvements féministes des années 1970, qui utilisaient des entretiens pour mettre en relief les voix des femmes "ordinaires". Par des entretiens, des récits de vie et des groupes de discussion, la recherche de cette époque mettait en lumière la dichotomie entre les grandes villes, en particulier Paris, et les régions plus éloignées, tout comme entre les figures emblématiques

du féminisme très médiatisées et les protagonistes du quotidien. La politisation de l'intime, la lutte contre les structures patriarcales, l'opposition aux normes sexistes et l'importance de l'intersectionnalité (Bilge (2009), Auclair (2017)), notamment les luttes de classe chez les femmes les plus marginalisées, étaient au cœur des préoccupations.

Ma propre analyse du mouvement #Metoo, en utilisant les données de Twitter, montre des parallèles certains avec ces constats sur les années 1970. Les voix médiatisées ont joué et continuent de jouer un rôle crucial pour initialiser ces mouvements, les propulser sur le devant de la scène, les démocratiser et les mettre en lumière dans les médias. Cependant, tant dans l'étude de 2010 que dans la mienne, les résultats s'accordent : si ces voix médiatisées sont essentielles pour initier le mouvement, la coexistence et l'harmonisation avec les voix "ordinaires" confèrent à ces mouvements leur durabilité, leur résonance et leur profondeur. Par ailleurs, l'accessibilité de Twitter et la résonance des mouvements numériques permettraient idéalement une meilleure communication avec les voix publiques, médiatiques et politiques.

Ainsi, bien que séparés par des époques et des outils, le mouvement féministe français des années 1970 et le mouvement #Metoo témoignent d'une constante : beaucoup de voix participent à la force motrice des mouvements, et il est nécessaire d'étudier ces voix dans leur ensemble. Cependant *beaucoup de voix* ne sont pas toutes les voix.

Même si Twitter permet une certaine démocratisation de la parole, moins contrainte par une logique de visibilisation des paroles de personnalités publiques, cette démocratisation est loin de représenter la diversité des violences (voir chapitre 4.1) et des victimes. Cela est vrai pour Twitter et pour les médias tradition-

nels : la représentation de l'intersectionnalité et de l'hétérogénéité des victimes est lacunaire.

Comme le soulignent Pollak and Heinich (1986) en évoquant l'histoire des camps de concentration :

Dans la constitution d'un corpus de biographies, l'histoire orale permet donc d'élargir le matériel de recherche vers le bas de l'échelle sociale. Mais il n'en reste pas moins que, dans l'un et l'autre cas, l'origine sociale et le niveau d'études sont bien plus élevés que dans la recherche sociographique évoquée précédemment. Cette sur-représentation du haut de l'échelle sociale dans le matériel de type biographique reflète une réalité propre à tout échantillon spontané, et relativement indépendante du phénomène concentrationnaire auquel elle ne fait qu'ajouter des contraintes d'énonciation supplémentaires : à savoir le silence des dominés, que rien n'autorise ou n'incite à raconter une vie à laquelle la qualité de leur propre personne ne semble pas suffire à conférer un intérêt d'ordre plus général.

Michael Pollak et Nathalie Heinich, *Le témoignage*, 1986

En travaillant spécifiquement sur *Le témoignage*, leur approche met en évidence le "silence des dominés", c'est-à-dire le fait que les personnes issues des milieux les plus défavorisés sont moins enclines ou moins autorisées à partager leurs expériences, car elles estiment que leur parcours n'a pas suffisamment d'intérêt général. Ce phénomène se reflète également dans les témoignages sur Twitter et dans les médias traditionnels, où les voix des personnes marginalisées ou d'origine moins valorisées socialement sont sous-représentées, bien que l'oralité et la spontanéité de Twitter permettrait *d'élargir le matériel de recherche vers le bas de l'échelle sociale*. Par ailleurs, leur travail met en avant l'importance d'étudier ce

rapport au témoignage qu'ont (ou non) les victimes et les différentes pratiques de celui-ci (judiciaire, autobiographique, ...), notamment pour étudier un problème de société mal représenté statistiquement et soumis aux biais du silence et du traumatisme.

Conclusion

La conclusion de ce mémoire revient sur les principaux enjeux et résultats abordés tout au long de l'étude, en les replaçant dans une perspective plus large afin de souligner les implications à la fois sociales, juridiques et politiques des témoignages de violences sexistes et sexuelles en ligne.

Les résultats principaux se déclinent en plusieurs aspects clés : d'abord, la volumétrie des témoignages montre une ampleur considérable des récits numériques attestant de l'importance de ces espaces numériques pour la prise de parole de certaines victimes, bien que relativement minoritaire vis à vis des tweets d'opinion et des partages médiatiques. Ensuite, l'analyse du contenu des témoignages révèle des dynamiques variées, notamment une diversité de vécus et de formes d'expression, mettant en avant les interprétations variées de ce qu'est un témoignage de violences, du récit d'un fait à la dénonciation d'un système de silenciation. Par ailleurs, la porosité entre les espaces numériques et les espaces plus institutionnels a été mise en évidence, les témoignages en ligne ayant parfois conduit à des réponses de la part d'acteurs institutionnels ou à une médiatisation accrue de certaines affaires. Ainsi, le numérique a joué un rôle central dans la démultiplication des appels à témoigner, en offrant des espaces permettant aux victimes de violences sexuelles de s'exprimer en dehors des contraintes temporelles et juridiques.

Le mouvement #Metoo, ainsi que d'autres initiatives plus institutionnalisées (la Ciivise ou la CIASE par exemple), permettent aussi de réfléchir à la place des témoignages dans le cadre de l'action publique et juridique. Au regard du constat appuyé par la psychologie (Salmona, 2018), les données judiciaires et l'analyse des témoignages depuis 2017, les victimes témoignent très souvent avec un certain recul temporel et expriment des besoins et conséquences différentes, mettant en avant la nécessité de penser différemment la prise en charge des victimes mais aussi la production de données les concernant. Le numérique a permis d'accéder à un espace de témoignages plus libres, confirmant certains constats établis dans d'autres disciplines et mettant en avant des dynamiques de déclaration propres à ce contexte collectif en ligne.

Cependant, il est essentiel de rappeler que beaucoup de victimes choisissent de ne pas témoigner ou de le faire en se protégeant, soit en évitant des interactions non souhaitées, soit en utilisant l'anonymat pour s'intégrer à une communauté de soutien tout en anticipant de potentielles violences numériques (Mendes et al., 2018). Cette précaution illustre à la fois la volonté des victimes de participer à une prise de parole collective et leur nécessité de se protéger dans un environnement où les violences, même numériques, peuvent persister.

L'une des limites importantes de ce mouvement reste l'intersectionnalité et la représentativité lacunaires de la somme des témoignages produits. Comme le souligne l'ouvrage de Souffrant (2023), il existe *Le privilège de dénoncer* auxquelles certaines femmes noires, immigrées, et les LGBT, n'ont pas accès. Ce manque de représentativité est nécessairement lié au caractère libre et public de ces espaces, où se jouent les rapports de pouvoir présents dans la société. Les biais d'accès à la parole sont multiples : d'une part, ils concernent l'accès pratique aux différents

médias, qu'il s'agisse des plateformes numériques ou des médias traditionnels, et d'autre part, ils se manifestent dans le processus de reconnaissance en tant que victime et de légitimation de sa propre voix (Pollak and Heinich, 1986). Ainsi la production du récit de soi et l'accès aux médias pour le partager constitue un exercice qui n'est pas possible pour toutes et tous.

En conclusion, les réseaux sociaux ont permis à certaines victimes de prendre de nouvelles voies d'expression, ce qui n'est pas nouveau en soi, mais chaque nouvel espace permet à différentes personnes de s'y imposer et de faire perdurer l'expression *MeToo* jusqu'à aujourd'hui, aussi bien en ligne qu'hors ligne. Cependant, cette prise de parole numérique pose des questions quant à la nature de la justice que ces témoignages permettent d'obtenir, et quant aux inégalités persistantes face à l'accès à la parole et à la reconnaissance de celle-ci.

Annexes

Les mesures de performance des modèles

Rapport de classification du premier modèle

label	precision	recall	f1-score	effectif de validation
Autre (0)	0.46	0.19	0.27	32
Médias (1)	0.86	0.85	0.85	165
Opinion (2)	0.91	0.91	0.91	533
Témoignage (3)	0.75	0.92	0.83	73
Moyenne pondérée	0.86	0.87	0.86	803

Matrice de confusion du premier modèle

TABLE 10 – Matrice de Confusion

		Prédiction			
		0	1	2	3
Vrai label	0	6	3	22	1
	1	1	140	24	0
	2	5	20	487	21
	3	1	0	5	67

Rapport de classification du second modèle

label	precision	recall	f1-score	effectif de validation
Autre (0)	0.69	0.78	0.73	161
Connaissance (1)	0.90	0.81	0.85	150
Inceste (2)	0.99	0.99	0.99	1171
Professionnel (3)	0.96	0.96	0.96	451
Scolaire (4)	0.92	0.94	0.93	572
Espace public (5)	0.97	0.95	0.96	776
Couple (6)	0.85	0.81	0.83	42
Total	0.90	0.89	0.89	3323

		Prédiction					
		0	1	2	3	4	5
Vrai label	0	126	3	5	1	13	11
	1	15	122	3	2	7	1
	2	6	2	1158	0	3	2
	3	9	1	2	433	6	0
	4	9	2	2	12	537	6
	5	13	5	2	2	15	739

TABLE 11 – Matrice de Confusion

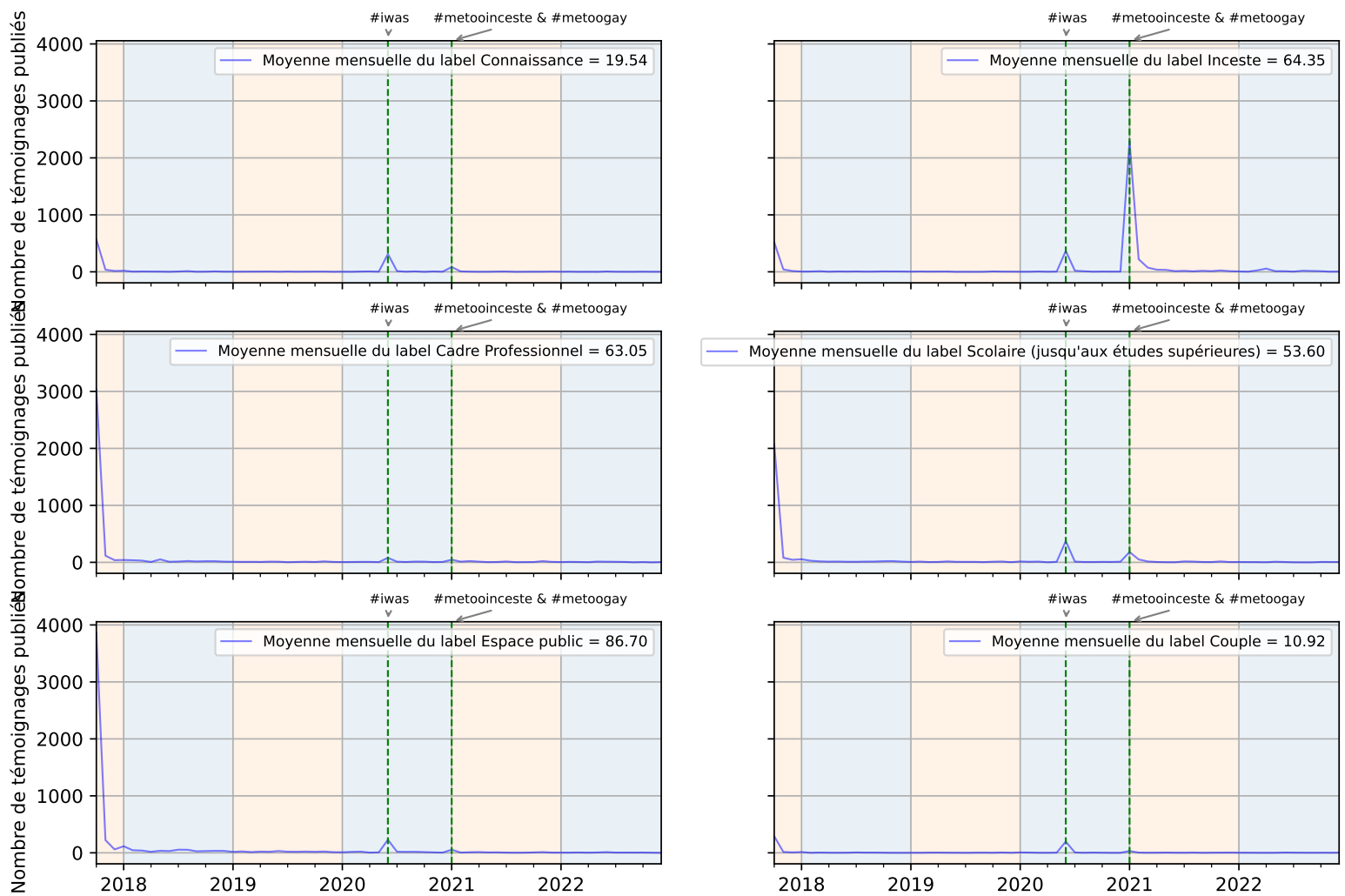
Analyse lexical du second modèle

Connaissance	Inceste	Professionnel	Etudes	Espace public	Couple
ami	metooinceste	balancetonporc	balancetonporc	balancetonporc	copain
voisin	ans	collègue	prof	mec	ex
balancetonporc	père	chef	ans	rue	iwas
iwas	mère	ans	collège	métro	balancetonporc
ans	famille	dit	école	ans	mari
pote	plus	quand	classe	bus	ans
quand	frère	travail	quand	homme	premier
amie	ça	stage	metoo	metoo	ça
chez	grand	patron	cours	quand	non
ça	quand	metoo	lycée	voiture	plus
fait	oncle	fait	iwas	main	quand
soirée	cousin	si	mec	mecs	faire
copine	inceste	bureau	fait	gars	metoo
mec	co	client	dit	dit	si
lit	https	ça	filles	fois	fois
amis	vie	faire	faire	faire	dit
metoo	fait	collègues	ça	fesses	petit
plus	10	main	plus	fait	16
dit	jamais	job	main	chez	envie
faire	tout	boulot	si	ça	époque

TABLE 12 – Mots clefs associés à chacun des contextes d’agressions étudiés

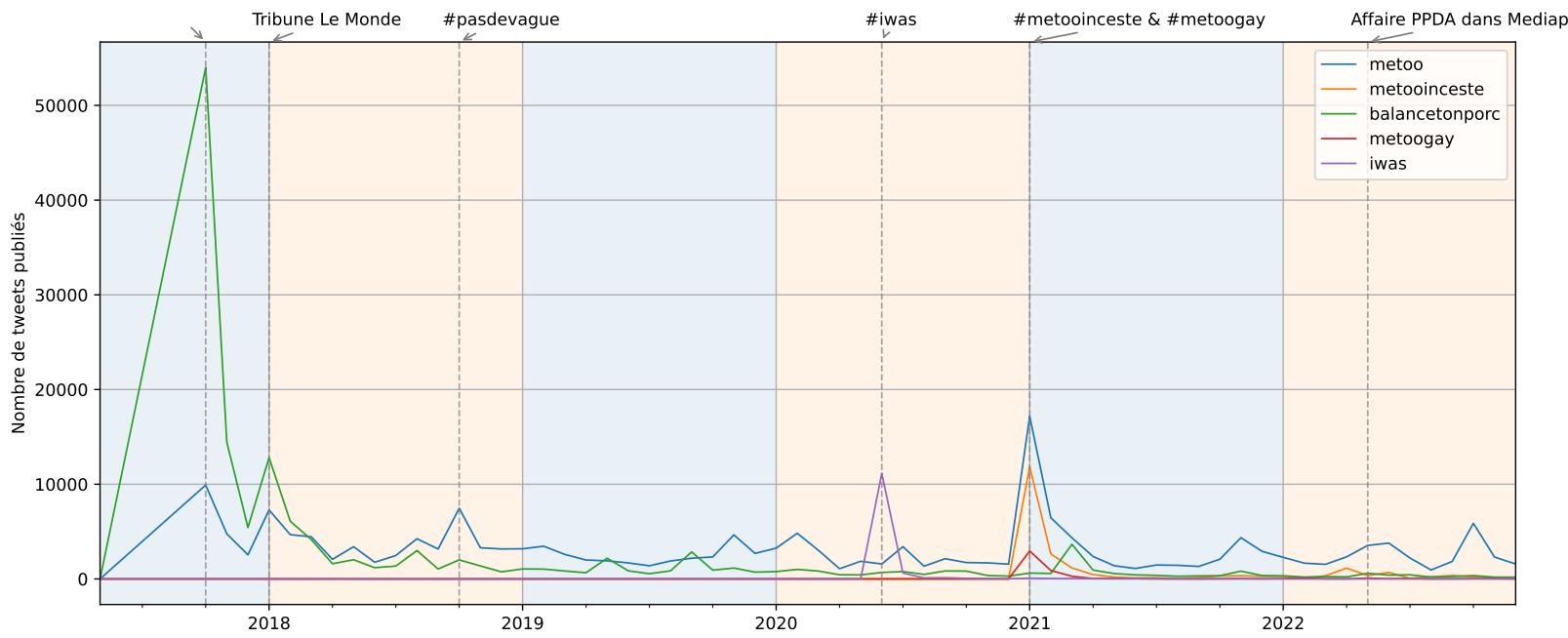
Evolution du nombre des témoignages en fonction du contexte des violences

FIGURE 12 – Témoignages mensuels par contexte déclaré



Evolution du nombre de tweets en fonction du hashtag

FIGURE 13 – Témoignages mensuels par hashtag



Description des témoignages les plus médiatisés sur Twitter

Tweet	Date	Statut	Nombre de RT
"Mon violeur vient d'être condamné à 25 ans..."	01-2021	Profil avec nom et photo	18,730
"Plusieurs personnes m'ont convaincu..."	06-2020	Profil avec nom et photo	4,266
"TW // Twitter j'ai besoin de vous..."	10-2021	Profil avec nom et photo	1,997
"Agressée ce jour par les manutentionnaires..."	07-2022	Profil d'association	1,391
"#Iwas 7. Je m'en souviendrai toujours..."	06-2020	Profil avec nom	1,197
"J'avais entre 11 et 14 ans..."	01-2021	Profil avec nom et photo	954
"(agression sexuelle) Mon ex vient de m'écrire..."	04-2020	Profil anonyme actif et suivi	890
"Merci au chef de bord pour sa réactivité..."	09-2022	Journaliste	779
"Animateur-prod tele dont je refusais les avances..."	10-2017	Actrice, chroniqueuse	702
"#Iwas 16, j'ai subi un attouchement sexuel..."	06-2020	Profil avec nom et photo	605

TABLE 13 – Témoignages les plus retweetés parmi les données étudiées

Références

- Achin*, C. and Naudier**, D. (2010). Trajectoires de femmes “ordinaires” dans les années 1970. la fabrique de la puissance d’agir féministe. *Sociologie*, (1) :77–93.
- Akrich, M. and Méadel, C. (2007). De l’interaction à l’engagement : les collectifs électroniques, nouveaux militants de la santé. *Hermès*, (1) :145–153.
- Auclair, I. (2017). L’analyse intersectionnelle du continuum des violences en situation de refuge : les discriminations en emploi 1. *Recherches féministes*, 30(2) :217–234.
- Bajos, N., Bozon, M., et al. (2008). Les violences sexuelles en France : quand la parole se libère. *Population Sociétés*, (5) :1–4.
- Bajos, N. and Spira, A. (1991). Analyse des comportements sexuels en France (acsf). *Sciences sociales et santé*, 9(4) :57–68.
- Bard, C., Boivineau, P., Charpenel, M., Grailles, B., and Lasserre, A. (2023). *Les féministes et leurs archives*. Rennes (Presses universitaires de).
- Beaulieu, L. (2022). Avant et après# metoo : traiter des violences sexistes et sexuelles en tant que journaliste. *Les violences sexistes après# MeToo*, pages 61–70.
- Beccarelli, M. (2021). *Micros de nuit : histoire de la radio nocturne en France, 1945-2012*. Presses universitaires de Rennes.
- Bertrand, D. (2018). L’essor du féminisme en ligne. *Rezeaux*, (2) :232–257.

- Bilge, S. (2009). Théorisations féministes de l'intersectionnalité. *Diogène*, (1) :070–088.
- Brown, E., Debauche, A., Hamel, C., and Mazuy, M. (2020). Violences et rapports de genre. *Enquête sur les violences de genre en France, Ined éditions, Grandes enquêtes*.
- Cardon, D. (2018). Le pouvoir des algorithmes. *Pouvoirs*, (1) :63–73.
- Cavalin, C., Delage, P., Lefèvre, I. D., Lacombe, D., Pavard, B., et al. (2022). *Les violences sexistes après# MeToo*. Presses des Mines via OpenEdition.
- Chaperon, S. (2002). L'histoire contemporaine des sexualités en France. *Vingtième siècle*, (3) :47–59.
- Charpenel, M. (2014). «*Le privé est politique !*» *Sociologie des mémoires féministes en France*. PhD thesis, Institut d'Etudes Politiques de Paris.
- Clark-Parsons, R. (2021). “i see you, i believe you, i stand with you” :# metoo and the performance of networked feminist visibility. *Feminist Media Studies*, 21(3) :362–380.
- Corrin, C., Hugon, A., and Marro, C. (1997). La violence masculine contre les femmes : résistance et recherche féministes. *Nouvelles questions féministes*, pages 9–47.
- Da Silva, J. (2022). La production et circulation d'un savoir collectif, profane et intersectionnel sur les plateformes numériques. *Balisages*, (5).

- Dauphin, S. (2023). Le féminisme d'état et les violences de genre en France : avancées et limites de la politique de lutte contre les violences conjugales. *Nouvelles questions féministes*, 42(1) :101–116.
- Debarbieux, É. (2004). Les enquêtes de victimation en milieu scolaire : leçons critiques et innovations méthodologiques. *Déviance et société*, 28(3) :317–333.
- Debauche, A. (2011). *Viol et rapports de genre : émergence, enregistrements et contestations d'un crime contre la personne*. PhD thesis, Paris, Institut d'études politiques.
- Delage, P. (2017). *Violences conjugales. Du combat féministe à la cause publique*. Paris : Presses de Sciences Po.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert : Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv :1810.04805*.
- Dotson, K. (2011). Tracking epistemic violence, tracking practices of silencing. *Hypatia*, 26(2) :236–257.
- Flament, F. (2016). *La consolation*. JC Lattès.
- Fouquet*, A. (2010). L'usage des statistiques : de l'aide à la décision à l'évaluation des politiques publiques. *Revue française des affaires sociales*, (1) :307–322.
- Fraisse, G. (2008). À contre-temps. *Genre & histoire*, (2).
- Grisson, T., Julliard, V., Alié, F., and Ecrement, V. (2023). La modération abusive sur twitter : Étude de cas sur l'invisibilisation des contenus lgbt et tds en ligne. *Rezeaux*, (1) :119–149.

- Guillaumin, C. (1978). Pratique du pouvoir et idée de nature (1) l'appropriation des femmes. *Questions féministes*, pages 5–30.
- Halbwachs, M. (1994). Les cadres sociaux de la mémoire.
- Hanmer, J. (1977). Violence et contrôle social des femmes. *Questions féministes*, (1) :68–88.
- Henneron, L. (2005). Être jeune féministe aujourd'hui : les rapports de génération dans le mouvement féministe contemporain. *L'Homme et la société*, (4) :93–111.
- Husson, A.-C. (2016). Les hashtags militants, des mots-arguments. *Fragmentum*, (48) :105–127.
- Ingenito, L. (2019). Rendre justice aux victimes de violences à caractère sexuel : Étude avec des femmes immigrantes et des femmes racisées.
- Jaspard, M. (2001). Nommer et compter les violences envers les femmes : une première enquête nationale en France. *Population et sociétés*, (364).
- Jaspard, M. (2007). L'enquête nationale sur les violences envers les femmes en France (enveff) : Historique et contextes. *Violence envers les femmes. Trois pas en avant deux pas en arrière*, pages 25–39.
- Jaspard, M. (2011). Les violences contre les femmes, Paris, la découverte.
- Jouët, J., Niemeyer, K., and Pavard, B. (2017). Faire des vagues. *Réseaux*, 201(1) :21–57.
- Kelly, L. (1987). The continuum of sexual violence. *Women, violence and social control*, pages 46–60.

- Kouchner, C. (2021). *La familia grande*.
- Le Goaziou, V. (2011). *Le viol, aspects sociologiques d'un crime : une étude de viols jugés en cour d'assises*. La documentation française.
- Le Querrec, B. (2023). Modération des contenus en ligne : l'échec des réseaux sociaux centralisés. *Après-demain*, (2) :34–36.
- Lévy-Guillain, R., Sponton, A., and Wicky, L. (2022). L'intime au bout du fil. enjeux méthodologiques de l'entretien biographique à distance. *Revue française de sociologie*, (2) :311–332.
- Linet, T. and Nizard, J. (2004). Constats de violences sexuelles : rédaction d'un protocole d'accueil et mise en pratique. *Journal de gynécologie obstétrique et biologie de la reproduction*, 33(2) :99–109.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019). Roberta : A robustly optimized bert pretraining approach. *arXiv preprint arXiv :1907.11692*.
- Lochon, A. (2021). Trente ans de médiatisation des violences sexistes et sexuelles. *Émulations : Revue des jeunes chercheuses et chercheurs en sciences sociales*, (38) :21–p.
- Lopez, I., Quillivic, R., Evans, H., and Arriaga, R. I. (2019). Denouncing sexual violence : A cross-language and cross-cultural analysis of# metoo and# balancetonporc. In *Human-Computer Interaction–INTERACT 2019 : 17th IFIP TC 13 International Conference, Paphos, Cyprus, September 2–6, 2019, Proceedings, Part II 17*, pages 733–743. Springer.

- Louis, M.-V. (1994). *Le droit de cuissage : France, 1860-1930*. Editions de l'atelier.
- Martin, J.-C. (1996). Violences sexuelles, etude des archives pratiques de l'histoire. In *Annales. Histoire, Sciences Sociales*, volume 51, pages 643–661. Cambridge University Press.
- Martin, L., Muller, B., Suárez, P. J. O., Dupont, Y., Romary, L., de La Clergerie, É. V., Seddah, D., and Sagot, B. (2019). Camembert : a tasty french language model. *arXiv preprint arXiv :1911.03894*.
- Méadel, C. (2022). Balance ton porc ou la fabrique ordinaire du continuum.
- Mendes, K., Ringrose, J., and Keller, J. (2018). # metoo and the promise and pitfalls of challenging rape culture through digital feminist activism. *European Journal of Women's Studies*, 25(2) :236–246.
- Mossuz-Lavau, J. (2002). *Les lois de l'amour, 2e édition*. PayotRivages.
- Parratt, K. A. and Pina, A. (2017). From “real rape” to real justice : A systematic review of police officers' rape myth beliefs. *Aggression and Violent Behavior*, 34 :68–83.
- Paveau, M.-A. (2013). Technodiscursivités natives sur Twitter. Une écologie du discours numérique. *Epistémè : revue internationale de sciences humaines et sociales appliquées /*, 9 :139–176.
- Perona, O. (2022). La police du consentement : La qualification policière des récits de violences sexuelles. *Sociétés contemporaines*, 125(1) :147–173.
- Perrot, M. (2020). 1. pratiques de la mémoire féminine. *Champs-Histoire*, pages 47–58.

- Pollak, M. and Heinich, N. (1986). Le témoignage. *Actes de la recherche en sciences sociales*, 62(1) :3–29.
- Ricordeau*, G. (2004). Enquêter sur l’homosexualité et les violences sexuelles en détention. *Déviance et société*, 28(2) :233–253.
- Robert, P., Zauberman, R., Pottier, M.-L., and Lagrange, H. (1999). Mesurer le crime : entre statistiques de police et enquêtes de victimation (1985-1995). *Revue française de sociologie*, pages 255–294.
- Rouch, M. (2017). «vous ne me connaissez pas mais ne jetez pas tout de suite ma lettre». le courrier des lecteurs et lectrices de simone de beauvoir.
- Ruffio, C. (2020). De « faits divers » à fait de société, comment le viol est peu à peu devenu un sujet politique.
- Ruffio, C. (2022). Politiser le «crime de l’intime». ce que# metoo fait à la médiatisation du viol en france (2000-2020). *Les violences sexistes après# Me-Too*, pages 71–89.
- Salmon, A. (2017). Misères de la vie conjugale à la belle époque. un courrier des lecteurs du matin sur le divorce (1908). *Revue historique*, pages 619–650.
- Salmona, M. (2018). La mémoire traumatique : violences sexuelles et psychotrauma. *Les Cahiers de la justice*, (1) :69–87.
- Salmona, M. (2021). *Violences sexuelles-2e éd. : Les 40 questions-réponses incontournables*. Dunod.
- Salmona, M. (2022). *Le livre noir des violences sexuelles-3e éd.* Dunod.

- Savoie, L., Pelland, M.-A., Morin, S., Boudreau, L. C., and Grandisson, S. (2018). L'invisibilité de la violence sexuelle ordinaire chez les étudiantes universitaires : des expériences à comprendre. *Recherches féministes*, 31(2) :141–158.
- Schwartz, P. (2002). Women's studies, gender studies. le contexte américain. *Vingtième siècle*, (3) :15–20.
- Sipes, J. B., Roberts, L. D., and Mullan, B. (2022). Voice-only skype for use in researching sensitive topics : A research note. *Qualitative Research in Psychology*, 19(1) :204–220.
- Sohn, A.-M. (1981). Les rôles féminins dans la vie privée : approche méthodologique et bilan de recherches. *Revue d'histoire moderne et contemporaine (1954-)*, 28(4) :597–623.
- Souffrant, K.-A. (2023). *Le privilège de dénoncer : Justice pour toutes les victimes de violences sexuelles*. Éditions du remue-ménage.
- Springora, V. (2020). *Le consentement*. Grasset.
- Thomas, E. (1986). *Le viol du silence*. Editions Aubier.
- Trovato, N. (2022). La tentative de construction d'une contre-réalité des violences sexuelles en France : l'échec discursif du mouvement #MeToo? Master's thesis.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- Vernier, B. (2022). Témoigner pour instruire. la petite fille sur la banquise d'adélaïde bon. *Revue critique de fiction française contemporaine*, (24).

Vigarello, G. (1999). Georges vigarello, histoire du viol xvie-xxe siècle, paris, seuil, 1998, 357 p.

Wicky, L. (2020). Parcours d'hommes se déclarant victimes de violences sexuelles dans l'enfance. # 1257, (2) :88–91.

Zauberman, R. (2015). Les enquêtes de victimation : Une brève histoire, quelques usages. *Idées économiques et sociales*, (3) :8–21.