



## Les jeunes ont-il vraiment la haine ?

### Exploration des définitions et expériences du discours de haine en ligne chez les Belges âgés de 15 à 25 ans

Océane Gangi<sup>1</sup> et Cécile Mathys<sup>1,2</sup>

<sup>1</sup>Département de Criminologie de l'Université de Liège (Belgique)

<sup>2</sup>PhD

Contact : [oceane.gangi@uliege.be](mailto:oceane.gangi@uliege.be)

#### Résumé

Étant donné la forte exposition des jeunes aux violences en ligne, cette étude cherche à approfondir la compréhension du discours de haine en ligne en examinant les définitions et expériences individuelles tout en tenant compte de caractéristiques individuelles. Ainsi, vingt-quatre Belges de 15 à 25 ans confrontés à une expérience au moins de discours de haine en ligne et revêtant des profils variés en termes de genre, d'orientation sexuelle, d'origine culturelle et de statut auto-rapporté ont été rencontrés par le biais d'entretien qualitatifs. L'entièreté de ces répondants se considère à minima comme témoins de discours de haine en ligne (n=23) et également comme victimes (n=13), auteur (n=1) ou encore auteurs/victimes (n=2). L'analyse thématique a permis de mettre en évidence deux catégories principales et exclusives basées sur les définitions formulées et les expériences vécues : «les contenus à caractère haineux» et «les contenus à caractère agressif» ainsi que des spécificités basées sur les caractéristiques individuelles. Ces résultats sont discutés selon la perspective du discours de haine en ligne basé sur un continuum, débutant par les micro-agressions. Des implications sont proposées, que ce soit au niveau conceptuel, juridique ou encore concernant la modération au sein du numérique.

#### Mots clés

Discours de haine en ligne, adolescents, jeunes adultes, caractéristiques individuelles, micro-agressions

Do young people really feel hate? Exploring the definitions and experiences of online hate speech among Belgians aged 15 to 25

#### Abstract

Given young people's high exposure to online violence, this study aims to deepen the understanding of online hate speech by examining individual definitions and experiences while taking personal characteristics into account. Twenty-four Belgians aged 15 to 25, who have encountered at least one instance of online hate speech and who vary in terms of gender, sexual orientation, cultural background, and self-reported status, were interviewed using qualitative methods. All respondents identified themselves at a minimum as witnesses of online hate speech (n=23) and additionally as victims (n=13), perpetrator (n=1), or both perpetrators and victims (n=2). Thematic analysis revealed two main and mutually exclusive categories based on the definitions provided and experiences reported: 'hate-related content' and 'aggressive content,' along with specificities linked to individual characteristics. These findings are discussed within the framework of a continuum of online hate speech, beginning with micro-aggressions. Conceptual, legal, and digital moderation implications are also proposed.

#### Keywords

Online hate speech, adolescents, young adults, individual characteristics, microaggressions

## Introduction

La mise en lumière de phénomènes de violences en ligne chez les jeunes a permis tant de les rendre visibles que de jeter l'opprobre sur le numérique et ses dangers. En parallèle, le regain d'intérêt pour la lutte contre la discrimination a ravivé le débat autour du concept de discours de haine, y compris sa manifestation en ligne. Notre étude se situe à l'intersection de ces préoccupations, en explorant la compréhension du discours de haine en ligne d'un échantillon de jeunes belges, au départ de leurs définitions et expériences autour du discours de haine en ligne.

Étant donné son caractère numérique, le discours de haine en ligne est un phénomène qui traverse les frontières. Si le contenu peut différer en fonction de la culture et de la langue, il n'en reste pas moins présent sur l'ensemble des plateformes en ligne (Costello et Hawdon, 2020). Toutefois, la définition du discours de haine en ligne varie selon les disciplines et les auteurs affiliés. Il apert de multiples tentatives pour circonscrire le phénomène et mieux appréhender ce qui peut être considéré comme de la haine (Seoane et Monnier, 2019). Nous pouvons par exemple retrouver une qualification des conduites, provenant des sciences juridiques, où le discours de haine en ligne serait perçu en termes de continuum et de gravité de la conduite posée (Cramer *et al.*, 2022; Schewpke et Perry, 2022). Par ailleurs, d'autres auteurs vont mobiliser l'utilisation d'un langage (agressif ou offensant) comme modalité d'expression spécifique (Zhang et Luo, 2018) tandis que certains tendent à déplorer que le recours au numérique ne soit qu'un moyen de diffusion et non une caractéristique distinctive des violences en ligne et du discours de haine en ligne (Crespi et Hellsten, 2022).

Incidentement, les jeunes âgés de 15 à 25 ans<sup>1</sup> seraient reconnus comme utilisant massivement les réseaux sociaux et autres plateformes en ligne (Costello *et al.*, 2016). La présence en ligne de ces adultes émergents contribuerait à développer des aptitudes en ligne, notamment au niveau technique (réseaux sociaux, algorithmes), stimulerait la créativité (photos, vidéos) et également augmenterait les capacités sociales et relationnelles (interactions, communication) (Aranda Juárez *et al.*, 2020). Soulignons néanmoins que ces compétences sont nuancées par la littérature. En effet, une étude de 2023 a mis en évidence certaines lacunes des jeunes de 16 à 18 ans, notamment vis-à-vis de la production de publications, des paramétrages et de l'utilisation d'outils améliorant la confidentialité et l'anonymat (Estanyol *et al.*, 2023). Il y aurait ainsi un écart entre la perception qu'ont les jeunes de leurs compétences, voire la perception qu'en ont les adultes, et leurs compétences réelles. En outre, ce public est également reconnu comme étant particulièrement exposé aux violences en ligne (Costello et Hawdon, 2020). A cet égard, certaines caractéristiques individuelles sont identifiées comme étant liées aux victimisations en ligne telles que l'orientation sexuelle (Baider, 2019; Jane, 2018; Soral *et al.*, 2018), le genre (Döring et Mohseni, 2019; Jane, 2018; Venäläinen et Virkki, 2019), ou encore l'origine culturelle (Awan, 2014; Jane, 2018; Ortiz, 2021; Venäläinen et Virkki, 2019).

Dans ce contexte, la présente étude revêt un triple objectif. En premier lieu, nous souhaitons examiner les différentes définitions du discours de haine en ligne que formulent un public âgé de 15 à 25 ans, confronté à une expérience au moins de discours de haine en ligne (témoin, auteur et/ou victime). Ainsi, nous serons particulièrement attentifs aux modalités d'expression utilisées et aux effets perçus (ou interprétation des effets). Deuxièmement, poursuivre cette exploration des définitions en intégrant plusieurs variables individuelles que sont l'orientation sexuelle, le genre, l'origine culturelle et l'âge. Troisièmement, nous intéresser, à un sous-échantillon spécifique et directement ciblé par le discours de haine en ligne, c'est-à-dire les auteurs et/ou les victimes. Dans ce cadre, nous examinerons les expériences vécues en fonction des définitions préalablement formulées et également en fonction des caractéristiques individuelles étudiées.

## Aspects théoriques

### Qu'est-ce que le discours de haine en ligne ?

Le discours de haine en ligne fait partie des conduites de violences en ligne, que l'on peut définir comme « l'utilisation de systèmes informatiques pour causer, faciliter ou menacer de violence des individus » (Crespi et Hellsten, 2022, p. 392). Les violences en ligne revêtiraient plusieurs formes, qu'il s'agisse de contenus audios, vidéos ou encore textuels, et se produiraient selon des moyens variés, notamment par le biais de messageries instantanées et des réseaux sociaux. Leurs conséquences peuvent être d'ordre physique, sexuel, psychologique et/ou économique (Crespi et Hellsten, 2022). En outre, ces violences en ligne recouvriraient à la fois l'expression d'insultes, la propagation de rumeurs, la diffusion d'images, l'usurpation d'identité ou encore l'usage frauduleux d'informations personnelles (Willard, 2004). Ainsi, toute infraction peut comporter une modalité « en ligne » qui en modifie la nature et la portée (Harmer et Lumdsen, 2019). C'est le cas du discours de haine à travers le discours de haine en ligne.

Le discours de haine, qu'il soit présent dans l'espace numérique ou non, peut être défini de multiples façons. Pour certains, il s'agit de l'utilisation d'un langage (agressif ou offensant) visant un groupe spécifique de personnes partageant une caractéristique commune (Zhang et Luo, 2018). Pour d'autres, il s'agirait plus largement d'exprimer sa haine d'un collectif, avec comme objectif ou conséquence l'exclusion (Hawdon *et al.*, 2017). Monnier et ses collègues parlent, eux, de groupes-cibles qui « sont essentialisés et renvoyés à leur supposée extériorité à la norme sociale » (2021, p. 9). Une multitude de concepts apparentés sont par ailleurs mobilisés. Ainsi, Chetty et Alathur (2018) définissent le «discours de haine» comme des propos tenus par des personnes qui se perçoivent comme supérieures et qui visent des individus considérés par eux comme ignorants, un symbole, une activité ou encore un événement. De leur côté, Costello et ses collègues (2019) emploient le terme « contenus haineux » pour qualifier les

<sup>1</sup> Ces jeunes, nés entre 1997 et 2007, sont aussi appelés des digital natives et font écho aux personnes «utilisant internet et les ressources digitales dans l'apprentissage à l'école primaire» (Harrison *et al.*, 2020, p.61).

attitudes négatives dirigées contre des groupes en raison d'une caractéristique particulière. Jacks et Adler (2016) utilisent le terme de « crimes haineux », en y ajoutant la dimension d'une agression physique motivée par la haine envers une personne ou un groupe en raison de certaines caractéristiques.

Enfin, au-delà de ces modalités techniques et énonciatives d'expression, une série d'auteurs proposent d'analyser les comportements motivés par la haine selon un continuum et qualification juridique des conduites (Cramer *et al.*, 2022; Schweppe et Perry, 2022). Plus spécifiquement, certains auteurs et organisations parlent d'un système pyramidal de « gravité » qui partirait de conduites de préjugés (stéréotypes, remarques, blagues), pour continuer vers des actes individuels guidés par les préjugés (harcèlement, exclusion, déshumanisation) et aboutir potentiellement à de la discrimination (politique, économique...). Ce déroulé participerait enfin à une violence motivée par les préjugés (envers un individu ou un groupe), s'exprimant au sein de l'espace public, avec à l'extrême, le génocide (Anti-Defamation League, 2018). Au sein de celui-ci, le discours de haine se situerait après les micro-agressions et avant les crimes haineux, le terrorisme et le génocide (Cramer *et al.*, 2022; Schweppe et Perry, 2022). Schweppe et Perry (2022) utilisent même le terme « expression de la haine » comme un élément transversal au sein de ce continuum qui irait d'un commentaire futile à l'incitation à commettre un génocide.

Sur le plan juridique, en Belgique, contexte géographique de cette étude, la recommandation 97(20) du Comité des Ministres sur le discours de haine considère que « [l]e terme « discours de haine » doit être compris comme couvrant toutes formes d'expression qui propagent, incitent à, promeuvent ou justifient la haine raciale, la xénophobie, l'antisémitisme ou d'autres formes de haine fondées sur l'intolérance, y compris l'intolérance qui s'exprime sous forme de nationalisme agressif et d'ethnocentrisme, de discrimination et d'hostilité à l'encontre des minorités, des immigrés et des personnes issues de l'immigration ». Par ailleurs, le Conseil de l'Europe le définit comme : « tout type d'expression qui incite à, promeut, diffuse ou justifie la violence, la haine ou la discrimination à l'encontre d'une personne ou d'un groupe de personnes, ou qui les dénigre, en raison de leurs caractéristiques personnelles ou de leur statut réels ou attribués telles que la « race », la couleur, la langue, la religion, la nationalité, l'origine nationale ou ethnique, l'âge, le handicap, le sexe, l'identité de genre et l'orientation sexuelle ». Si nous nous attardons sur ces définitions et recommandations, nous constatons un manque de textes légaux, belge et européen, spécifiques au discours de haine en ligne. Ceci illustre les propos de Crespi et Hellsten (2022) selon lesquels les violences en ligne seraient plutôt considérées comme des comportements utilisant le numérique comme moyen ou technique, sans prendre suffisamment en compte leurs spécificités.

In fine, ces définitions laissent entrevoir la disparité de concepts apparentés au discours de haine en ligne et la difficulté de positionner clairement le discours de haine en ligne. Dans ce contexte, plusieurs questionnements émergent, dont ceux liés aux modalités d'expression du discours de haine en ligne : est-ce que l'évocation de la sphère numérique s'avérerait prédominante ? Serait-ce plutôt le caractère haineux des propos qui prévaudrait ? Ou encore, le type de conduite, s'apparentant aux violences

verbales, qui deviendrait central ? En outre, qu'en est-il des motivations et/ou conséquences sous-jacentes, telle l'exclusion ? Enfin, serions-nous en présence d'une interrelation de ces différentes considérations plutôt que face à une caractéristique spécifique ?

### **Comment est perçu le discours de haine en ligne parmi les 15 à 25 ans ? Et quelles caractéristiques individuelles distinctives ?**

La tranche d'âge se situant entre l'adolescence et le début de l'âge adulte semble particulièrement exposée (au sens large) au discours de haine en ligne (Bautista-Ortuño *et al.*, 2018; al Serhan et Elareshi, 2019). En effet, ce public possède deux caractéristiques qui le rendent spécifique : sa forte présence sur les réseaux sociaux et sa recherche d'identité (Costello et Hawdon, 2020; Keipi *et al.*, 2017). Ceci est confirmé par les résultats d'études quantitatives. Par exemple, une étude belge récente note que près d'un tiers des jeunes de 15 à 25 ans (échantillon de 2819 jeunes) aurait déjà été victime de discours de haine en ligne basés sur le genre, l'orientation sexuelle ou l'origine culturelle, et près d'un quart en aurait déjà été auteur. En outre, il apert que les contenus haineux cibleraient davantage l'origine culturelle (Walrave *et al.*, 2023). Ensuite, les questionnaires auto-rapportés d'étudiants universitaires jordaniens, âgés de 18 à 26 ans, ont démontré qu'ils auraient des facilités à identifier les discours de haine du fait de leur haute exposition à ceux-ci (al Serhan et Elareshi, 2019). Cette étude explique notamment que la majorité des personnes interrogées ont défini le discours de haine en ligne comme tout contenu susceptible de nuire à un groupe ou à un individu de ce groupe en raison de ses caractéristiques. En appui, une étude qualitative, menée auprès d'un échantillon de quatorze belges francophones âgés de 15 à 25 ans et confrontés à minima à une expérience de discours de haine, observe que ces jeunes se représentent le discours de haine en ligne comme un comportement qui cible prioritairement un individu, et non spécifiquement un groupe (Gangi *et al.*, 2023). Cette étude a également mis en évidence que les participants percevaient des interrelations entre le discours de haine en ligne et le cyberharcèlement et que l'une des modalités d'expression principales est ce que l'on pourrait appeler communément la violence verbale par l'usage des insultes notamment (Gangi *et al.*, 2023). Cette dernière étude n'examinait pas les caractéristiques individuelles des répondants en détails.

En ce qui concerne les expériences personnelles portant sur le discours de haine en ligne, plusieurs recherches quantitatives s'intéressant aux jeunes indiquent des corrélations entre le fait d'être témoin et auteur (Bernatzky *et al.*, 2022; Evans *et al.*, 2019) et le fait d'être victime et auteur (Costello et Hawdon, 2020; Wachs, Bilz, *et al.*, 2022). Ainsi, selon une étude de Costello et ses collègues (2019), les personnes victimisées en ligne seraient plus enclines à définir un discours comme blessant ou nocif, du fait de leur expérience passée.

D'autres études quantitatives mettent en évidence les caractéristiques individuelles comme marqueur central du discours de haine en ligne, notamment au travers du sentiment d'appartenance à un groupe (Bautista-Ortuño *et al.*, 2018). Dans ce contexte, le genre de la victime peut être identifié, les

jeunes femmes étant plus susceptibles d'être victimes (Döring et Mohseni, 2019; Jane, 2018; Venäläinen et Virkki, 2019) ; également l'orientation sexuelle, les personnes non hétérosexuelles seraient davantage ciblées (Baider, 2019; Jane, 2018; Soral et al., 2018), ou encore l'origine culturelle, les personnes issues de minorités culturelles seraient plus victimisées (Awan, 2014; Jane, 2018; Ortiz, 2021; Venäläinen et Virkki, 2019).

En conclusion, selon la littérature scientifique et les recherches centrées sur la jeunesse, le discours de haine en ligne serait tout contenu susceptible de nuire à un individu en raison des caractéristiques du groupe auquel il appartient. En outre, les études quantitatives réalisées mettent en évidence les caractéristiques individuelles alors que peu est détaillé quant aux motivations sous-jacentes (préjugé, intolérance, discrimination). Tel qu'exposé précédemment, ces motivations sont pourtant le socle des définitions juridiques (belge et européenne), participant à distinguer le discours de haine en ligne d'autres formes de violences en ligne.

Sur base de ces constats émanant d'une majorité d'études quantitatives et, face à la multitude de concepts apparentés au discours de haine en ligne, la présente étude, de nature qualitative, examine les trois questions principales suivantes :

1) Comment les jeunes âgés de 15 à 25 ans définissent-ils le discours de haine en ligne ? Grâce à la richesse du matériel qualitatif, nous supposons que les définitions pourront reposer sur un continuum plus large allant des insultes aux attaques visant des groupes/communautés (Cramer et al., 2022; Schweppe et Perry, 2022; al Serhan et Elareshi, 2019) plutôt que se restreindre aux catégories définies par un questionnaire. Ensuite, que ces définitions pourront varier selon les modalités d'expression et l'interprétation des effets (Gangi et al., 2023) qui seront ici plus facile à identifier par l'entremise des verbatims

2) Quelles variabilités individuelles apparaissent dans ces définitions (en termes de genre, orientation sexuelle, origine culturelle, âge) ? Il a ainsi été mis en évidence que les groupes plus fréquemment ciblés (femmes, jeunes non hétérosexuels, jeunes issus de minorités culturelles) avaient tendance à proposer une définition plus large du discours de haine en ligne (Awan, 2014; al Serhan et Elareshi, 2019 ; Baider, 2019; Costello et al., 2019; Döring et Mohseni, 2019). Dès lors, il semble aussi utile d'examiner si cette variabilité est également présente au sein de matériaux qualitatifs.

3) Comment, en particulier, les jeunes âgés de 15 à 25 ans auteurs et/ou victimes de discours de haine en ligne définissent le discours de haine en ligne ? Costello et al. (2019) ont pu montrer que les jeunes ayant des expériences de victimisation proposaient une définition assez large du discours de haine en ligne. De nouveau, l'ouverture offerte par un guide d'entretien permettra de vérifier si des définitions moins restrictives, ou non, sont exprimées.

## Méthodologie

Cette présente étude<sup>2</sup> porte sur un échantillon de vingt-quatre Belges (francophones et néerlandophones) âgés de 15 à 25 ans (moyenne d'âge de 19,83 ans et écart-type de 3,18) et confrontés à une expérience au moins de discours de haine en ligne. Les autres critères de sélection portaient sur la recherche de profils variés en termes de genre, d'orientation sexuelle, d'origine culturelle et de statut auto-rapporté en lien avec le discours de haine en ligne, comme mis en évidence par la littérature (Bautista-Ortuño et al., 2018; Costello et al., 2019). Cet échantillon est un échantillon de cas multiple (Yin, 2017), c'est-à-dire que la volonté est d'inclure une diversité de profils plutôt que d'atteindre une représentativité complète de la population des 15-25 ans. L'entièreté de ces répondants se considère comme témoins de discours de haine en ligne<sup>3</sup>. Parmi eux, treize se sont définis également comme victimes et trois comme auteurs, dont deux combinant ces deux statuts. Concernant les autres caractéristiques individuelles, l'échantillon est composé de dix personnes de sexe féminin et treize de sexe masculin. Treize définissent leur orientation sexuelle comme hétérosexuelle, trois comme homosexuelle, trois comme bisexuelle, une comme pansexuelle et trois comme « autres ». Enfin, dix personnes sont exclusivement d'origine<sup>4</sup> belge, une personne est d'origine belge et européenne (non belge), une personne est d'origine belge et africaine, quatre sont d'origine européenne (non belge) et sept sont d'origine africaine. Pour la suite des analyses, à l'instar de la variable genrée, nous avons choisi de dichotomiser certaines de nos variables telles l'orientation sexuelle (hétérosexuel, n=13 et non hétérosexuel, n=10), l'origine culturelle (exclusivement d'origine européenne, n=15 et non européenne, n=8) et l'âge (15-17ans ; n=9 et 18-25 ans, n=14)

Afin de parvenir à la diversité de profils attendue, les participants ont été recrutés grâce à des publications sur internet, ainsi que des affiches positionnées en des lieux stratégiques, notamment dans des établissements scolaires (école de promotion sociale, université et école secondaire). Il était stipulé clairement que le public visé était tout jeune ayant une expérience avec le discours de haine en ligne, sans préciser quel type d'expérience. Par ailleurs, le recours à des intermédiaires a permis d'accéder à un public plus sensible, comme des jeunes auteurs de faits qualifiés infractions en milieu de placement (sous garde) ou des jeunes pris en charge dans des centres d'hébergement relevant du secteur de l'aide à la jeunesse.

Les entrevues individuelles ont duré de 45 minutes à 1 heure 30. Tous les participants ont accepté l'enregistrement audio des échanges, ce qui a permis d'obtenir une retranscription fidèle des propos tenus. Les participants ont été informés de manière claire et détaillée des objectifs de l'étude, des conditions de participation et de leur droit de se retirer à tout moment, sans justification. Un consentement éclairé a été obtenu par écrit avant chaque entretien sur base d'un document rédigé et lu par la chercheuse principale, incluant tant un consentement

<sup>2</sup> Cette étude est l'un des axes de recherche poursuivis au sein du projet belge @ntidote par les universités de Liège, d'Anvers et de Saint-Louis (Bruxelles).

<sup>3</sup> L'une des répondantes n'a pas souhaité donner de définition du discours de haine en ligne, contenu pourtant attendu pour les analyses. Cette participante a été écartée de la présente étude, les résultats proviennent donc de vingt-trois répondants.

<sup>4</sup> L'origine des répondants dépend du lieu de naissance de leurs parents.

pour la participation effective à l'étude que l'autorisation de publier ensuite. Afin de garantir l'anonymat et la confidentialité, les entretiens ont été transcrits en supprimant ou modifiant toute information nominative ou contextuelle susceptible d'identifier un participant. Les données brutes sont conservées sur un support sécurisé, accessible uniquement aux chercheurs impliqués dans l'étude. Tous les éléments liés à la récolte des données ont été validés par le comité d'éthique de l'Université d'Anvers (Belgique). Le règlement de ce comité prévoit, entre autres, des précautions spécifiques concernant l'inclusion de mineurs (entre 14 et 18 ans). Les entrevues ont été menées grâce à un guide d'entretien rédigé au départ de la littérature sur le discours de haine en ligne ainsi que sur les violences en ligne (al Serhan et Elareshi, 2019; Bautista-Ortuño *et al.*, 2018; Castano-Pulgarín *et al.*, 2021; Costello et Hawdon, 2020; Keipi *et al.*, 2017; Meyer, 2010; Ortiz, 2021). Les questions principales portaient sur les définitions données du discours de haine en ligne et sur les expériences vécues (selon le statut auto-rapporté<sup>5</sup>). Le guide d'entretien a fait l'objet d'un pré-test avec des jeunes présentant des caractéristiques similaires quant à l'échantillon visé.

Une analyse thématique a été menée afin d'identifier les catégories de définition du discours de haine en ligne à partir des propos recueillis (Miles et Huberman, 2003). Après une première lecture intégrale des verbatim, des unités de sens ont été identifiées puis regroupées en catégories. Deux catégories ont émergé : les contenus à caractère agressif, correspondant à l'expression de l'agressivité envers une personne, et les contenus à caractère haineux, renvoyant à l'exclusion d'un représentant d'un groupe ou d'un groupe entier. Afin de guider la catégorisation, les critères examinés concernaient la cible (individu identifié ou groupe d'appartenance) et les motivations (volonté de blesser ou volonté

d'exclure). Dans certains cas, les éléments pouvaient relever des deux catégories simultanément, par exemple lorsqu'une insulte individuelle mobilisait une stigmatisation de groupe. Nous avons alors privilégié la classification en « contenu à caractère haineux ». Ce choix repose sur le fait que les répondants, en intégrant ces éléments, montrent qu'ils prennent en compte les dynamiques sociales plus larges, où un groupe ou un représentant de groupe est spécifiquement visé et où il existe une volonté d'exclure ou de stigmatiser une communauté. En ce sens, l'agressivité envers l'individu existe au travers du caractère haineux. En outre, en lien avec nos questions de recherche (notamment la première), deux critères d'analyse ont été ajoutés et utilisés à priori pour examiner les définitions du discours de haine en ligne, à savoir : les modalités d'expression (type de conduites, actions spécifiques, évocation du discours, présence du numérique) et l'interprétation des effets (conséquences réelles ou supposées).

Par ailleurs, pour rendre compte de la diversité des vécus, dix-sept expériences issues de notre sous-échantillon de quatorze jeunes auteurs et/ou victimes ont été retenues, allant au-delà des seules définitions proposées.

## Résultats

### Les définitions du discours de haine en ligne formulées par les participants (témoins, auteurs et/ou victimes) et les liens avec leurs caractéristiques individuelles

Tout d'abord, au sein de l'échantillon au complet (n=23), le tableau 1 met en exergue les éléments qui semblent rassembler

Catégories de définitions	Eléments de définition	Mots ou groupes de mots présents dans la définition des participants	Participants
Contenu à caractère agressif	Modalité d'expression du discours (type de conduites, actions spécifiques, évocation du discours, présence du numérique)	Acte conscient, réfléchi ou impulsif ; Agressivité ; Blagues innocentes ; Contenus qui continuent dans l' <i>offline</i> ; Cyberviolence ; Discours préjudiciable (Forme de) Harcèlement ; Insultes ; Opposé de la liberté d'expression ; Pas joli ; Quelque chose qui arrive plus sur les réseaux sociaux ; Sous-entendus	<b>n=13</b> B ; E ; J ; L ; N ; P ; Q ; R ; S ; T ; U ; V ; W 8 non hétérosexuels 5 hétérosexuels 7 de genre féminin 6 de genre masculin 4 d'origine non européenne 9 d'origine européenne 9 de 18 à 25 ans 4 de 15 à 17 ans
	Interprétation des effets (conséquences réelles ou supposées)	Blesser quelqu'un ; Se sentir moins en sécurité	
Contenu à caractère haineux	Modalité d'expression (type de conduites, actions spécifiques, évocation du discours, présence du numérique)	Avoir de la haine et l'afficher ; Discours qui incite à la haine ; Discours avec du racisme ou contre le genre ; Discrimination ; Haine pure ; Non-acceptation Ignorance ; Pensées clairement haineuses contre quelqu'un	<b>n=10</b> A ; C ; D ; F ; G ; H ; I ; K ; M ; O 2 non hétérosexuels 8 hétérosexuels 3 de genre féminin 7 de genre masculin 4 d'origine non européenne 6 d'origine européenne 5 de 18 à 25 ans 5 de 15 à 17 ans
	Interprétation des effets (conséquences réelles ou supposées)	Inciter à la violence, physique ou verbale ; Exclure des gens qui sont dans une minorité ; Exprimer de la haine ; Essayer de convaincre les gens avec de la haine ; Appel au racisme	

Tableau 1 : Définitions du discours de haine en ligne par l'ensemble des participants (participants témoins, auteurs et/ou victimes)

<sup>5</sup> Par exemple, « Après notre discussion concernant tes expériences en ligne, dans quel(s) rôle(s) te reconnaîtraitas-tu le plus ? Une victime ? Un auteur ? Un témoin ? Une combinaison de plusieurs rôles ? »

les différentes définitions du discours de haine en ligne.

Ainsi, au sein de la catégorie «contenus à caractère agressif» et relatif aux modalités d'expression, nous pouvons relever différentes conduites telles que de l'agressivité, du harcèlement et de la cyberviolence : «*Le discours de haine en ligne et le harcèlement, c'est la même chose parce qu'en fait le harcèlement affecte Internet mais il affecte aussi les gens en face à face, donc cela regroupe simplement différentes possibilités de harcèlement*» (L). Également des actions spécifiques comme insulter, blaguer ou faire des sous-entendus. L'usage des mots et le recours au discours sont aussi très présents pour définir le discours de haine en ligne tel que «*[c'est] un discours qui incite à la haine et qui va souvent va dans des propos injurieux*» (H). En outre, des qualificatifs sont mentionnés pour appuyer le caractère déplaisant de la conduite «*pas joli*», «*préjudiciable*», «*opposé à la liberté d'expression*» (P). Par ailleurs, le contexte d'expression est également évoqué, que celui-ci soit dissimulé à l'instar de «*Cela peut se faire à travers des insultes pour moi, et aussi de manière plus cachée*» (J), ou que le numérique soit spécifiquement nommé. Dans ce cas, le numérique peut constituer un terrain qui favorise la conduite «*Quelque chose qui n'arrive pas souvent hors ligne mais plutôt sur les réseaux sociaux*» (R). Les effets des discours de haine sont également discutés : «*Un discours de haine, c'est dire quelque chose qui ne va pas plaire à la personne dans le but de lui faire du mal, c'est tout*» (W). Ainsi, au sein de cette catégorie, nous identifions un large panel de conduites et d'actions citées en tant que modalités d'expression du discours de haine en ligne. En effet, les répondants identifient tant des blagues, des remarques, des insultes que des conduites de harcèlement. Cependant, peu de conséquences réelles ou perçues sont identifiées, si ce ne sont les mentions de se sentir moins en sécurité et le fait de blesser l'autre, qui semblent concerner plus spécifiquement la sphère interpersonnelle proximale (soi ou un autre).

En parallèle, la catégorie «contenu à caractère haineux» comporte un éventail plus restreint de modalités d'expression s'ancrant au départ de la conduite de discrimination, couplé à des d'action spécifiques incluant le fait d'ignorer, de ne pas accepter ou encore de ressentir de la haine. Les discours de haine «*incluent la haine pure, le refus d'accepter une autre personne, la discrimination, ou même l'ignorance*» (O). Ces modalités d'expression font référence de façon plus explicite au discours ainsi qu'à l'expression de la haine : «*ce sont plutôt les mots, donc tout ce que l'on peut voir dans les commentaires*» (L). Incidemment, les effets perçus, réels ou supposés, s'assortissent d'exclusion, d'incitation à la violence, d'expression de la haine, d'une volonté de convaincre et d'un appel au racisme telle ces définitions : «*C'est une incitation à la violence, qu'elle soit verbale ou physique, en relation avec des propos irrespectueux, ou des remarques à propos d'une personne, pour inciter à la violence en rapport avec la race*» (I); «*C'est des gens qui par exemple veulent exprimer leur haine et essayer de convaincre les gens que ce qu'il dit lui, avec de [la] haine, en étant énervé, que c'est juste, que c'est vrai*» (E); «*Je viens de trouver le mot... Exclure, c'est ça, exclure les gens en fait, les personnes qui sont en minorité, que ce soit culturellement ou en termes d'orientation sexuelle aussi*» (M). Dans cette catégorie de définition, les modalités d'expression semblent donc plus ciblées, au départ de la conduite de discrimination, et intègrent plus clairement la portée dirigée des actions envers un tiers qui prendrait la forme d'un collectif ou d'une communauté.

Relatif aux caractéristiques individuelles, les participants présentant

une orientation non hétérosexuelle, un genre féminin et étant âgé entre 18-25 ans sont ceux qui ont davantage formulé une définition appartenant à la catégorie «contenu à caractère agressif». Relevons également que pour le genre et l'orientation sexuelle, leurs homologues (masculin et/ou hétérosexuel) se retrouvent majoritairement dans le «contenu à caractère haineux». L'origine culturelle est la seule variable où les répondants semblent se répartir assez équitablement entre les deux catégories de discours de haine en ligne identifiées, que ce soit pour les jeunes d'origine européenne que pour les jeunes d'origine non européenne.

### **Les expériences de discours de haine en ligne vécues par les participants auteurs et/ou victimes**

Au sein de ce sous-échantillon spécifique (n=14), les quatorze jeunes ont au moins mentionné une expérience relevant du «contenu à caractère agressif». Parmi eux et comme illustré au sein du tableau 2, douze ont exclusivement mentionné une expérience de «contenu à caractère agressif», et deux ont mentionné aussi au moins une expérience de «contenu à caractère haineux». Ce sous-échantillon de quatorze jeunes a ainsi mis en évidence dix-sept expériences comme auteur ou comme victime.

Parmi les expériences de «contenu à caractère agressif», nous retrouvons douze personnes se qualifiant de victime et deux d'auteur. Il est intéressant de noter que les douze participants ayant partagé ce type d'expériences comme victimes ont tous proposé en amont une définition relevant de la même catégorie. Par exemple, un jeune homosexuel ayant défini le discours de haine en ligne comme des insultes a confié recevoir régulièrement des insultes homophobes, souvent justifiées par l'humour. Nous entrevoyons déjà une certaine homogénéité entre les définitions formulées d'une part et, d'autre part, les expériences vécues de discours de haine en ligne, ce en lien avec des «contenu à caractère agressif». Au contraire, les deux personnes s'identifiant comme auteurs d'une expérience que nous avons qualifiée de «contenu à caractère agressif» avaient proposé une définition relevant de la catégorie «contenu à caractère haineux». Ainsi, l'un de ces jeunes a raconté quant à son expérience en tant qu'auteur : «*Je suis dans mon lit, je critique... ça m'arrive d'écrire « sale PD », c'est normal madame. Ça a jamais existé un être humain qui sort avec... Un homme qui sort avec un homme, une femme qui sort avec une femme. Un qui veut se faire transformer ses parties génitales d'une femme, un qui veut devenir enceinte. Il y a aucun sens, madame*» (F).

Parmi les expériences relevant du «contenu à caractère haineux», nous identifions deux personnes, l'une comme victime et auteure et l'autre comme victime. Ce dernier raconte une expérience en tant que victime de «discrimination contre ses origines maghrébines» (C), notamment pour trouver un appartement au départ de recherches en ligne. Il est intéressant de noter que ce même jeune a aussi proposé une expérience en lien avec un «contenu à caractère agressif» en tant qu'auteur : «*j'ai déjà partagé des propos qui disait explicitement que les gays étaient très souvent des pédales (efféminés), ou en tout cas, que c'est comme ça qu'on veut les catégoriser dans les médias. Pourquoi je fais ça ? Parce que je considère que c'est pas sain pour la société, et j'ai vraiment peur pour mes éventuels*

futurs enfants, alors même si ça sert à rien, j'ai besoin de cracher ce que j'ai dans le ventre, de m'exprimer, pour faire un minimum barrière. Alors évidemment, il m'arrive d'insulter quand on me répond méchamment, c'est typique de cette communauté, mais la plupart du temps, j'essaye d'argumenter avec eux» (C). Nous pouvons ajouter en outre que ce jeune a proposé une définition initiale associée à la catégorie «contenu à caractère haineux».

L'autre jeune relevant du «contenu à caractère haineux» nous a confié deux expériences marquantes pour elle. En tant qu'auteure de discours de haine en ligne, la répondante explique qu'elle a pris part à un débat entre une personnalité et «des fachos». Elle a ainsi relayé des propos «qui peuvent être considérés comme de la haine anti-blanc», par exemple «[les blancs, faudrait les gazer parce que parfois ils font des trucs vraiment trop abusé» (J). En tant que victime de discours de haine en ligne, elle explique : «C'était un mec de 30 balais, quelque chose comme ça, et ça avait monté jusqu'à ce qu'il me dise que moi je devais me faire exciser pour comprendre ce que c'est les vraies mutilations génitales et donc là j'ai compris que ouais non c'était vraiment parti loin et c'est le plus hardcore que j'ai eu je pense» (J). Il est à noter que cette même jeune fille a proposé une troisième expérience, relevant elle du «contenu à caractère agressif» : «Sinon, c'était des trucs assez banals en soi sur le fait que je défende les personnes LGBT, que je sois parfois un peu incisive avec les entre guillemets mâles dominants, des choses comme ça, j'ai reçu quelques insultes là-dessus, j'ai déjà entendu 's salope', 'connasse', des choses comme ça mais j'ai plus les phrases exactes en tête, ça commence à dater maintenant» (J). Notons que cette jeune a proposé une définition initiale entrant dans la catégorie «contenu à caractère agressif».

Ces matériaux relatifs aux expériences des auteurs et/ou des victimes du discours de haine en ligne indiquent une tendance marquée d'expériences vécues associées à la catégorie des «contenus à caractère agressif». Ces expériences mettent en exergue des conduites de harcèlement et des actions d'insultes, de moqueries reçues, de chantage ou encore la formulation de commentaires négatifs et de critiques. Par ailleurs, ces expériences sont, en majorité, reliées aux définitions initiales exprimées, étant davantage à caractère agressif. Ceci est en particulier observé pour les jeunes qui se disent victime de discours de haine en ligne. Si nous regardons du côté des deux personnes s'attribuant le statut d'auteur (C ; F), nous relevons que les expériences mentionnées sont assez similaires, ces participants évoquant la commission d'insultes et de critiques à l'encontre d'un individu ciblé. Toutefois, les définitions données dans un premier temps s'apparentaient plutôt à des contenus à caractère haineux touchant à la discrimination d'une communauté et visant l'exclusion de ses membres. Il y a donc ici, pour ces deux auteurs, une tension entre, d'une part, la façon d'envisager le discours de haine en ligne, comme étant une conduite discriminante et, d'autre part, le vécu d'expériences qui s'exprime par des conduites ciblées d'agressivité verbale au sein du numérique. Nous retrouvons également cette tension pour une personne (J) décrivant des expériences de discours de haine en ligne, en tant qu'auteur et victime, s'apparentant à des contenus à caractère haineux, alors que la définition initiale formulée est davantage tournée vers des contenus à caractère agressif. Il est à noter que l'un de ces participants (C) rapporte aussi des comportements de victimisation entrant dans la catégorie des contenus à caractère haineux, créant alors une relative homogénéité avec la définition initiale donnée.

#### Sous-échantillon (n=14)

A ; B ; C ; E ; F ; G ; H ; I ; J ; L ; M ; N ; R ; W

6 non hétérosexuels et 8 hétérosexuels ; 7 de genre féminin et 7 de genre masculin ; 7 d'origine non européenne et 7 d'origine européenne ; 8 de 18 à 25 ans et 6 de 15 à 17 ans

Catégories de définitions	Statut auto-rapporté	Mots ou groupes de mots décrivant l'expérience des participants	Définitions initiales	Participants
Contenu à caractère agressif	Victime (n=12) <b>A ; B ; E ; G ; H ; I ; J ; L ; M ; N ; R ; W</b>	Insultes ; Mauvaises expériences ; Attaques ; Harcèlement ; Chantage ; Moqueries ; Commentaires négatifs ; Critiques ;	Contenu à caractère agressif	<b>n=14</b> 6 non hétérosexuels 8 hétérosexuels 7 de genre féminin 7 de genre masculin 7 d'origine non européenne 7 d'origine européenne 8 de 18 à 25 ans 6 de 15 à 17 ans
Contenu à caractère haineux	Auteur (n=2) <b>C ; F</b>	Insultes ; Critiques	Contenu à caractère haineux	
	Victime (n=1) <b>C</b>	Discrimination	Contenu à caractère haineux	<b>n=2</b> 1 non hétérosexuel 1 hétérosexuel 1 de genre féminin 1 de genre masculin
	Victime (n=1) <b>J</b>	Appel à la violence/ au meurtre	Contenu à caractère agressif	1 d'origine non européenne 1 d'origine européenne 2 de 18 à 25 ans 0 de 15 à 17 ans
	Auteur (n=1) <b>J</b>	Appel à la mutilation		

Tableau 2 : Expériences de discours de haine en ligne par un sous-échantillon (participants auteurs et/ou victimes)

Relatif aux caractéristiques individuelles, nous pouvons identifier un étonnant équilibre en termes d'orientation sexuelle, de genre, d'origine culturelle et d'âge, que ce soit au sein du sous-échantillon en tant que tel, mais aussi spécifiquement parmi les deux groupes auxquels se rapportent les expériences de « contenu à caractère agressif » et « contenu à caractère haineux ». Précisons aussi que les deux participants décrivant des expériences à caractère haineux se situent uniquement entre 18 et 25 ans. En ce qui concerne le statut auto-rapporté, nous pouvons mettre en évidence la présence d'auteurs et de victimes au sein des deux catégories, avec une grande majorité de victimes relatant une expérience de « contenus à caractère agressif ». Il est également intéressant de noter qu'une caractéristique individuelle émerge au sein des participants se présentant comme « auteur et/ou victime », à savoir l'origine culturelle. Ainsi, il apparaît que parmi les huit répondants d'origine non-européenne, sept d'entre eux se retrouvent au sein de ce sous-échantillon ciblé.

## Discussion des résultats de recherche

Nos résultats de recherche, issus de matériaux qualitatifs, ont permis de mettre en exergue non pas un continuum clair, mais deux catégories distinctes de définitions et d'expériences vécues du discours de haine en ligne au sein d'un public de 15 à 25 ans confronté à une expérience à minima de discours de haine en ligne.

Premièrement, la catégorie la plus représentée, le « contenu à caractère agressif », qui est l'expression de l'agressivité envers une personne (formulation large). Deuxièmement, le « contenu à caractère haineux », qui vise l'exclusion d'un représentant d'un groupe ou du groupe entier (formulation plus restrictive). En outre, nous avons pu examiner deux éléments de définition que sont les modalités d'expression et l'interprétation des effets.

Ensuite, nous avons analysé les tendances des définitions partagées par nos participants en fonction de caractéristiques individuelles précitées. Il est intéressant d'identifier que seule la variable liée à l'origine culturelle converge vers des tendances plutôt similaires parmi les jeunes d'origine européenne et non européenne entre les deux catégories identifiées. Au sein des autres variables, l'orientation non hétérosexuelle, le genre féminin et la tranche d'âge 18-25 ans sont les groupes qui ont davantage formulé une définition appartenant à la catégorie « contenu à caractère agressif ». Par ailleurs, leurs homologues masculins et/ou hétérosexuels ont plutôt formulé des définitions associées au « contenu à caractère haineux ». Enfin, et en adéquation avec les définitions partagées, nous constatons une forte présence des expériences liées à la catégorie « contenu à caractère agressif » plutôt que celle « contenu à caractère haineux », ce au sein de notre sous-échantillon spécifique de victimes et/ou d'auteur de discours de haine en ligne. Pour finir, nous pouvons noter une présence forte des jeunes répondants non européens au sein de notre sous-échantillon auteur et/ou victime.

Dans la suite de cette section, nous allons revenir sur les matériaux récoltés en discutant premièrement des définitions formulées par les participants rencontrés et en cherchant à rendre compte des nuances exprimées. Ensuite, nous nous attarderons sur les profils de nos participants pour tenter de comprendre les définitions et expériences vécues. Enfin, nous clôturerons en

identifiant les forces et limites de cette étude, ainsi qu'une série d'implications à visée scientifique et appliquée.

## Tentative de compréhension du discours de haine en ligne : des micro-agressions jusqu'à l'exclusion ?

Nos matériaux nous questionnent premièrement quant aux définitions proposées, notamment sur les objectifs potentiellement poursuivis par le discours de haine en ligne (que nous avons nommé «interprétation des effets»). Ainsi, Hawdon et ses collègues (2017) présentent le discours de haine en ligne comme l'expression de la haine d'un collectif, avec l'exclusion du groupe comme objectif ou conséquence. En ce sens, seule la catégorie de définitions «contenu à caractère haineux» appuie la volonté d'exclusion du groupe comme conséquence réelle ou supposée alors que la catégorie «contenu à caractère agressif» renseigne davantage sur la sphère interpersonnelle.

Si l'on s'attache à la modalité discursive du terme «discours de haine», il est intéressant de préciser que le terme « *hate speech* », largement mobilisé dans la littérature, fait référence, en français, à trois facettes distinctes : le « discours haineux » comme porteur de haine, le discours « provoquant la haine » en tant qu'incitation à la violence, ainsi que le « discours de haine » issu d'un sentiment (présumé de haine) (Seoane et Monnier, 2019). Cette tendance reflète la volonté croissante, observée chez les auteurs anglophones ces dernières années, de mieux comprendre et affiner la signification de ce terme. Ainsi, Sponholz (2017) a identifié et différencié le discours haineux (*hateful speech*), le discours incitant à la haine (*hate-fomenting speech*) et le discours dangereux (*dangerous speech*). Pour lui, le discours haineux serait prioritairement émotionnel et basé sur l'utilisation d'un langage désobligeant (comme des insultes), réalisé par des personnes avec peu d'expériences dans la sphère publique (Gagliardone *et al.*, 2015, cités dans Sponholz, 2017). L'objectif serait alors d'exprimer ses émotions. Le discours incitant à la haine serait, lui, davantage rationnel et intentionnel (volonté de blesser ou d'obtenir de l'attention), sous-tendu par une argumentation, et serait plutôt communiqué à un public qui partage les griefs contre ce groupe (par exemple, au sein de groupes en ligne). Alors que ces deux comportements provoqueraient une violence symbolique, le discours dangereux serait lui un appel à de la violence physique au sein de l'espace public (Sponholz, 2017). Les réflexions autour des catégorisations émanant de nos résultats nous ont conduit à rapprocher notre catégorie «contenu à caractère agressif» du «discours haineux» et du «discours incitant à la haine» de Sponholz (2017), en précisant que tant les définitions que les expériences relèvent davantage du «discours haineux», notamment au vu du manque d'argumentation. Cependant, l'un de nos répondants a pris le temps de mettre en évidence le caractère intentionnel et rationnel des propos insultants qu'il partage aux personnes homosexuelles.

Notre catégorie «contenu à caractère haineux» serait, elle, principalement rattachée au «discours incitant à la haine» et au «discours dangereux». Mentionnons que si les définitions se rapportent davantage au «discours incitant à la haine», une majorité des exemples évoqués partie de ce «discours dangereux».

En résumant, les expériences vécues par les participants seraient presque exclusivement d'ordre émotionnel puisque issues de la catégorie « contenu à caractère agressif » (au sein de notre étude) ou « discours haineux » (selon Sponholz). Le « discours incitant à la haine » serait par ailleurs connu de nos répondants, car présent au sein des deux définitions formulées, mais peu expérimenté, in fine, par le sous-échantillon ciblé que sont les auteurs et/ou les victimes (si ce n'est deux expériences rapportées). Ainsi, nous pouvons nous représenter le « discours incitant à la haine » comme à l'interstice du discours plus émotionnel et de l'appel à la violence. En allant plus loin, nous pouvons faire l'hypothèse que cet interstice pourrait être une sorte de tremplin amenant à la violence. De manière plus concrète, il est pertinent d'examiner comment les émotions négatives telles que la colère ou la frustration, exprimées dans les « discours haineux » peuvent alimenter un discours devenant plus rationnel et argumenté, celui-ci incarnant alors une forme d'appel à la violence. Cette évolution peut résulter d'une façon de donner sens aux émotions initiales. En d'autres termes, des ressentis intenses, tels que la colère ou la frustration, peuvent se transformer en arguments justifiant des comportements hostiles. Comme le soulignent Smith et Mackie (2015), ces émotions ont le pouvoir de forger un narratif qui unit un groupe contre un autre, consolidant ainsi des identités collectives fondées sur l'exclusion. Leur théorie des émotions intergroupes met en lumière le fait que les émotions ne se limitent pas à des expériences individuelles ; elles peuvent également émerger de l'appartenance à des groupes sociaux, engendrant des réactions émotionnelles face à des événements qui impactent ces groupes, même lorsque l'individu n'est pas directement affecté (Smith et Mackie, 2015).

Par conséquent, nous nous interrogeons sur les liens entre les deux catégories proposées, « contenu à caractère agressif » et « contenu à caractère haineux ». Au départ, nous avions imaginé retrouver un continuum large et indistinct de conduites de haine en ligne (Cramer *et al.*, 2022; Schweppe et Perry, 2022). La richesse des matériaux qualitatifs nous a offert la possibilité de dépasser la qualification juridique et d'identifier deux catégories précises. Aussi, si nous nous situons dans une perspective criminologique, la compréhension du discours de haine en ligne peut être facilitée par celle des micro-agressions (Clark *et al.*, 2011; Sue *et al.*, 2008). En effet, les micro-agressions sont tous les comportements et langages quotidiens intentionnels et non intentionnels subis par les personnes minoritaires et/ou opprimées (Clark *et al.*, 2011; Sue *et al.*, 2008). Dans ce contexte, Sue et ses collègues (2008) identifient trois sortes de micro-agressions : les micro-injures (intentionnelles), les micro-insultes (non-intentionnelles) et les micro-invalidations (banalisation des expériences vécues). D'un point de vue juridique, les micro-agressions peuvent parfois être prises en compte par la loi, selon le préjudice et le contexte (par exemple, le lieu de travail), mais sont souvent considérées comme non infractionnelles (n'entrant pas dans les lois anti-discrimination). Les micro-agressions vont ainsi participer à un environnement propice aux préjugés (Schweppe et Perry, 2022) et certains groupements internationaux demandent qu'elles soient criminalisées (Schweppe et Perry, 2022). D'ailleurs, en ce qui concerne plus spécifiquement les jeunes, plusieurs auteurs indiquent qu'une approbation tacite de ces micro-agressions peut mener à une intensification du langage ou des comportements agressifs (Hall, 2009; Wieland, 2007).

Sur cette base, il semblerait que nos participants de la catégorie « contenu à caractère agressif » intègrent, au sein de leur définition du discours de haine en ligne, la notion de micro-agressions, de préjugés et des actes individuels guidés par les préjugés. Des exemples tels que formuler des insultes, des blagues, des remarques et des conduites de harcèlement au sein de la sphère interpersonnelle et envers un individu ont ainsi pu être cités. La catégorie « contenu à caractère haineux » correspondrait à la fois aux éléments de qualification relatifs à de la discrimination, mais aussi à l'exclusion et à la déshumanisation présentes dans les actes individuels guidés par les préjugés. Les propos de nos participants relatifs à cette catégorie sont à la fois plus précis dans leurs modalités d'expression (expression de la haine, usage du discours, ...) et les intentions perçues (exclusion, incitation à la violence, ...) et en même temps se concentreraient sur une cible plus large, un collectif ou une communauté et non un individu isolé.

Ainsi, en vue d'enrichir le continuum existant, les micro-agressions incluant les commentaires, les remarques, l'agressivité verbale, les blagues, les insultes ou encore le harcèlement dans un contexte plus restreint, touchant un individu, peuvent constituer un contexte propice pour un glissement vers un niveau supérieur de gravité renvoyant à la discrimination, l'exclusion, l'incitation à la violence voire l'appel au racisme envers un groupe (Schweppe et Perry, 2022). En outre, et de façon rétroactive, l'existence de comportements exprimant des « contenus à caractère haineux » pourraient autoriser et légitimer des comportements relatifs au « contenu à caractère agressif », soutenu par un processus de normalisation de ces conduites (Costello et Hawdon, 2020). Dès lors, plus qu'un continuum linéaire, les interrelations entre les différentes formes de discours de haine en ligne pourraient davantage s'intégrer dans une boucle, non loin sans rappeler un cercle vicieux qui s'installera.

Enfin, alors que ces actes n'auraient pas, pour tous, une visée discriminatoire, l'objet relatif à l'identité de la victime, individualisée, couplé à leur fréquence et à leur caractère public peuvent possiblement créer un sentiment d'exclusion et de discrimination, ce indépendamment de la volonté de l'auteur. En effet, les matériaux de recherche associés à la catégorie de définitions « contenu à caractère agressif » démontrent que la motivation de nuire, en excluant la personne ciblée, n'est pas forcément présente. Comme discuté, la volonté perçue serait plutôt associée à une forme de régulation émotionnelle chez l'auteur (de type exutoire) (Sponholz, 2017). Certains de nos répondants ont en effet souligné le caractère humoristique de certains propos. Aussi, il apparaît que l'humour peut exercer un rôle, en servant non seulement de mécanisme de régulation émotionnelle mais aussi de moyen de communication relationnelle au sein de cette catégorie davantage associée à la sphère interpersonnelle. Il serait ainsi montré que l'une des principales fonctions de l'humour serait de se reconnaître entre membres d'un même groupe social en fonction des propos qui font rire et ceux qui ne font pas rire (Argüello-Gutiérrez *et al.*, 2018). En outre, selon le modèle des fonctions sociales de l'humour proposé par Martineau en 1972 (encore largement mobilisé par les recherches récentes), l'humour revêt deux principaux objectifs contradictoires : l'humour pourrait servir de « lubrifiant » social, facilitant la cohésion et

l'interaction entre les membres d'un groupe, ou de matière «abrasive», entraînant des tensions et des conflits, notamment lorsque des propos dénigrants sont employés. Cette dualité souligne la complexité de l'humour dans les relations interpersonnelles et son potentiel à renforcer les liens sociaux tout en créant des divisions.

Alors que nos résultats semblent en partie s'éloigner des définitions classiques du discours de haine en ligne (voir Hawdon *et al.*, 2017), ils confirment en réalité le caractère complexe et graduel de ces discours, voire rétroactif, tributaire, peut-être, du contexte dans lequel ils s'expriment (prolongement numérique de la sphère interpersonnelle «proximale» ou espace public numérique plus vaste) ainsi que la séparation très fine entre des conduites d'apparence semblables mais pourtant légalement différentes.

### **Tendances relatives aux définitions et expériences du discours de haine en ligne selon le profil des participants et leurs caractéristiques individuelles**

Nos matériaux indiquent que tant les personnes non hétérosexuelles que les personnes de genre féminin, tout comme les participants plus âgés (18-25 ans) ont davantage tendance à décrire le discours de haine en ligne avec une formulation que nous qualifions de plus large, associée aux «contenus à caractère agressif». Nous pouvons ainsi faire les liens avec nos deuxième et troisième questions de recherche et tenter de comprendre pourquoi pour certains participants spécifiques, le fait d'être régulièrement sujets à de la victimisation en ligne (Costello et Hawdon, 2020) concourt à leur manière de définir le discours de haine en ligne. Ainsi, comme dit précédemment, selon une étude quantitative de Costello et ses collègues (2019), les personnes victimisées en ligne seraient plus enclines à définir un discours comme blessant ou nocif, du fait de leur expérience passée. Au sein de notre étude, cette victimisation pourrait ainsi être un élément qui expliquerait le choix d'une formulation plus large du discours de haine en ligne, considérant aussi les micro-agressions comme du discours de haine en ligne et finalement renforçant une forme d'amplification du phénomène, intégrant le contexte interpersonnel plus restreint, par exemple une personne visée, au discours de haine en ligne. Par ailleurs, nous relevons que les participants de sexe masculin tout comme ceux se présentant comme hétérosexuel ont davantage tendance à rapporter des définitions associées à la catégorie «contenu à caractère haineux». Dans la littérature, nous constatons que ces profils sont davantage reliés au statut d'auteur de discours de haine en ligne (Bernatzky *et al.*, 2022). Au sein de notre étude, nous remarquons également ces tendances, deux des trois auteurs étant des hommes hétérosexuels. En ce qui concerne l'origine culturelle, nous remarquons que les répondants d'origine non européenne se répartissent équitablement entre les deux catégories quant aux définitions formulées, alors que ceux d'origine européenne ont proposé majoritairement une définition relative aux «contenus à caractère agressif». Toutefois, quand nous regardons les expériences vécues par les répondants d'origine non européenne, nous observons que sept répondants sur huit ont rapporté avoir été victime et/ou auteur de discours de haine en ligne. Ces résultats peuvent faire écho aux recherches portant sur le numérique qui identifient une grande quantité de contenus haineux en lien avec l'origine culturelle (Costello *et al.*, 2016; Reichelmann *et al.*, 2021; Walrave *et al.*, 2023). A l'instar de

ce qui était formulé dans notre troisième questionnement, les expériences vécues par les auteurs et/ou victimes au sein de notre étude ont eu pour conséquence de proposer une définition large du discours de haine en ligne. En effet, la majorité de ce sous-échantillon se retrouve dans la catégorie «contenu à caractère agressif», tant pour les définitions formulées que pour les expériences vécues. D'ailleurs, deux participants ayant défini le discours de haine en ligne de manière restrictive («contenu à caractère haineux») décrivent une expérience vécue comme auteur qui entre dans la catégorie des contenus à caractère agressif. Ce décalage peut refléter une tension entre représentations sociales et vécu personne, mais aussi comme l'effet de mécanismes d'adaptation telles que la normalisation de certaines conduites (Costello et Hawdon, 2020). Il est intéressant de noter que ces deux participants sont des jeunes hommes d'origine non européenne. Une étude de Ortiz (2019) explique ainsi que les personnes issues de minorités culturelles, en l'absence de soutien social de leurs pairs et lassés de répondre aux contenus dénigrants, vont se tourner vers la désensibilisation comme stratégie d'adaptation. Certains jeunes tendent à minimiser les comportements de victimisation qu'ils subissent – de manière directe ou indirecte, en tant que membres d'un groupe – pour se protéger. Cette minimisation ou banalisation peut également s'étendre sur leurs propres comportements agissants. Une telle dynamique fait écho aux études mettant en lumière les liens étroits entre les statut de victime et d'auteur (Costello et Hawdon, 2020; Wachs, Bilz, *et al.*, 2022).

### **Forces et limites de l'étude**

Premièrement, notre étude a inclus un nombre élevé de participants ( $n=23$ ) directement ciblés par le discours de haine en ligne et faisant partie de la tranche d'âge (15-25 ans) exposée à ce type de violence en ligne (Bautista-Ortuño *et al.*, 2018; al Serhan et Elareshi, 2019). Nos participants démontrent une diversité en termes de statut auto-rapporté et de variabilité de profils (orientation sexuelle, genre, origine culturelle) et ce, grâce à une variété de lieux de recrutement. Toutefois, comme mentionné dès le début de cet écrit, il est important de noter que, bien que cette diversité enrichisse la compréhension des expériences individuelles, les résultats ne peuvent pas être généralisés au-delà de cet échantillon spécifique et des caractéristiques qui lui sont propres. Ensuite, notre méthodologie repose sur une solide base scientifique, soutenue par un guide d'entretien rigoureusement élaboré, en adéquation avec les recherches antérieures tout en cherchant à les dépasser. Ainsi, le recours à des entretiens individuels, entrant progressivement dans l'objet d'étude examiné, le discours de haine en ligne, sur base des expériences racontées des participants, a permis de mener une analyse plus détaillée et nuancée que celle offerte par les études quantitatives précédentes (Awan, 2014; al Serhan et Elareshi, 2019; Reichelmann *et al.*, 2021; Walrave *et al.*, 2023). En effet, le discours de haine en ligne est souvent examiné par l'analyse des contenus publiés (Awan, 2014) ou encore des récurrences ou des statistiques de perpétration et de victimisation (al Serhan et Elareshi, 2019; Costello *et al.*, 2016; Reichelmann *et al.*, 2021; Walrave *et al.*, 2023). Le recours à des médias et à des tendances chiffrées pourraient parfois véhiculer des représentations qui ne reflètent pas fidèlement la réalité des jeunes. C'est pourquoi notre étude a visé à donner directement la parole aux jeunes concernés.

personnelles, de contribuer à l'élaboration de définitions et, surtout, de se sentir écoutés et valorisés dans leur témoignage : leur parole était attendue et entendue.

En lien, une des limites de cette étude est le choix, assumé, de présenter une vision plus quantifiée et descriptive de nos matériaux et analyses qualitatives. Si nous n'avions pas l'intention de compter, nos matériaux nous ont poussé à recenser et comparer les catégories ayant émergé de ceux-ci de façon à décrire les définitions et expériences vécues de la façon la plus objectivée possible. Nous n'avons toutefois pas procédé à une analyse discursive. En effet, ce n'était pas dans nos questions de recherche de vérifier le vocable utilisé de façon préférentielle, ni de procéder à un comptage strict visant à quantifier totalement nos matériaux de recherche afin de les comparer statistiquement entre eux. Par ailleurs, notre étude s'est également concentrée sur les perceptions et expériences des répondants à un moment donné. Il serait pertinent de compléter cette recherche par des études longitudinales. Celles-ci permettraient, entre autre, de vérifier l'hypothèse d'un continuum, soit linéaire, où les micro-agressions (contenu à caractère agressif) pourraient évoluer vers des discours de haine (contenu à caractère haineux) et/ou de déterminer si nous observons aussi une boucle rétroactive avec une normalisation des conduites et une amplification du phénomène, au sens où les participants pourraient définir le discours de haine en ligne en termes de contenus à caractère agressif alors qu'ils vivent des expériences répétées illustrant des contenus à caractère haineux. En lien, les tensions relevées au sein de certains auteurs de discours de haine en ligne, entre d'une part, les expériences vécues, et les définitions données, seraient pertinentes à examiner sur du long terme. Par exemple, est-ce que nous assisterons à une réduction de la dissonance se marquant par une amplification des conduites (intentionnalité, exclusion...) à la faveur des définitions à caractère haineuse ou à l'inverse, observerons-nous une restructuration cognitive amenant à percevoir le discours de haine en ligne comme étant plutôt une agression verbale ? Enfin, lors des entretiens, nous avons demandé aux répondants de déclarer leur statut en tant que témoin, auteur ou victime. Il est important de noter que l'adhésion au statut auto-rapporté peut introduire des biais, qui plus est au sein de cette tranche d'âge (Krumpal, 2013). En effet, certains jeunes ont pu minimiser leurs conduites, tandis que d'autres ont pu être tentés d'adopter fièrement l'étiquette d'auteur. En vue d'atténuer l'effet de la désirabilité sociale sur les propos partagés, une question neutre et ouverte a été posée («*Après notre discussion concernant tes expériences en ligne, dans quel(s) rôle(s) te reconnaîtrait-tu le plus ? Une victime ? Un auteur ? Un témoin ? Une combinaison de plusieurs rôles ?*»). De plus, en posant cette question à la fin de l'entretien, nous avons permis aux répondants de s'appuyer sur les expériences discutées précédemment pour formuler des réponses correspondant à leur réalité.

## Implications et conclusion

Cette étude a pu mettre en évidence l'intérêt de considérer les caractéristiques individuelles des jeunes exposés au discours de haine en ligne. Ainsi, nos résultats révèlent que certains groupes considérés dans la littérature comme davantage ciblés par de la haine en ligne tendent à adopter une définition plus large du discours de haine en ligne, alors que les hommes hétérosexuels, considérés plutôt comme auteurs par la littérature, proposent

une définition orientée sur les contenus explicitement haineux. Nous constatons également des décalages entre les définitions du discours de haine en ligne et les expériences rapportées par les participants. Ces écarts peuvent refléter une tension entre représentations sociales et vécu personnel, mais aussi l'effet de mécanismes d'adaptation tels que la désensibilisation. Plus largement, ces différences reflètent des expériences vécues diversifiées, allant de la victimisation à l'adoption de mécanismes d'adaptation tels que la désensibilisation ou l'exutoire émotionnel, où l'humour peut occuper un rôle ambivalent. Ceci questionne donc sur les facettes de l'humour en tant que potentiel outil de régulation émotionnelle, capable de dédramatiser certaines situations, mais également en tant que mécanisme susceptible d'alimenter des tensions ou des divisions.

Par ailleurs, nous avons observé que la catégorie des contenus à caractère agressif est une catégorie large, composée notamment de micro-agressions, de préjugés et d'actes individuels guidés par les préjugés, qui sont des comportements répétés et parfois réguliers (al Serhan et Elareshi, 2019; Anti-Defamation League, 2018), particulièrement mis en évidence au sein de la sphère interpersonnelle. Les réseaux sociaux seraient un lieu propice pour l'apparition et la diffusion de ces «contenus à caractère agressif», où les micro-agressions seraient fortement présentes dans les normes de communication du numérique (Eschmann *et al.*, 2020). Tant des recherches qualitatives que quantitatives soulignent l'importance des normes sociales dans la perpétration des discours de haine (Krause *et al.*, 2021; Wachs, Wettstein *et al.*, 2022). Suivant cette logique, si les réseaux sociaux acceptent davantage les micro-agressions et les contenus à caractère agressif que l'espace non numérique, cela peut favoriser un contexte propice au glissement à travers le continuum de la haine, tels que le discours de haine et le «contenu à caractère haineux», lesquels peuvent alors nourrir et légitimer les contenus à caractère agressifs. Dès lors, l'éducation numérique à destination des jeunes devrait intégrer le continuum des manifestations du discours en ligne, en mettant l'accent sur les micro-agressions et leur progression possible vers des discours explicitement haineux, et inversement. En complément, l'apparition de la haine en ligne dépendrait, entre autres, des normes de modération et de la philosophie qui les sous-tendent (par exemple, la liberté d'expression prônée par des plateformes comme X et Telegram). Il serait dès lors intéressant que des recherches futures s'intéressent davantage aux liens qui existent entre la présence de micro-agressions, la présence de discours de haine en ligne et les normes de modération. A noter que lorsque des plateformes imposent des règles strictes, de nouvelles plateformes plus permissives émergent, ce qui incite les premières à éviter des régulations plus sévères pour ne pas perdre d'utilisateurs et subir des pertes financières (Gillespie, 2018).

Nous pouvons également nous interroger sur le vocable choisi et plus précisément sur le «contenu à caractère agressif» qui lie plusieurs concepts pourtant distincts que sont les micro-agressions, le cyberharcèlement, les violences en ligne et les violences verbales. Ces amalgames ont déjà été pointés au sein d'autres études, mettant en évidence la difficulté pour les répondants à identifier le discours de haine en ligne (Ross *et al.*, 2017). Or, si ce regroupement semble rendre compte du

vécu et des expériences relatés par les jeunes dans notre étude, il comporte néanmoins le risque de diluer la spécificité du discours de haine. Aussi, même si la communauté scientifique, à laquelle nous nous affiliions ici, fait le choix de recourir au vocable ciblé du discours de haine en ligne, cette même communauté invite également à ne pas s'arrêter à la vision restrictive véhiculée par le discours de haine. Au contraire, elle encourage une réflexion plus poussée sur les fondements qui participent à définir directement ces phénomènes (Schweppe et Perry, 2022). Dans ce cadre, il est utile de s'interroger sur les raisons de ces difficultés chez les jeunes eux-mêmes : résultent-elles d'un manque de connaissances, d'une absence d'expériences personnelles, ou encore des mécanismes de distanciation et de banalisation ? Répondre à ces interrogations permettrait de déterminer si la spécificité du «discours de haine» est insuffisamment perçue par les acteurs eux-mêmes, et dès lors à préciser, ou si elle se trouve brouillée par la proximité conceptuelle avec d'autres formes de violences en ligne, qui seraient à intégrer et clarifier. En lien, il est nécessaire de penser les actions de prévention autour du discours de haine en ligne auprès des 15-25 ans en tenant compte des différents vocables utilisés et qualifications associées. Il pourrait ainsi être utile de discuter de la progressivité des conduites, à l'instar du système pyramidal de l'Anti-Defamation League (2018), ou encore du continuum de Sponholz (2017), voire de potentielles boucles rétroactives. Ensuite, en écho à Crespi et Hellsten (2022), la modalité numérique, associée à une multiplicité d'expression des contenus, doit être intégrée au sein des études empiriques et questionnée. En effet, cette modalité d'expression est bien présente dans les définitions de nos participants, même si davantage marquée pour la catégorie «contenu à caractère haineux». Notre étude met donc en évidence les nuances du discours de haine en ligne, exprimant de l'agressivité et de la haine, à visée interpersonnelle ou communautaire. Ces nuances nous semblent capitales sous peine de ne pas percevoir les réalités vécues par les victimes, voire de les banaliser et les minimiser. Concrètement, ces distinctions devraient être prises en compte dans le secteur de la modération des contenus en ligne : la non-adéquation entre le vécu des utilisateurs et les règles de modération pouvant accentuer le sentiment de banalisation des contenus (Gagliardone *et al.*, 2015). Enfin, le manque d'outils légaux et judiciaires pertinents pour traiter ces contenus entraîne un sentiment d'impunité pour leurs auteurs et un sentiment d'insécurité et de frustration pour les victimes (Duggan, 2017).

## Références

- al Serhan, F., et Elareshi, M. (2019). University Students' Awareness of Social Media Use and Hate Speech in Jordan. *International Journal of Cyber Criminology*, 13(2), 548–563. <https://doi.org/10.5281/zenodo.3709236>
- Anti-Defamation League (2018). *The Pyramid of hate*. Retrieved from <https://www.adl.org/sites/default/files/documents/pyramid-of-hate.pdf>
- Aranda Juárez, D., Sánchez-Navarro, J., et Mohammadi, L. (2020). Perception and self-assessment of digital skills and gaming among youth: A dataset from Spain. *Data in Brief*, 28, 104957–104957. <https://doi.org/10.1016/j.dib.2019.104957>
- Argüello-Gutiérrez, C., et Romero-Sánchez, M. (2018). Psicología social y humor: aproximaciones desde el humor de denigración. Dans J. Benavides-Delgado (Ed.), *Psicología y filosofía del humor* (pp. [non spécifié]). Bogotá: Ediciones Universidad Cooperativa de Colombia. <https://doi.org/10.16925/9789587600865>
- Awan, I. (2014). Islamophobia and Twitter: A Typology of Online Hate Against Muslims on Social Media: Islamophobia and Twitter. *Policy and Internet*, 6(2), 133–150. <https://doi.org/10.1002/1944-2866.POI364>
- Baider, F. (2019). Le discours de haine dissimulée : le mépris pour humilier. *Déviance et société*, 43(3), 359–387. <https://doi.org/10.3917/ds.433.0359>
- Bautista-Ortuño, R., Perea García, J., Rodríguez Gómez, N. et Castro Toledo, F. (2018). “May I offend you?” An experimental study on perceived offensiveness in online violent communication and hate speech. *International e-journal of criminal sciences*, 12.
- Bernatzky, C., Costello, M., et Hawdon, J. (2022). Who Produces Online Hate?: An Examination of the Effects of Self-Control, Social Structure, & Social Learning. *American Journal of Criminal Justice*, 47(3), 421–440. <https://doi.org/10.1007/s12103-020-09597-3>
- Castaña-Pulgarín, S., Suárez-Betancur, N., Vega, L. et López, H. (2021). Internet, social media and online hate speech. Systematic review. *Aggression and Violent Behavior*, 58, 101608. <https://doi.org/10.1016/j.avb.2021.101608>
- Chetty, N. et Alathur, S. (2018). Hate speech review in the context of online social networks. *Aggression and Violent Behavior*, 40, 108–118. <https://doi.org/10.1016/j.avb.2018.05.003>
- Clark, D. A., Spanierman, L. B., Reed, T. D., Soble, J. R., et Cabana, S. (2011). Documenting weblog expressions of racial microaggressions that target American Indians. *Journal of Diversity in Higher Education*, 4(1), 39–50. <http://dx.doi.org/10.1037/a0021762>
- Costello, M. et Hawdon, J. (2020). Hate Speech in Online Spaces. In The Palgrave Handbook of International Cybercrime and Cyberdeviance (p.1397–1416). Springer International Publishing. [https://doi.org/10.1007/978-3-319-78440-3\\_60](https://doi.org/10.1007/978-3-319-78440-3_60)
- Costello, M., Hawdon, J., Bernatzky, C. et Mendes, K. (2019). Social Group Identity and Perceptions of Online Hate. *Sociological Inquiry*, 89(3), 427–452. <https://doi.org/10.1111/soin.12274>
- Costello, M., Hawdon, J., Ratliff, T., et Grantham, T. (2016). Who views online extremism? Individual attributes leading to exposure. *Computers in Human Behavior*, 63, 311–320
- Cramer, R., Cacace, S., Sorby, M., Adrian, M., Kehn, A., et Wilsey, C. N. (2022). A Psychometric Investigation of the Hate-Motivated Behavior Checklist. *Journal of Interpersonal Violence*, 38(7-8), p. 5638–5660. <https://doi.org/10.1177/08862605221127196>
- Crespi, I., et Hellsten, L. (2022). Cyberviolence and the digital experience: reflections on a problematic issue for youth. *International Review of Sociology*, 32(3), 391–399. <https://doi.org/10.1080/03906701.2022.2133404>

- Döring, N., et Mohseni, M. (2019). Fail videos and related video comments on YouTube: a case of sexualization of women and gendered hate speech? *Communication Research Reports*, 36(3), 254–264. <https://doi.org/10.1080/08824096.2019.1634533>
- Duggan, M. (2017). *Online harassment 2017*. Pew Research Center. <https://www.pewresearch.org/fact-tank/2017/07/11/americans-are-worried-about-online-harassment-but-they-dont-feel-safe-reporting-it/>
- Eschmann, R., Groshek, J., Chanderdatt, R., Chang, K., et Whyte, M. (2020). Making a Microaggression: Using Big Data and Qualitative Analysis to Map the Reproduction and Disruption of Microaggressions through Social Media. *Social Media + Society*, 6(4), <https://doi.org/10.1177/2056305120975716>
- Estanyol, E., Montaña, M., Fernández-de-Castro, P., Aranda, D., et Mohammadi, L. (2023). Competencias digitales de la juventud en España: Un análisis de la brecha de género. *Comunicar (Huelva, Spain)*, 31(74), 113–123. <https://doi.org/10.3916/C74-2023-09>
- Evans, C. B. R., Smokowski, P. R., Rose, R. A., Mercado, M. C., et Marshall, K. J. (2019). Cumulative Bullying Experiences, Adolescent Behavioral and Mental Health, and Academic Achievement: An Integrative Model of Perpetration, Victimization, and Bystander Behavior. *Journal of Child and Family Studies*, 28(9), 2415–2428. <https://doi.org/10.1007/s10826-018-1078-4>
- Gagliardone, I., Gal, D., Alves, T., et Martinez, G. (2015). *Combattre les discours de haine sur internet. Collection UNESCO sur la liberté de l'Internet*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000234620>
- Gangi, O., Brassine, N. et Mathys, C. (2023) Entre discours de haine en ligne et cyberharcèlement chez un public belge de 15 à 25 ans : une distinction de fait et de droit, mais une distinction pertinente en criminologie ? *Criminologie, Forensique et Sécurité*, 1 (1). 3620.
- Gillespie, T. (2018). *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions*. That Shape Social Media. London: Yale University Press.
- Hall, C. (2009). Sticks and stones may break bones but will the law ever protect me? Ensuring educational access through federal prohibition of peer-on-peer harassment. *Children's Legal Rights Journal*, 29(4), 42–
- Harmer, E. et Lumsden, K. (2019) 'Introduction: Online Othering.' In: K. Lumsden and E. Harmer (eds) *Online Othering: Exploring Violence and Discrimination on the Web*. Basingstoke: Palgrave Macmillan.
- Harrison, J., Chapleau, A., Schell, M., Vaughan, V., et Colzin, C. (2020). Leveraging technology in field education for digital natives: Using goal attainment scaling. *Social Work Education*, 39(1), 60–70. <https://doi.org/10.1080/02615479.2019.1642318>
- Hawdon, J., Oksanen, A. et Räsänen, P. (2017). *Exposure to Online Hate in Four Nations: A Cross-National Consideration*. *Deviant Behavior*, 38(3), 254–266. <https://doi.org/10.1080/01639625.2016.1196985>
- Jacks, W., et Adler, J. R. (2016). A proposed typology of online hate crime. *Open Access Journal of Forensic Psychology*, 7, 64–89.
- Jane, E. (2018). Systemic misogyny exposed: Translating Rapeglash from the Manosphere with a Random Rape Threat Generator. *International Journal of Cultural Studies*, 21(6), 661–680. <https://doi.org/10.1177/1367877917734042>
- Keipi, T., Näsi, M., Oksanen, A., et Räsänen, P. (2017). *Online hate and harmful content: Cross-national perspectives*. London: Routledge.
- Krause, N., Ballaschk, C., Schulze-Reichelt, F., Kansok-Dusche, J., Wachs, S., Schubarth, W., et Bilz, L. (2021). « Ich lass mich da nicht klein machen! » Eine qualitative Studie zur Bewältigung von Hatespeech durch Schüler/innen. *Zeitschrift für Bildungsforschung (Internet)*, 11(1), 169–185. <https://doi.org/10.1007/s35834-021-00291-w>
- Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity*, 47(4), 2025–2047.
- Martineau, W. H. (1972). A model of social functions of humor. In H. Goldstein & P.E. McGhee, *The Psychology of Humor*. 101–125. New York : NYA <https://doi.org/10.1353/rmr.1985.0073>
- Meyer, D. (2010). Evaluating the Severity of Hate-motivated Violence: Intersectional Differences among LGBT Hate Crime Victims. *Sociology (Oxford)*, 44(5), 980–995. <https://doi.org/10.1177/0038038510375737>
- Miles, M. et A. Huberman. (2003). *Analyse des données qualitatives*. Paris : De Boeck Université.
- Monnier, A., et Seoane, A. (2019). Discours de haine sur l'internet. Dans *Publicationnaire. Dictionnaire encyclopédique et critique des publics*. Misen ligne le 12 juin 2019. <http://publicationnaire.huma-num.fr/notice/discours-de-haine-sur-linternet/>
- Monnier, A., Seoane, A., Hubé, N., et Leroux, P. (2021). Discours de haine dans les réseaux socionumériques. *Mots. Les langages du politique*, (125), 9–14
- Ortiz, S. (2019). « You Can Say I Got Desensitized to It »: How Men of Color Cope with Everyday Racism in Online Gaming. *Sociological Perspectives*, 62(4), 572–588. <https://doi.org/10.1177/0731121419837588>
- Ortiz, S. (2021). Racists without racism? From colourblind to entitlement racism online. *Ethnic and Racial Studies*, 44(14), 2637–2657. <https://doi.org/10.1080/01419870.2020.1825758>
- Reichelmann, A., Hawdon, J., Costello, M., Ryan, J., Blaya, C., Llorent, V., Oksanen, A., Räsänen, P., et Zych, I. (2021). Hate Knows No Boundaries: Online Hate in Six Nations. *Deviant Behavior*, 42(9), 1100–1111. <https://doi.org/10.1080/01639625.2020.1722337>
- Ross, B., Rist, M., Carbonell, G., Cabrera, B., Kurowsky, N., et Wojatzki, M. (2017). *Measuring the reliability of hate speech annotations: The case of the European refugee crisis*. [Communication présentée à The 3rd Workshop on Natural Language Processing for Computer Mediated Communication, Bochum, Germany.]
- Schweppé, J., et Perry, B. (2022). A continuum of hate: delimiting the field of hate studies. *Crime, Law, and Social Change*, 77(5), 503–528. <https://doi.org/10.1007/s10611-021-09978-7>

Smith, E. R., Et Mackie, D. M. (2015). Dynamics of Group-Based Emotions: Insights From Intergroup Emotions Theory. *Emotion Review*, 7(4), 349–354. <https://doi.org/10.1177/1754073915590614>

Soral, W., Bilewicz, M., et Winiewski, M. (2018). Exposure to hate speech increases prejudice through desensitization. *Aggressive Behavior*, 44(2), 136–146. <https://doi.org/10.1002/ab.21737>

Sponholz, L. (2017). *Tackling hate speech with counter speech? Practices of contradiction and their effects*. [Communication présentée à International Conference Worlds of Contradiction, Bremen, Germany.]

Sue, D. W., Capodilupo, C. M., et Holder, A. M. B. (2008). Racial Microaggressions in the Life Experience of Black Americans. *Professional Psychology, Research and Practice*, 39(3), 329–336. <https://doi.org/10.1037/0735-7028.39.3.329>

Venäläinen, S., et Virkki, T. (2019). Struggles for moral value and the reproduction of gendered and racialised hierarchies in online discussions of violence. *The Sociological Review*, 67(6), 1367–1382. <https://doi.org/10.1177/0038026119866635>

Wachs, S., Bilz, L., Wettstein, A., Wright, M. F., Krause, N., Ballaschk, C., et Kansok-Dusche, J. (2022). The Online Hate Speech Cycle of Violence: Moderating Effects of Moral Disengagement and Empathy in the Victim-to-Perpetrator Relationship. *Cyberpsychology, Behavior and Social Networking*, 25(4), 223–229. <https://doi.org/10.1089/cyber.2021.0159>

Wachs, S., Wettstein, A., Bilz, L., Krause, N., Ballaschk, C., Kansok-Dusche, J., et Wright, M. (2022). Playing by the Rules? An Investigation of the relationship between social norms and adolescents' hate speech perpetration in schools. *Journal of Interpersonal Violence*, 37(21-22), 21143–21164. <https://doi.org/10.1177/08862605211056032>

Walrave, M., Franssen, V., Mathys, C., Van de Heyning, C., Vrielink, J., Giacometti, M., Gangi, O., Gilen, A. (2023). *Cyberviolence: defining borders on permissibility and accountability - @ntidote 2.0, Final Report*. Brussels: Belgian Science Policy Office 2021–2023 (BRAIN-be 2.0- (Belgian Research Action through Interdisciplinary Networks)

Wieland, J. (2007). Peer-on-peer hate crime and hate-motivated incidents involving children in California's public schools: Contemporary issues in prevalence, response and prevention. *UC Davis Journal of Juvenile Law et Policy*, 11(2), 235–269

Willard, N. (2004). *An educator's guide to cyberbullying and cyberthreats*. Center for Safe and Responsible Use of the Internet. <http://cyberbully.org>

Yin, R. (2017). *Case study research and applications: Design and methods*. Los Angeles, CA: Sage.

Zhang, Z., & Luo, L. (2018). Hate speech detection: a solved problem? The challenging case of Long Tail on Twitter. *Semantic Web*, 10(5), 925–945. <https://doi.org/10.3233/SW-180338>

Ziv, A. (2010). The Social Function of Humor in Interpersonal Relationships. *Society* (New Brunswick), 47(1), 11–18. <https://doi.org/10.1007/s12115-009-9283-9>