



Éditions  
UNESCO

Organisation  
des Nations Unies  
pour l'éducation,  
la science et la culture

# COMBATTRE LES DISCOURS DE HAINE SUR INTERNET

Iginio Gagliardone • Danit Gal • Thiago Alves • Gabriela Martinez

COLLECTION UNESCO SUR LA LIBERTÉ DE L'INTERNET



# COMBATTRE LES DISCOURS DE HAINE SUR INTERNET



Iginio Gagliardone • Danit Gal • Thiago Alves • Gabriela Martinez

---

Publié en 2015 par l'Organisation des Nations Unies pour l'éducation, la science et la culture  
7, place de Fontenoy, 75352 Paris 07 SP, France

© UNESCO 2015

ISBN 978-92-3-200071-2



Ceuvre publiée en libre accès sous la licence Attribution-ShareAlike 3.0 IGO ([CC-BY-SA 3.0 IGO](http://creativecommons.org/licenses/by-sa/3.0/igo)) (<http://creativecommons.org/licenses/by-sa/3.0/igo>). Les utilisateurs du contenu de la présente publication acceptent les termes d'utilisation de l'Archive ouverte de libre accès UNESCO ([www.unesco.org/open-access/terms-use-ccbysa-fr](http://www.unesco.org/open-access/terms-use-ccbysa-fr)).

Titre original: *Countering Online Hate Speech*

Publié en 2015 par l'Organisation des Nations Unies pour l'éducation, la science et la culture

Les désignations employées dans cette publication et la présentation des données qui y figurent n'impliquent de la part de l'UNESCO aucune prise de position quant au statut juridique des pays, territoires, villes ou zones, ou de leurs autorités, ni quant au tracé de leurs frontières ou limites.

Les idées et les opinions exprimées dans cette publication sont celles des auteurs ; elles ne reflètent pas nécessairement les points de vue de l'UNESCO et n'engagent en aucune façon l'Organisation.

La présente publication a été rendue possible grâce à une contribution de la Suède

Mise en pages et impression par UNESCO

*Imprimé en France*

# TABLE DES MATIÈRES

|   |           |
|---|-----------|
| <b>AVANT-PROPOS</b>   | <b>5</b>  |
| <b>1. RÉSUMÉ ANALYTIQUE</b>   | <b>7</b>  |
| <b>2. INTRODUCTION</b>  | <b>10</b> |
| Le débat sur les discours de haine  | 11        |
| Ce qui caractérise les discours de haine en ligne   | 13        |
| Les réactions aux discours de haine en ligne : mesures juridiques et non juridiques                         | 16        |
| <b>3. MÉTHODOLOGIE</b>  | <b>19</b> |
| <b>4. CADRES NORMATIFS</b>  | <b>21</b> |
| Les discours de haine : étude des principes internationaux permettant de les identifier et de les combattre | 21        |
| <i>Les discours de haine et le Pacte international relatif aux droits civils et politiques</i>              | 22        |
| <i>Autres instruments juridiques internationaux</i>   | 25        |
| <i>Mesures régionales contre les discours de haine</i>  | 27        |
| <i>Résumé des dispositions internationales visant à limiter la liberté d'expression</i>                     | 31        |
| <i>Espaces d'expression privés et discours de haine</i>   | 32        |
| <i>La définition et la réglementation des discours de haine au sein des espaces d'expression privés</i>     | 34        |
| <i>Définition et mise en œuvre</i>  | 36        |
| Conclusion  | 38        |
| <b>5. ANALYSE DES RÉPONSES DE LA SOCIÉTÉ</b>  | <b>39</b> |
| Suivi du discours de haine et discussions à son sujet   | 40        |
| Mobiliser la société civile   | 44        |
| Faire pression sur les entreprises privées  | 46        |
| <i>Différentes approches stratégiques</i>   | 47        |
| <i>Faire campagne sur le long terme</i>   | 49        |
| <i>Faire campagne contre les publicitaires</i>  | 50        |

|  |           |
|--|-----------|
| <i>Limites à la lutte contre le discours de haine en ligne et réactions des intermédiaires Internet</i>            | 53        |
| Lutter contre le discours de haine en ligne par l'éducation de base aux médias et à l'information                  | 54        |
| <i>Éducation à la citoyenneté et citoyenneté numérique</i>   | 54        |
| <i>L'éducation comme outil de lutte contre le discours de haine</i>  | 56        |
| <i>Développement de l'esprit critique pour contrer le discours de haine en ligne</i>                               | 57        |
| <i>Objectifs éducatifs de l'éducation de base aux médias et à l'information pour répondre du discours de haine</i> | 59        |
| <i>Évaluer l'éducation de base au bon usage des médias et de l'information et les initiatives éducatives</i>       | 61        |
| <b>6. CONCLUSION ET RECOMMANDATIONS</b>  | <b>62</b> |
| Définition   | 62        |
| Juridiction  | 64        |
| Compréhension  | 66        |
| Intervention   | 68        |
| <b>Auteurs et remerciements</b>  | <b>70</b> |
| <b>Liste des œuvres citées :</b>   | <b>71</b> |
| <b>Notes</b>   | <b>75</b> |

# AVANT-PROPOS

Les possibilités qu'offre Internet éclipsent largement les difficultés qui en résultent. Tout en gardant cela à l'esprit, nous devons néanmoins nous attaquer à certains de ces problèmes, et notamment celui des discours de haine en ligne. Mais en quoi consistent exactement les propos haineux sur Internet et comment les combattre efficacement ?

En matière de liberté d'expression, sur Internet ou hors Internet, l'UNESCO préconise que la libre circulation de l'information soit toujours la norme. Le contre-discours est généralement préférable à la répression de la liberté de parole. Et toute mesure de restriction de la liberté d'expression doit être soigneusement considérée afin de garantir que ce type d'action demeure tout à fait exceptionnel et que le débat légitime et approfondi ne soit pas entravé.

En ce qui concerne l'expression de la haine, les normes internationales peuvent nous aider, de manière générale, à identifier les discours haineux et à trouver des solutions pour les contrer :

- Dans le cas des messages haineux de nature raciale, La Convention internationale sur l'élimination de toutes les formes de discrimination raciale exige qu'on interdise l'expression d'idées de supériorité ou d'infériorité à l'encontre d'individus catégorisés selon leur « race ».

---

- Toute expression de haine fondée sur la nationalité ou la religion est interdite par la loi selon l'article 20 du Pacte international relatif aux droits civils et politiques – mais il est précisé qu'il doit s'agir d'un (i) appel à la haine, (ii) qui constitue une incitation à (iii) la discrimination, l'hostilité ou la violence.

---

- Il est possible, bien que non exigé, que les discours haineux, lorsqu'ils reposent sur le sexe, l'orientation sexuelle ou d'autres caractéristiques fassent l'objet de restrictions selon les termes du Pacte international relatif aux droits civils et politiques (article 19), dans l'intérêt du respect des droits ou de la réputation d'autrui.

---

Néanmoins, dans ces trois cas de figure, toute limitation, loin d'être arbitraire, doit être précisée par la loi. Elle doit également répondre au critère de « nécessité » – ce qui implique que la limitation soit proportionnée et ciblée afin d'éviter toute restriction collatérale injustifiée de propos légitimes.

Les normes internationales exigent également que toute restriction de la liberté d'expression soit caractérisée par un objectif légitime et ne puisse être en aucun cas un simple exercice de pouvoir. Outre la défense des droits d'autrui indiquée précédemment, les objectifs peuvent aussi être la sécurité nationale, la moralité publique ou la santé publique.

Au vu de la complexité du sujet, il nous semble nécessaire d'adopter une approche éclairée et nuancée afin d'identifier et de limiter les « discours de haine ». Il ne faudra rien de moins pour

assurer un juste équilibre avec la libre circulation des idées et de l'information. Cet équilibre est particulièrement crucial au regard des modes d'expression sur Internet, comme l'explique cette publication.

Si Internet n'échappe pas aux lois, la conception et l'application des mesures juridiques à l'encontre des propos perçus comme des discours de haine sont néanmoins complexes. C'est la raison pour laquelle la présente étude examine des solutions sociales qui pourraient être considérées comme complémentaires à toute restriction juridique mise en œuvre par un État.

Cette étude présente une typologie des différentes méthodes adoptées pour combattre les discours de haine sur Internet. La première consiste en un suivi et une analyse de ces discours par la société civile. La seconde est la promotion par les internautes du contre-discours en « peer-to-peer ». La troisième est une série de mesures prises par des ONG afin d'informer les autorités sur certains cas, et la quatrième consiste à favoriser des actions menées par les fournisseurs Internet qui hébergent les contenus, y compris les propos haineux. La cinquième, enfin, est structurelle : il s'agit de permettre aux utilisateurs d'exercer leur droit à la liberté d'expression sur Internet en leur offrant, à travers l'éducation et la formation, les connaissances, notamment en matière d'éthique, et les compétences nécessaires. C'est ce que l'UNESCO appelle l'initiation aux médias et à la maîtrise de l'information.

Ce sont des réponses créatives de la société comme celles que nous venons de citer qui auront toutes les chances de produire des résultats. Elles pourront en effet garantir qu'Internet demeure un espace de potentiel positif et que ce « réseau de réseaux » contribue à la construction des sociétés du savoir fondées sur la paix, le respect des droits de l'homme et le développement durable.

Le Directeur général adjoint  
Getachew Engida

# 1. RÉSUMÉ ANALYTIQUE

Les discours de haine sur Internet se situent au croisement de nombreuses tensions : ils sont l'expression de conflits entre différents groupes, au sein des sociétés mais aussi entre les sociétés ; ils révèlent clairement que des technologies recélant un potentiel de transformation telles qu'Internet sont à la fois sources de possibilités nouvelles et de problèmes, ce qui suppose de maintenir un équilibre subtil entre les droits et principes fondamentaux, notamment la liberté d'expression, et la défense de la dignité humaine.

En tant qu'organisme des Nations Unies dont le mandat spécifique est de promouvoir la liberté d'expression et ses corollaires, la liberté de la presse et la liberté de l'information, l'UNESCO contribue activement à favoriser la connaissance et la compréhension mutuelle des nations en utilisant tous les moyens de communication de masse, y compris Internet en général et les plates-formes des réseaux sociaux en particulier.

Les recherches exposées dans cette publication trouvent leur origine dans la mise en œuvre de la résolution 52, votée par les 195 États membres de l'Organisation lors de la 37<sup>e</sup> session de la Conférence générale de l'UNESCO qui s'est tenue en novembre 2013. Cette résolution demandait la réalisation, par le biais d'un processus de consultation multipartite, d'une étude d'ensemble, dans le cadre du mandat de l'UNESCO, sur les questions relatives à Internet concernant l'accès à l'information et au savoir, la liberté d'expression, le respect de la vie privée et les dimensions éthiques de la Société de l'information. Les recherches sur les discours de haine ont ainsi contribué à l'élaboration de cette étude plus vaste.<sup>1</sup>

Le présent rapport offre un panorama mondial des dynamiques caractéristiques des discours de haine en ligne et des mesures prises pour les combattre et les contenir, en mettant en évidence les bonnes pratiques qui ont émergé au niveau local comme au niveau international. Si l'étude fournit une analyse exhaustive des cadres normatifs internationaux, régionaux et nationaux mis en place pour lutter contre les discours de haine en ligne et de leurs répercussions sur la liberté d'expression, elle met particulièrement l'accent sur les mécanismes sociaux et non réglementaires qui peuvent contribuer à lutter contre la production, la diffusion et l'impact des messages de haine en ligne.

Les conclusions de cette étude s'articulent autour de quatre axes principaux : définition, compétence juridique, compréhension et intervention.

- **Définition.** Le terme « discours de haine » a une portée très large et il fait l'objet de contestations. Certains traités multilatéraux, tels que le Pacte international relatif aux droits civils et politiques, ont tenté d'en définir les contours. On a par ailleurs initié des processus multipartites (par exemple le Plan d'action de Rabat) afin de clarifier la nature des messages haineux et de suggérer l'adoption de mesures

permettant de les identifier. Cependant, on emploie toujours communément les mots « discours de haine » en tant que terme générique, alors qu'ils mettent sur un même plan des propos qui menacent concrètement la sécurité d'individus ou de groupes et des messages dans lesquels des individus ne font que laisser libre cours à leur colère contre l'autorité. Les intermédiaires de l'Internet – les sociétés qui assurent la médiation de la communication en ligne telles que Facebook, Twitter et Google – ont leurs propres définitions des discours de haine, en fonction desquelles ils imposent à leurs utilisateurs un ensemble de règles et permettent aux entreprises de restreindre certains types de discours. Des organismes nationaux et régionaux ont tenté de promouvoir une acception du terme davantage enracinée dans les traditions locales. Dans ce contexte, il semble improbable d'arriver à une définition universellement partagée, mais le souhait commun d'éviter la violence et de protéger la dignité humaine a permis aux participants aux débats sur les discours de haine, venus d'horizons divers, de se rapprocher afin de rechercher ensemble des solutions pertinentes sur le plan local.

- **Compétence juridique.** Du fait de la rapidité et de la portée d'Internet, les gouvernements se heurtent à des difficultés lorsqu'ils veulent faire appliquer la législation nationale dans le cyberspace. Les problèmes relatifs aux discours de haine sur Internet mettent en lumière l'émergence d'espaces d'expression privés qui remplissent une fonction de nature publique (par exemple Facebook ou Twitter) et les défis que posent ces espaces aux autorités de réglementation. Malgré des résistances initiales et face à la pression publique, certaines des entreprises propriétaires de ces espaces sont devenues plus réactives dans leur lutte contre les discours de haine en ligne, bien qu'elles n'aient pas (encore) été pleinement intégrées dans les débats internationaux (notamment le Plan d'action de Rabat) visant à identifier et combattre les discours de haine.
- **Compréhension.** La nature même des discours de haine en ligne et leurs liens avec les discours et les actes hors Internet sont mal compris. Ces thèmes sont très souvent abordés – notamment par des hommes politiques, des activistes ou des universitaires – mais les discussions dont ils font l'objet sont généralement dépourvues de données empiriques systématiques. En raison de la nature des discours de haine et de leurs conséquences éventuelles, on a mis l'accent sur la recherche de solutions à ce problème, solutions qui doivent se référer aux dispositions internationales en matière de droits de l'homme. Or cette approche a également entravé des recherches plus approfondies visant à comprendre les causes profondes de ce phénomène et la manière dont certains types de contenus apparaissent, sont diffusés et conduisent (ou non) à la discrimination, l'hostilité ou la violence. La présente étude offre différents exemples de recherches visant à cartographier l'émergence et la diffusion des propos haineux sur Internet, mais elle insiste également sur l'absence de travaux de recherche axés sur les liens entre les

discours de haine en ligne et d'autres phénomènes sociaux, tels que l'éducation ou les inégalités croissantes.

---

- **Intervention.** Dans cette étude, nous avons identifié une grande diversité de méthodes employées pour résoudre des problèmes spécifiques et contextuels. Il en résulte d'importants enseignements généraux. Premièrement, l'acceptation large du terme « discours de haine » et la gravité des dommages potentiels qu'ils peuvent engendrer incitent à débattre des définitions, du suivi et de la contextualisation. Deuxièmement, ce rapport étudie les mesures adoptées par les entreprises privées pour combattre des cas répertoriés de propos haineux en encourageant les utilisateurs à signaler les discours de haine et à recourir au contre-discours, et il examine par ailleurs les architectures de plusieurs plates-formes de réseaux sociaux et leur influence à la fois sur la diffusion des discours de haine en ligne et sur les réactions à ces discours. Troisièmement, ce rapport montre à quel point les organisations spécialisées dans ce combat sont importantes car elles permettent de faire converger les réactions individuelles tout en faisant pression sur les entreprises et les pouvoirs publics pour qu'ils agissent. Mais ce qui est plus important encore, c'est que plusieurs initiatives diverses peuvent être complémentaires. Ainsi, par exemple, les intermédiaires de l'Internet se montrent de plus en plus attentifs aux demandes provenant d'utilisateurs individuels. Cependant, comme les intermédiaires n'ont pas publié de résultats globaux, qui auraient permis une meilleure compréhension du phénomène, les organisations de la société civile ont cherché à combler ces lacunes en mettant à la disposition des internautes des plates-formes chargées de recueillir leurs signalements. Dans le même temps, des initiatives ont été lancées en matière d'éducation afin de permettre aux internautes, lorsqu'ils sont confrontés à des messages haineux, de savoir quoi signaler précisément et comment le faire. Comme le révèle cette étude, certains éléments particuliers, inhérents au problème des discours de haine sur Internet, suggèrent qu'il serait inefficace de se concentrer sur des actions isolées ou réalisées par un seul acteur. Des efforts concertés sont nécessaires pour trouver des solutions appropriées à ce phénomène émergent.
-

## 2. INTRODUCTION

Il existe un ensemble de liens très complexes entre les discours de haine d'une part et la liberté d'expression, les droits des individus, des groupes et des minorités d'autre part, sans oublier les concepts de dignité, de liberté et d'égalité. La définition du terme « discours de haine » est fréquemment contestée. Selon la loi, à l'échelle nationale et internationale, ces derniers constituent une incitation à causer du tort à autrui (en particulier une incitation à la discrimination, à l'hostilité et à la violence), dont la cible correspond à un certain groupe social ou démographique. La définition des discours de haine peut inclure, sans toutefois s'y limiter, des propos qui comportent une menace de violence ou préconisent l'usage de la violence. Pour certains, cependant, ce concept s'étend également aux propos qui encouragent l'installation d'un climat de préjugés et d'intolérance car ils partent du principe que les propos haineux sont susceptibles d'alimenter la discrimination ciblée, l'hostilité et les attaques violentes.

Toutefois, dans le langage courant, les définitions des discours de haine sont généralement plus larges, allant même parfois jusqu'à inclure des propos insultants à l'encontre des dirigeants au pouvoir ou bien désobligeants à l'égard de personnes particulièrement visibles. C'est surtout à des moments critiques, notamment pendant les élections, que le concept de discours de haine peut se prêter à la manipulation : les opposants politiques peuvent être accusés de préférer des propos haineux ou bien les dirigeants peuvent utiliser ces accusations afin de juguler toute critique et toute dissidence.

Tout au long de cette étude, nous examinerons différentes **définitions** des discours de haine, proposées par divers acteurs, des organisations internationales jusqu'aux plates-formes des réseaux sociaux, et nous expliquerons pourquoi le terme « discours de haine » est si difficile à cerner. Cependant, bien qu'il échappe à des définitions claires, ce terme (qu'il soit véhiculé par le texte, l'image ou le son) peut être identifié par approximation au travers des **fonctions** dégradantes ou déshumanisantes qu'il remplit. En nous référant aux travaux de Waldron, nous pouvons affirmer que des propos sont considérés comme haineux (qu'ils soient véhiculés par le texte, l'image ou le son) lorsqu'ils envoient deux sortes de message. Le premier est envoyé au groupe ciblé et sa fonction est de déshumaniser et d'humilier les membres attribués à ce groupe. On a souvent affaire à ce genre de discours :

Ne vous méprenez pas en pensant que vous êtes les bienvenus ici. [...] On ne veut pas de vous ; vos familles et vous, vous serez rejetés, exclus, battus et mis dehors, à chaque fois qu'on pourra le faire impunément. En ce moment, il se peut qu'on doive faire profil pas. Mais ne vous sentez pas trop à l'aise. [...] Craignez nous (Waldron 2012).

Les discours de haine peuvent avoir une deuxième fonction : celle de faire savoir aux personnes qui partagent vos opinions qu'elles ne sont pas seules, de renforcer leur sentiment

d'appartenance à un groupe (prétendument) menacé. Voici le genre de message envoyé cette fois-ci à des individus qui partagent les vues de l'expéditeur :

Nous savons que certains d'entre vous conviennent qu'on ne veut pas de ces gens-là ici. Nous savons que certains d'entre vous pensent qu'ils sont sales (ou dangereux, ou bien que ce sont des criminels ou des terroristes). Sachez maintenant que vous n'êtes pas seuls. [...] Nous sommes suffisamment nombreux ici pour veiller à ce que ces gens-là ne soient pas les bienvenus. Nous sommes suffisamment nombreux ici pour attirer l'attention sur la véritable nature de ces gens (Waldron 2012).

Les discours de haine reposent sur les tensions, qu'ils cherchent à reproduire et à amplifier. De tels propos unissent et divisent en même temps. Ils créent un « nous » et un « ils ». Dans la présente étude, nous employons généralement le terme « discours de haine » au sens large, sans limiter sa signification aux discours qui constituent une incitation à causer du tort à autrui.

Dans cette étude, nous présentons un certain nombre d'axes autour desquels s'articulent les propos haineux, mais sans explorer nécessairement toute la gamme des critères sociaux tels que la race, l'origine ethnique, la langue, le sexe, la religion, l'orientation sexuelle ou la nationalité. Nous reconnaissons que la notion de discours de haine, quelle qu'en soit la définition, n'est pas fondée sur des idées abstraites, telles que les idéologies, les convictions politiques ou les croyances – idées qui ne doivent pas être confondues avec certains groupes susceptibles d'y souscrire. Les discours de haine concernent les antagonismes entre les individus. La présente étude souligne par ailleurs que l'on rencontre actuellement les plus gros problèmes de discours de haine en ligne dans des pays où le degré de connectivité Internet est important. Dans le même temps, cette situation peut laisser présager une évolution semblable dans d'autres pays, puisque les individus sont de plus en plus connectés à travers le monde. Cette étude indique également qu'un grand nombre de solutions qui y sont évaluées sont issues de réactions à des cas de discours de haine en ligne. Dans cette optique, certaines de ces expériences pourraient être prises en compte en amont et de manière proactive, sans attendre l'apparition du problème. Cette étude a pour but de tirer des enseignements d'ordre général d'un large éventail de situations empiriques.

## Le débat sur les discours de haine

Le phénomène des discours de haine remet en question certains des principes fondamentaux sur lesquels reposent les sociétés. Les réponses que chaque société a élaborées afin de maintenir un équilibre entre la liberté d'expression et le respect de l'égalité et de la dignité ont provoqué des divisions et des alliances sans précédent sur le plan international. Une grande partie des recherches sur les discours de haine, par exemple, s'est concentrée sur le clivage qui existe entre l'approche américaine et l'approche européenne dans la réglementation des propos haineux (Rosenfeld 2012; Bleich 2013). Aux États-Unis, en effet, la protection

de la liberté d'expression s'étend bien au-delà des limites des propos tolérés en Europe. La caractéristique majeure de cette approche est qu'elle met l'accent sur la nécessité d'identifier un danger manifeste et précis afin de pouvoir interdire ou de sanctionner certains types de discours. De nombreux pays européens, y compris l'Allemagne et la France, ont opté en revanche pour une approche qui interdit certains types de discours non seulement en raison de la probabilité qu'ils nuisent à autrui, mais aussi à cause de leur contenu intrinsèque.

D'autres sociétés ont mis au point des méthodes originales permettant d'identifier et de combattre les discours de haine, qui peuvent combiner, de diverses manières, droit coutumier et loi formelle. En Somalie, par exemple, où la poésie est un moyen essentiel à la diffusion des idées et de l'information au sein de la population, les poètes qui ont composé plusieurs œuvres dont le contenu est considéré par les anciens comme désobligeant à l'égard de certains individus ou groupes peuvent se voir interdire la création de nouveaux poèmes (Stremlau 2012). L'analyse du rôle joué par les discours de haine dans les atrocités ou les explosions de violence a donné lieu à d'importants travaux de recherche (Kellow and Steeves 1998, Thompson 2007, Yanagizawa-Drott 2010). Mais les recherches systématiques portant sur le phénomène des discours de haine et leur réglementation hors des États-Unis et de l'Europe sont encore marginales.

On conteste aussi le concept de discours de haine en lui reprochant d'être trop vaste et donc sujet aux manipulations, et l'on propose à la place des concepts plus étroits, notamment ceux du « discours dangereux » et du « discours attisant la peur », qui mettent l'accent sur la capacité du discours à nuire à autrui et mener à des comportements violents. Si l'on trouve des discours de haine – quelle qu'en soit la forme – dans presque toutes les sociétés, y compris dans celles où le risque de violence est faible, le concept de « discours dangereux », lui, vise à isoler des actes qui présentent une forte probabilité de « catalyser ou d'amplifier la violence, celle-ci provenant d'un groupe et étant dirigée contre autre groupe » (Benesch 2012). Susan Benesch a proposé un cadre d'étude qui permet d'identifier un acte de langage dangereux en s'appuyant sur : i) la personnalité et la popularité du locuteur ; ii) l'état émotionnel du public ; iii) le contenu de l'acte de langage en tant qu'appel à l'action ; iv) le contexte historique et social dans lequel se produit l'acte de langage ; et enfin v) les moyens utilisés pour le diffuser (y compris le type de langage adopté, par exemple si le discours est formulé dans une langue locale, une personne issue de la région où l'on parle cette langue peut le percevoir différemment que s'il était formulé dans la langue nationale). Quant au concept de « discours attisant la peur » (Buyse 2014), il a été récemment mis en lumière afin d'examiner un type de langage susceptible de créer progressivement une mentalité d'assiégé et de conduire, au bout du compte, à légitimer des actes violents sous prétexte qu'ils sont commis pour assurer la sécurité d'un groupe ou protéger son intégrité. Fondée aussi sur l'étude des atrocités de masse, cette idée de « discours attisant la peur » nous permet de comprendre l'apparition progressive des conditions préalables au déclenchement de la violence et d'identifier, le cas échéant, les moments critiques où des contre-mesures pourraient être le plus efficaces.

Enfin, il y a eu quelques tentatives pour aller au-delà de l'identification, la réglementation des discours de haine et l'élaboration de contre-mesures : on a cherché en effet à comprendre qui étaient les personnes incitant à la haine et ce qui motivait leurs actes. Malheureusement, ce type de recherches demeure marginal, mais la permanence des contenus textuels et visuels qu'offre Internet multiplie les occasions de mener ces études. Karmen Erjavec et Melita Poler Kovačič (2012), par exemple, ont analysé les messages haineux contenus dans les commentaires sur les sites Internet d'information les plus populaires de Slovénie et ont pu interviewer certains de leurs auteurs. Leur méthode de recherche a permis d'identifier différentes catégories de locuteurs, chacune d'entre elles motivée par des facteurs spécifiques : des « soldats », qui appartiennent à des partis politiques ou à des organisations non-gouvernementales et qui utilisent systématiquement Internet pour diffuser des stéréotypes et porter atteinte à la réputation de leurs adversaires, jusqu'aux « gardiens » qui utilisent les discours de haine pour attirer l'attention des internautes sur les problèmes sociaux. Ce type de recherches aide à mieux comprendre ce qui pousse certains utilisateurs à recourir aux excès de langage. Voici, selon les auteurs de l'étude, comment les « soldats » justifient le recours aux discours de haine :

Ils affirment que les discours de haine en ligne ne peuvent pas être comparés aux discours de haine présents dans les médias traditionnels car ils sont, selon eux, le seul moyen possible de communication lorsqu'on fait des commentaires sur un site Internet : « C'est le seul moyen de communiquer ; sinon, on ne peut pas faire entendre sa voix. » Ainsi, ils justifient l'utilisation de propos haineux dans leurs commentaires en expliquant que des propos « cinglants », formulés sur un ton « militaire » sont « le seul moyen possible et convaincant » de communiquer, car c'est le seul que comprennent les « ennemis » (Erjavec & Kovačič 2012, p. 910).

Ces conclusions trouvent un écho dans les recherches menées sur « la perturbation sur Internet » (Buckels et al. 2014; Herring et al. 2002; Shin 2008), pratiquée par des utilisateurs délibérément provocateurs qui tiennent des propos incendiaires et diffusent des contenus dérangeants, et elles nous fournissent des indications permettant de comprendre comment le médium peut influencer le message. D'une manière plus générale, ces recherches visent à obtenir une meilleure compréhension des caractéristiques singulières et de certaines des causes d'un phénomène qui évolue rapidement, sans se limiter à chercher des « solutions » à ce phénomène.

## Ce qui caractérise les discours de haine en ligne

La prolifération des discours de haine en ligne, telle qu'elle a été observée par le Rapporteur spécial des Nations Unies sur les questions relatives aux minorités (HRC, 2015), pose de nouveaux défis. S'il n'existe aucune statistique offrant une vue d'ensemble du phénomène,

les plates-formes des réseaux sociaux tout comme les organisations créées dans le but de combattre les discours de haine ont reconnu, en revanche, que les messages haineux diffusés sur Internet sont de plus en plus courants, font l'objet d'une attention sans précédent et suscitent la mise en place de ripostes appropriées. Selon *HateBase*, une application Web qui répertorie des cas de propos haineux sur Internet à travers le monde, la plupart de ces cas visent des individus et sont fondés sur l'origine ethnique et la nationalité, mais les incitations à la haine fondées sur la religion ou la classe sociale sont également en augmentation.

Certaines mesures juridiques, élaborées pour d'autres médias, se révèlent inefficaces ou inadéquates lorsqu'on les applique aux discours de haine sur Internet. Pour les combattre efficacement, il est nécessaire de prendre en compte la nature spécifique des interactions propres aux technologies numériques de l'information et de la communication (TIC). Il faut notamment prendre en compte les risques de faire l'amalgame entre un tweet fulminant envoyé par un utilisateur qui n'a pas réfléchi aux conséquences de son acte et une menace réelle qui fait partie d'une campagne systématique d'incitation à la haine (Rowbottom 2012). Il faut aussi tenir compte de la différence qui existe entre un message qui ne suscite aucune réaction — ou très peu — et un message qui se propage parmi les internautes tel un virus. Il faut en outre prendre en considération la complexité des difficultés auxquelles sont confrontés les gouvernements et les tribunaux, notamment lorsqu'ils tentent de faire respecter une loi par une plate-forme de réseau social dont le siège se trouve dans un autre pays. Par conséquent, bien que les discours de haine ne soient pas intrinsèquement différents des autres discours semblables formulés hors Internet, il existe des problèmes propres aux contenus en ligne et à leur réglementation. Ces problèmes, liés à la **longévité**, l'**itinérance**, l'**anonymat** et le caractère **transnational** d'Internet, font partie des défis les plus complexes auxquels nous sommes confrontés :

La **longévité** des propos haineux sur Internet est liée à leur faible coût et à leur potentiel de réapparition immédiate. Les discours de haine peuvent rester en ligne pendant longtemps et sous différents formats sur de nombreuses plates-formes. Comme l'a noté Andre Oboler, le P-DG de Online Hate Prevention Institute : « Plus le contenu reste disponible sur le Web, plus les dommages causés aux victimes sont importants et plus le pouvoir des auteurs de ces messages est renforcé. Si l'on supprime le contenu très tôt, on peut limiter les risques. C'est comme si vous procédiez à l'enlèvement des ordures ; cela n'empêchera pas les gens de déposer leurs débris dans la rue, mais si l'on ne traite pas le problème, les ordures s'accumulent et la situation s'aggrave. »<sup>2</sup> En fonction de l'architecture d'une plate-forme, on peut développer un sujet pendant une période longue ou courte. Les conversations sur Twitter autour des sujets les plus tendance peuvent faciliter la diffusion rapide des messages haineux, mais elles offrent aussi la possibilité aux influenceurs d'ignorer certains messages et de mettre fin à des fils de discussion très suivis contenant des incitations à la haine. Sur Facebook, en revanche, de nombreux fils de discussion peuvent se dérouler en parallèle et passer inaperçus, permettant ainsi, pendant plus longtemps, que certains individus ou groupes soient ridiculisés, insultés ou fassent l'objet de discrimination.<sup>3</sup>

Les discours de haine en ligne sont aussi **itinérants**. Même si le contenu est supprimé, son auteur peut s'exprimer à nouveau ailleurs, éventuellement sur la même plate-forme mais sous un pseudonyme différent, ou bien sur d'autres cyberspaces. Lorsqu'un site Internet est fermé, on peut rapidement le rouvrir en utilisant un service d'hébergement dont la réglementation est moins stricte ou bien en hébergeant le site dans un pays où la législation est plus permissive à l'égard des discours de haine. La nature itinérante des messages haineux signifie que des idées mal formulées qui, par le passé, n'auraient pas attiré l'attention du public ni recueilli son soutien bénéficient actuellement d'une grande visibilité parmi les internautes.<sup>4</sup> Dans ce cas également, les architectures distinctes qui caractérisent les différentes plates-formes des réseaux sociaux favorisent ou restreignent diverses formes d'expression et de réactions. Du point de vue de l'utilisateur, Facebook se situe au croisement entre les sphères privée et publique, où les individus créent des espaces privés en ligne dans lesquels ils interagissent avec d'autres internautes en ligne. Et pourtant, un message partagé entre amis peut atteindre des publics auxquels on n'avait pas songé et avoir des conséquences imprévues. Twitter est plus clairement conçu comme un espace public qui offre la possibilité à un message d'être retransmis vers un large public. Des plates-formes telles que Snapchat, en revanche, en supprimant les conversations entre utilisateurs juste après qu'elles ont eu lieu, garantissent davantage que les propos demeurent à l'intérieur de cercles plus restreints.

L'**anonymat** pose également problème lorsqu'on traite des discours de haine en ligne. « Internet facilite l'expression anonyme et sous pseudonyme, qui peut aussi facilement accélérer les comportements destructifs qu'alimenter le débat public » (Citron & Norton 2011). Comme l'a déclaré Drew Boyd, directeur des opérations de The Sentinel Project : « Internet offre aux utilisateurs la possibilité de dire des choses épouvantables parce qu'ils croient qu'ils ne seront pas découverts. C'est ce qui rend les discours de haine en ligne si particuliers car leurs auteurs sont beaucoup plus à l'aise dans le monde virtuel que dans la vie réelle, où ils doivent faire face aux conséquences de leurs propos ». <sup>5</sup> Certains gouvernements et certaines plates-formes de réseaux sociaux ont tenté de faire appliquer des mesures exigeant que les internautes utilisent leur vrai nom. Mais elles ont été vivement contestées car elles se heurtent au droit du respect de la vie privée et au droit à la liberté d'expression. De surcroît, les actes de perturbation et les propos haineux en ligne proviennent, pour l'essentiel, de comptes sous pseudonymes qui ne sont pas nécessairement anonymes pour tout le monde.<sup>6</sup> Les communications sur Internet véritablement anonymes sont rares car elles nécessitent que l'utilisateur sache employer des méthodes très sophistiquées sur le plan technique lui permettant de dissimuler son identité.

La portée **transnationale** d'Internet constitue un obstacle supplémentaire à la lutte contre les discours de haine en ligne ; elle soulève en effet des problèmes de coopération transjuridictionnelle quant aux mécanismes juridiques visant à combattre les discours de haine. S'il existe des traités d'entraide judiciaire dans de nombreux pays, leur fonctionnement est extrêmement lent. La portée transnationale de nombreux intermédiaires privés de l'Internet pourrait fournir un moyen plus efficace pour résoudre certains problèmes, même si ces organismes sont souvent affectés par des demandes transjuridictionnelles de données

(notamment la véritable identité de l'auteur d'un contenu particulier). Contrairement à la diffusion des discours de haine par le biais des médias traditionnels, la diffusion des propos haineux en ligne implique souvent une multitude d'acteurs, sciemment ou non. Lorsque les auteurs de messages haineux utilisent une plate-forme en ligne pour les diffuser, ils ne causent pas seulement du tort à leurs victimes, ils enfreignent aussi souvent les conditions d'utilisation de cette plate-forme et parfois même la législation nationale. Les victimes, de leur côté, se sentent parfois démunies face au harcèlement en ligne, ignorant vers qui se tourner pour demander de l'aide. Lorsqu'on considère les différentes formes de réactions aux discours de haine examinées tout au long de cette étude, il en ressort que les actions collectives, généralement menées par des organisations non gouvernementales et par des groupes de pression, sont un moyen efficace de sensibiliser le public et d'inciter les différents acteurs concernés à agir.

## Les réactions aux discours de haine en ligne : mesures juridiques et non juridiques

Les réactions aux discours de haine sur Internet les plus débattues sont principalement axées sur la définition et les moyens juridiques, mais cette approche comprend des risques et des limites. Premièrement, il existe des liens inextricables entre législation et pouvoir. Comme l'a fait observer Robert Post : « La réglementation des discours de haine croit possible de faire respecter les normes données et naturelles d'une société respectable [...] ; mais d'un point de vue sociologique ou anthropologique, on sait que la loi fait toujours respecter les mœurs du groupe dominant qui contrôle le contenu de la loi. Pour chaque Machiavel qui exhorte la loi à interdire des discours contenant une incitation à la haine de l'État, il y aura un Walt Whitman qui nous exhortera à 'haïr les tyrans' » (Post et al. 2009). Remarquons, par exemple, que la législation relative aux discours de haine en Afrique du Sud, du temps de l'apartheid, servait à criminaliser toute critique de la domination du pays par les Blancs, ce qui prouve les risques d'abus que le pouvoir politique est susceptible de commettre en limitant les propos haineux.<sup>7</sup>

Deuxièmement, en examinant la question uniquement sous l'angle juridique, on risque d'omettre que les sociétés évoluent par le biais de la contestation et des dissensions. Bien que les discours de haine soient des formes d'expression du désaccord blessantes et méprisables, ils peuvent aussi être considérés comme le symptôme de tensions et d'inégalités profondément enracinées dans une société, lesquelles doivent être traitées au-delà des questions relevant purement de l'expression et au-delà des limites d'Internet.

Troisièmement, comme l'explique le Rapporteur spécial des Nations Unies sur les questions relatives aux minorités, les crimes motivés par la haine se produisent rarement sans qu'il y ait eu au préalable des manifestations de stigmatisation et de déshumanisation à l'égard de groupes ciblés et des incitations à commettre des actes inspirés par la haine. Dans le même

temps, « seuls les propos haineux les plus extrêmes, c'est-à-dire ceux qui constituent une incitation à la discrimination, à l'hostilité ou à la violence, sont généralement considérés comme illicites. » (HRC, 2015). Le Rapporteur souligne en outre qu'il importe de différencier trois catégories de discours : (a) le discours qui constitue une infraction selon le droit international et qui peut faire l'objet de poursuites pénales ; (b) le discours qui n'est pas passible de sanctions pénales mais peut justifier une restriction et une action civile ; (c) le discours qui ne donne pas lieu à des sanctions pénales ou civiles, mais qui suscite néanmoins des inquiétudes en matière de tolérance, de civilité et de respect d'autrui. Il ressort de ces réflexions que, tandis que la loi joue parfaitement son rôle en réprimant la catégorie (a), les mesures juridiques ne sauraient être considérées comme des réponses suffisantes pour combattre toute la diversité des propos susceptibles de créer un climat propice aux crimes haineux.

C'est dans ce contexte que la présente étude, tout en offrant un vaste aperçu des principaux instruments juridiques qui permettent de réglementer les discours de haine, met particulièrement l'accent sur la société civile et sur les initiatives sociales plutôt que sur les mesures juridiques engagées par les États. Parmi ces initiatives, citons l'exemple d'une communauté d'internautes qui s'est mobilisée pour contrer et marginaliser les messages haineux. Comme l'a déclaré le Rapporteur spécial des Nations Unies sur les questions relatives aux minorités : « Si l'on accorde beaucoup d'attention, à juste titre, aux mesures juridiques prises contre les discours de haine, il faudrait s'intéresser tout autant aux réactions non-juridiques et sociales. » (HRC, 2015).

Voici comment est structurée la présente étude. Après une brève discussion sur la méthodologie au chapitre 3, les débats principaux relatifs aux discours de haine sont examinés au chapitre 4, qui propose une cartographie des différentes réponses internationales et régionales aux discours de haine et s'interroge sur le nouveau rôle que jouent les entreprises dans la réglementation du discours en ligne. Nous expliquons, par exemple, dans quelle mesure les organisations internationales se sont lancées dans des processus consultatifs visant à identifier plus clairement et à combattre les discours de haine. Nous soulignons que la tendance à désigner les gouvernements comme les principaux titulaires de devoirs et de responsabilités au sein de la vie publique risque de minimiser le rôle des nouveaux cyberspaces privés qui doivent, eux aussi, remplir une fonction de nature publique, en influençant les modes et les contenus de communication. Les utilisateurs des réseaux sociaux tels que Facebook ou Twitter sont soumis non seulement à la législation nationale, mais aussi aux conditions générales d'utilisation de chaque réseau social et aux mécanismes que les réseaux sociaux mettent en place pour les faire appliquer.

Au chapitre 5, nous examinons quatre types d'initiatives lancées dans le but de contrer l'émergence et/ou la propagation des messages de haine : i) des efforts de recherche permettant de suivre l'émergence et la diffusion des discours de haine en ligne, en développant un système d'alerte précoce et des méthodes qui facilitent la distinction entre les différents types d'actes de parole ; ii) des actions coordonnées menées par des membres de la société civile qui cherchent à créer des coalitions nationales et internationales afin de répondre aux

menaces émergentes qui lient la haine en ligne à la violence hors ligne ; iii) des initiatives encourageant les réseaux sociaux et les fournisseurs d'accès à Internet à renforcer leur lutte contre les discours de haine en ligne ; et iv) des campagnes et des initiatives d'éducation aux médias afin de préparer les utilisateurs à interpréter les discours de haine et à réagir aux messages de haine. En nous appuyant sur ces cas et sur l'analyse des débats en cours relatifs aux réactions à envisager face aux discours de haine en ligne, nous concluons notre étude au chapitre 6 par une série de recommandations qui pourront être adoptées par différents acteurs pour élaborer des solutions pratiques et adaptées.

# 3. MÉTHODOLOGIE

Le présent rapport se concentre sur les questions relatives aux discours de haine sur Internet et aux réactions sociales à ce phénomène. De l'examen d'un grand nombre de données empiriques résulte une typologie approximative des différentes ripostes sociales les plus pertinentes, même si, dans la pratique, il peut exister quelques chevauchements entre ces catégories.

La stratégie de recherche qui sous-tend cette étude combine plusieurs techniques de collecte et d'analyse de données. Nous avons commencé nos recherches par une consultation approfondie de la documentation existante sur le sujet, en l'abordant sous diverses perspectives complémentaires, notamment à travers l'approche juridique, en observant comment on traite le problème des propos haineux dans différents pays et sur différents continents, ainsi qu'à travers les études ethnographiques qui analysent les comportements des utilisateurs de cyberspaces diffusant de la propagande haineuse. Compte tenu de la nouveauté du phénomène et de la rapidité de son évolution, notre analyse documentaire comprend aussi des articles qui ne proviennent pas de travaux universitaires, comme par exemple des extraits de blogs rédigés par des spécialistes de la question ou des articles publiés dans des revues spécialisées ou bien dans la presse en ligne. Cette première étape a pour objet de délimiter les contours du débat et d'identifier les questions les plus préoccupantes mises en évidence par différents acteurs, issus des organisations de la société civile, des gouvernements ou des entreprises privées.

La partie essentielle de ce rapport s'appuie sur une analyse de plusieurs cas montrant différents aspects du discours de haine sur Internet, et elle expose les stratégies mises en place pour les contrer, notamment à travers des actions organisées en ligne, souvent promues par des organisations de la société civile et des citoyens préoccupés. Des entretiens semi-structurés ont été réalisés auprès d'un vaste éventail d'acteurs concernés, comprenant des représentants des plates-formes de réseaux sociaux, y compris Facebook et Google, des membres des organisations de la société civile, des responsables politiques ainsi que des experts techniques. Nous avons également analysé des contenus produits par des organisations non gouvernementales (ONG) qui ont lancé des campagnes d'initiation aux médias visant à combattre les discours de haine en ligne, et examiné les conditions générales d'utilisation des plates-formes des médias sociaux en ligne, tels que Facebook, Twitter et YouTube. L'objectif était de comprendre comment fonctionnent précisément la surveillance et la gestion des contenus en ligne. En ce qui concerne les propos haineux et les élections, l'initiation aux médias et à la maîtrise de l'information et les efforts coordonnés des ONG, nous avons aussi effectué des comparaisons entre les méthodes de surveillances mises en places dans différents pays ; nous avons souligné que les campagnes d'initiation aux médias et à la maîtrise de l'information visaient divers publics et rendu compte des résultats ; enfin,

nous avons considéré les stratégies adoptées par les groupes ou coalitions luttant contre la discrimination pour faire pression sur les médias sociaux.

Vu l'étendue des diverses conceptions des discours de haine en ligne et celle des ripostes élaborées pour les combattre, nous avons posé les mêmes questions dans chaque cas. Le processus de collecte des données visait à comprendre l'équilibre délicat qu'il convient de maintenir entre surveillance et réglementation du discours d'un côté, et respect des droits fondamentaux, notamment celui de la liberté d'expression, de l'autre. Compte tenu de la complexité propre à Internet, à la fois dans les formes d'expression et leur réglementation, nous avons pris soin de souligner que l'espace virtuel n'a pas été traité comme un espace social à part.

## 4. CADRES NORMATIFS

Le débat autour des discours de haine aborde les enjeux controversés de la dignité, de la liberté d'expression, de la liberté et de la démocratie. Ce chapitre décrit des concepts clés qui portent sur les tensions qui existent entre discours de haine et liberté d'expression. Il examine les solutions que proposent les principaux instruments juridiques internationaux et régionaux pour résoudre ces tensions et les méthodes utilisées par les propriétaires des nouveaux espaces privés d'expression en ligne pour régler les propos haineux, dont chacun a sa propre définition. Ce sont ces divers cadres qui constituent fréquemment un contexte et servent de références propices à l'émergence de réponses sociales efficaces.

### Les discours de haine : étude des principes internationaux permettant de les identifier et de les combattre

Si les discours de haine ne sont pas mentionnés explicitement dans de nombreux documents et traités relatifs aux droits de l'homme, certains principes portant sur la dignité humaine et la liberté d'expression s'y réfèrent indirectement. Par exemple, la Déclaration universelle des droits de l'homme de 1948, rédigée en réaction aux atrocités de la Seconde Guerre mondiale, contient le droit à une égale protection de la loi selon l'article 7, qui stipule : « Tous ont droit à une protection égale contre toute discrimination qui violerait la présente Déclaration et contre toute provocation à une telle discrimination ».<sup>8</sup> La Déclaration universelle des droits de l'homme énonce aussi que tout individu a droit à la liberté d'expression, ce qui implique « le droit de ne pas être inquiété pour ses opinions et celui de chercher, de recevoir et de répandre, sans considérations de frontières, les informations et les idées par quelque moyen d'expression que ce soit ».<sup>9</sup> Si l'on réunit ces articles, on en déduit que tout individu a droit à la liberté d'expression et à une protection contre toute discrimination. En d'autres termes, chacun a le droit d'être protégé contre les discours de haine dans la mesure où ces discours incluent des objectifs discriminatoires (Morsink 1999).

La Déclaration universelle des droits de l'homme a permis d'établir un cadre en matière de protection des droits de l'homme, mais elle est non contraignante. Un certain nombre de documents contraignants ont donc vu le jour par la suite afin d'offrir une protection plus solide à la liberté d'expression et de renforcer la protection contre la discrimination. Parmi ces documents, le Pacte international relatif aux droits civils et politiques est le plus important et le plus exhaustif en matière de propos haineux ; à l'article 19, il mentionne en effet le droit à la liberté d'expression, et à l'article 20, il énonce que tout appel à la haine qui constitue une incitation à la discrimination, à l'hostilité ou à la violence est interdit par la loi. D'autres instruments juridiques internationaux plus adaptés contiennent des dispositions qui ont des

répercussions sur la définition des discours de haine et sur l'identification des moyens de les combattre, notamment la Convention pour la prévention et la répression du crime de génocide (1951), la Convention internationale sur l'élimination de toutes les formes de discrimination raciale (1969), et enfin, quoique dans une moindre mesure, la Convention sur l'élimination de toutes les formes de discrimination à l'égard des femmes (1981).

### *Les discours de haine et le Pacte international relatif aux droits civils et politiques*

Le Pacte international relatif aux droits civils et politiques est l'instrument juridique auquel on se réfère le plus couramment dans les discussions portant sur les discours de haine et leur réglementation, même si ce document ne mentionne pas explicitement le terme « discours de haine ». L'article 19, souvent considéré comme « la partie essentielle du Pacte » (Nowak 1993), garantit le droit à la liberté d'expression et inclut également des dispositions générales auxquelles doit se conformer toute restriction de ce droit afin d'être légitime.

L'article 19 est toutefois suivi de l'article 20, qui restreint expressément la liberté d'expression dans les cas d' « appel à la haine nationale, raciale ou religieuse qui constitue une incitation à la discrimination, à l'hostilité ou à la violence » (Leo et al. 2011). La décision d'inclure cette disposition, qui peut être considérée comme le reflet d'une conception particulière des discours de haine, a fait l'objet de vives contestations. Les annotations du projet du texte de 1955 révèlent des divergences d'opinions.<sup>10</sup> Tandis que certains pays estimaient que la clause de limitation plus générale à la liberté d'expression (à l'article 19, paragraphe 3) suffisait à traiter la question des discours de haine, d'autres plaidaient en faveur d'une disposition autonome (article 20) qui interdise expressément tout propos haineux constituant une incitation à causer du tort à autrui (Nowak 1993). Même lorsque le document final est parvenu à l'étape de la ratification, certains signataires du Pacte international relatif aux droits civils et politiques ont émis des réserves concernant l'article 20.<sup>11</sup> Il s'agit là d'une pratique courante : de nombreux pays formulent en effet des réserves lorsqu'ils signent des traités internationaux, néanmoins, ce fait met en lumière les tensions provoquées par l'article 20 et révèle à quel point il contraste avec l'article 19. Le Comité des droits de l'homme des Nations unies, créé pour superviser la mise en œuvre du Pacte international relatif aux droits civils et politiques, conscient de ces tensions, a tenu à souligner que l'article 20 était parfaitement compatible avec le droit à la liberté d'expression.<sup>12</sup>

Afin d'approfondir ces questions et leur application à l'expression sur Internet, examinons plus en détail le Pacte international relatif aux droits civils et politiques. L'article 19 (2) énonce que « toute personne a droit à la liberté d'expression ; ce droit comprend la liberté de rechercher, de recevoir et de répandre des informations et des idées de toute espèce, sans considération de frontières, sous une forme orale, écrite, imprimée ou artistique, ou par tout autre moyen de son choix ». <sup>13</sup> L'expression « de toute espèce » implique que les États ne peuvent pas

exclure des contenus indésirables du champ d'application de la protection du droit à la liberté d'expression (Nowak 1993).<sup>14</sup> De plus, l'allusion à « tout autre moyen de son choix »<sup>15</sup> permet d'étendre la liberté d'expression à toutes les nouvelles technologies, notamment Internet.<sup>16</sup> Pour clarifier encore davantage ce point, le Comité des droits de l'homme des Nations unies, dans son Observation générale n°34, se réfère spécifiquement à Internet et aux autres technologies mobiles et exhorte les États à « prendre toutes les mesures voulues pour favoriser l'indépendance de ces nouveaux moyens et garantir l'accès des particuliers à ceux-ci ».<sup>17</sup>

Dans le Pacte international relatif aux droits civils et politiques, le droit à la liberté d'expression n'est pas un droit absolu. Il peut être légitimement restreint par les États dans certaines circonstances bien définies :

3. L'exercice des libertés prévues au paragraphe 2 du présent article comporte des devoirs spéciaux et des responsabilités spéciales. Il peut en conséquence être soumis à certaines restrictions qui doivent toutefois être expressément fixées par la loi et qui sont nécessaires :

(a) Au respect des droits ou de la réputation d'autrui ;

(b) A la sauvegarde de la sécurité nationale, de l'ordre public, de la santé ou de la moralité publiques.<sup>18</sup>

Les restrictions ne doivent pas porter atteinte à l'essence même de ce droit.<sup>19</sup> À titre de principe général, les restrictions imposées en matière de droits de l'homme en vertu du Pacte « doivent constituer une exception à la règle et se limiter au minimum nécessaire afin de poursuivre le but légitime de sauvegarder les autres droits de l'homme établis dans le Pacte ».<sup>20</sup> Les restrictions imposées par les États peuvent inclure des propos en ligne selon l'article 19 (3) du Pacte international relatif aux droits civils et politiques, comme l'explique le Comité des droits de l'homme dans son Observation générale n°34 :

Toute restriction imposée au fonctionnement des sites Web, des blogs et de tout autre système de diffusion de l'information par le biais de l'Internet, de moyens électroniques ou autres, y compris les systèmes d'appui connexes à ces moyens de communication, comme les fournisseurs d'accès à Internet ou les moteurs de recherche, n'est licite que dans la mesure où elle est compatible avec le paragraphe 3. Les restrictions licites devraient d'une manière générale viser un contenu spécifique ; les interdictions générales de fonctionnement frappant certains sites et systèmes ne sont pas compatibles avec le paragraphe 3. Interdire à un site ou à un système de diffusion de l'information de publier un contenu uniquement au motif qu'il peut être critique à l'égard du gouvernement ou du système politique et social épousé par le gouvernement est tout aussi incompatible avec le paragraphe 3.<sup>21</sup>

Entre l'article 19 (3) et l'article 20, on remarque une différence entre les restrictions optionnelles ou obligatoires à l'exercice du droit à la liberté d'expression. L'article 19 (3) énonce que l'exercice de la liberté d'expression « peut en conséquence être soumis à certaines restrictions », dans la mesure où elles sont fixées par la loi et nécessaires à certains buts

légitimes. L'article 20 stipule, lui, que tout appel à certaines formes de haine qui constitue une incitation à la discrimination, à l'hostilité ou à la violence « est interdit par la loi ».

Malgré les indications sur la gravité des délits d'expression qui doivent être interdits selon l'article 20, le sujet demeure complexe.<sup>22</sup> Il existe notamment une zone grise qui empêche d'établir des distinctions claires entre (i) des propos haineux, (ii) des propos incitant à la haine et (iii) des propos haineux qui constituent tout particulièrement une incitation aux préjudices que sont la discrimination, l'hostilité ou la violence. Ainsi, tandis que les États ont l'obligation d'interdire des propos considérés comme « un appel à la haine qui constitue une incitation à la discrimination, à l'hostilité ou à la violence », conformément à l'article 20 (2),<sup>23</sup> la façon dont on doit interpréter celui-ci n'est pas clairement définie.<sup>24</sup> En conséquence, les restrictions de la liberté d'expression, fondées sur les dispositions du Pacte international relatif aux droits civils et politiques, pourraient donner lieu à des abus,<sup>25</sup> comme l'ont signalé certaines ONG.<sup>26</sup> Les Principes de Camden sur la liberté d'expression et l'égalité, élaborés par l'ONG Article 19 en concertation avec des spécialistes en matière de législation sur les droits de l'homme, définissent des critères précis visant à éviter une mauvaise application de l'article 20 (2), en déclarant que : « Les États ne doivent pas interdire la critique, ou les débats, sur des idées, des croyances ou des idéologies, des religions ou des institutions religieuses particulières tant que ces discours ne constituent pas un discours de haine ».<sup>27</sup> Il importe de souligner que l'article 20 doit faire l'objet d'une interprétation étroite (Eltayeb 2008) afin d'éviter tout usage abusif de ce concept. Comme l'indique Ghanea, « tout en reconnaissant que les discours de haine qui incitent à la discrimination sont interdits par la loi, il est nécessaire de bien doser le processus de riposte contre ces derniers afin de s'assurer que les sanctions adoptées à chaque étape soient réellement 'appropriées'. » (Ghanea 2008).

Étant donné la complexité de la situation et les risques d'un usage abusif des normes internationales en matière de restriction du discours légitime, les Nations Unies ont cherché à créer des espaces visant à promouvoir une compréhension commune de ce que sont les discours de haine et de la façon dont on doit les combattre, tout en tenant compte des sensibilités régionales, nationales et locales. Le Haut-Commissariat des Nations Unies aux droits de l'homme a notamment organisé une série de réunions consultatives qui ont conduit en 2012 à la mise en œuvre du Plan d'action de Rabat sur l'interdiction de « toute expression de haine nationale, raciale ou religieuse qui constitue une incitation à la discrimination, l'hostilité ou la violence. »<sup>28</sup> Le Plan d'action de Rabat reconnaît que, malgré les obligations pour les États signataires du Pacte international relatif aux droits civils et politiques, de nombreux cadres juridiques ne contiennent pas l'interdiction légale de ce type d'incitation et que certaines lois qui l'interdisent emploient une terminologie qui est en contradiction avec l'article 20 du Pacte international relatif aux droits civils et politiques.<sup>29</sup> Il propose également de fixer un critère préliminaire en six parties visant à identifier les messages haineux en prenant en compte le contexte, le locuteur, l'intention, le contenu, l'étendue du discours et la probabilité selon laquelle le discours puisse réellement inciter à causer du tort à autrui.<sup>30</sup> Toutefois, comme le souligne plus loin ce chapitre et plus généralement la présente étude, dans le cas des

discours de haine en ligne, l'accent mis par le Plan d'action de Rabat sur les acteurs au niveau national, et en particulier sur les États, risque de minimiser l'importance des plates-formes des réseaux sociaux, qui opèrent à un niveau transnational. Ces acteurs peuvent jouer un rôle crucial dans l'interprétation des discours de haine ainsi que dans l'autorisation ou la limitation de l'expression. En outre, le Plan de Rabat n'accorde que peu d'attention au problème de l'incitation à la haine pour des motifs liés au sexe, à l'orientation sexuelle ou à la langue.

### *Autres instruments juridiques internationaux*

La Convention internationale sur l'élimination de toutes les formes de discrimination raciale, entrée en vigueur en 1969, a également des incidences sur la conceptualisation des différentes formes de discours de haine. Cette Convention et le Pacte international relatif aux droits civils et politiques diffèrent sur trois points.

Premièrement, la Convention prend en compte les discours de haine uniquement lorsque ceux-ci font référence à la race ou l'origine ethnique.

Deuxièmement, elle affirme au paragraphe (a) de l'article 4 que les États parties :

Devront déclarer délits punissables par la loi toute diffusion d'idées fondées sur la supériorité ou la haine raciale, toute incitation à la discrimination raciale, ainsi que tous actes de violence, ou provocation à de tels actes, dirigés contre toute race ou tout groupe de personnes d'une autre couleur ou d'une autre origine ethnique, de même que toute assistance apportée à des activités racistes, y compris leur financement ;

Cette obligation imposée aux États parties par la Convention internationale sur l'élimination de toutes les formes de discrimination raciale est également plus stricte que l'article 20 du Pacte international relatif aux droits civils et politiques, qui couvre l'incrimination d'idées racistes qui ne sont pas nécessairement une incitation à la discrimination, à l'hostilité ou à la violence.

Troisièmement, les deux textes diffèrent considérablement sur la question de l'intention. Le concept d'« appel à la haine » contenu dans le Pacte international relatif aux droits civils et politiques est plus précis que les propos discriminatoires décrits dans la Convention internationale sur l'élimination de toutes les formes de discrimination raciale, dans la mesure où ce concept nécessite la prise en compte de l'intention de l'auteur et non l'expression prise isolément — parce que le terme « appel » est interprété dans le Pacte comme impliquant l'intention de semer la haine.<sup>31</sup> La simple diffusion d'idées fondées sur la supériorité ou la haine raciales, ou même toute incitation à la discrimination ou à la violence raciale, sont passibles de sanctions conformément à la Convention. Mais dans le Pacte international relatif aux droits civils et politiques, selon l'article 20 (2), il est nécessaire de prouver l'intention d'inciter à la haine pour que l'infraction soit sanctionnée.

Le Comité pour l'élimination de la discrimination raciale a vigoureusement dénoncé les discours de haine dans sa Recommandation générale n° 29, qui recommande aux États parties de :

- (r) Prendre des mesures contre toute diffusion d'idées prônant la supériorité ou l'infériorité liée à la caste ou tentant de justifier la violence, la haine ou la discrimination à l'encontre de communautés fondées sur l'ascendance ;
- (s) Prendre des mesures strictes contre toute incitation à la discrimination, à la violence contre les communautés, y compris par l'Internet;
- (t) Prendre des mesures pour sensibiliser les professionnels des médias à la nature et aux conséquences de la discrimination fondée sur l'ascendance ;<sup>32</sup>

Ces points, qui reflètent les références faites à la diffusion de l'expression par la Convention internationale sur l'élimination de toutes les formes de discrimination raciale, revêtent une grande importance pour Internet. Exprimer des idées dans certains contextes en ligne peut immédiatement revenir à les propager. Ceci concerne tout particulièrement les espaces privés qui ont commencé à jouer un rôle public, comme par exemple les plates-formes des réseaux sociaux.

Tout comme la Convention internationale sur l'élimination de toutes les formes de discrimination raciale, la Convention pour la prévention et la répression du crime de génocide vise à protéger des groupes définis par la race, la nationalité ou l'origine ethnique, bien qu'elle étende aussi ses dispositions aux groupes religieux. Cependant, en matière de discours de haine, la Convention sur le génocide ne s'applique qu'à des actes qui incitent publiquement au génocide, reconnus comme « des actes commis dans l'intention de détruire, ou tout ou en partie, un groupe national, ethnique, racial ou religieux », que ces actes soient commis en période de paix ou en période de guerre (Defeis 1992).

Les discours de haine fondés spécifiquement sur le sexe (contrairement aux actes discriminatoires) ne sont pas traités en profondeur par la législation internationale. La Convention sur l'élimination de toutes les formes de discrimination à l'égard des femmes, entrée en vigueur en 1981, impose aux États l'obligation de condamner la discrimination à l'égard des femmes sous toutes ses formes<sup>33</sup> et de « prévenir les actes de violence sexiste, d'enquêter sur ces actes et de les sanctionner ». <sup>34</sup> Le Comité des droits de l'homme a également exprimé « sa vive inquiétude face aux actes de violence et de discrimination commis dans toutes les régions du monde contre des individus en raison de leur orientation sexuelle ou de leur identité sexuelle. »<sup>35</sup> Dans quelle mesure l'expression haineuse est-elle liée à de tels actes ? Cette question fait débat. Néanmoins, le Comité des droits de l'homme des Nations Unies, dans son Observation générale n°28, enjoint les États parties à « fournir des renseignements sur les mesures légales prises pour limiter la publication et la diffusion » de matériels pornographiques qui présentent les femmes comme des objets de traitement dégradant.<sup>36</sup>

Cet examen des différents instruments juridiques, loin d'être exhaustif, révèle que la législation internationale permet aux États de prendre des mesures pour limiter le discours de haine. Il existe aussi certaines dispositions dans différents traités prévoyant que toute personne peut déposer une plainte contre des discours haineux auprès d'institutions visant à la protection des droits de l'homme : Le Comité des droits de l'homme reçoit et examine les plaintes individuelles relatives au Pacte international relatif aux droits civils et politiques ; le Comité pour l'élimination de la discrimination raciale reçoit et examine les plaintes relatives à la Convention internationale sur l'élimination de toutes les formes de discrimination raciale ; et dans le cas de la Convention sur l'élimination de toutes les formes de discrimination à l'égard des femmes, c'est le Comité pour l'élimination de la discrimination à l'égard des femmes qui est chargé d'examiner les plaintes. Notons cependant qu'on ne peut déposer une plainte que contre un État qui a expressément autorisé (par le biais de la ratification d'un protocole optionnel) le fonctionnement de ces mécanismes.

Le tableau ci-dessous permet de mettre en évidence les différentes catégories dans lesquelles certaines normes internationales rangent les propos haineux (d'après Ghanea 2013) :

| CATÉGORIE | Diffusion d'idées fondées sur la supériorité (raciale/ ethnique) et sur la haine                  | Appel à la haine qui constitue une incitation à l'hostilité, à la discrimination ou à la violence | Incitation au terrorisme   | Incitation au génocide   |
|-----------|---|---|--|--|
| EXEMPLE   | (voir la Recommandation générale n° 29 du Comité pour l'élimination de la discrimination raciale) | (voir l'article 20 du Pacte international relatif aux droits civils et politiques)                | (voir « l'incitation à commettre un ou des actes terroristes » à l'article 1 (1) de la Résolution 1624 (2005) du Conseil de sécurité | (voir « l'incitation directe et publique à commettre le génocide » à l'article 3(c) de la Convention de 1948 pour la prévention et la répression du crime de génocide) |

### *Mesures régionales contre les discours de haine*

Les différents avis sur la nécessité de maintenir un équilibre entre la liberté d'expression et les restrictions en matière de propos haineux trouvent leur concrétisation dans les instruments régionaux normatifs relatifs aux droits de l'homme. Ces documents complètent les traités internationaux en reflétant les particularités régionales qui ne sont pas mentionnées dans les traités universels. Ces instruments régionaux peuvent s'avérer particulièrement efficaces pour faire respecter la protection des droits de l'homme, comme par exemple la Cour européenne des droits de l'homme, qui statue sur davantage d'affaires de discours haineux que le Comité

des droits de l'homme des Nations Unies. Toutefois, les instruments régionaux relatifs aux droits de l'homme ne doivent pas être en contradiction avec les normes internationales, ni imposer des restrictions plus sévères aux droits fondamentaux. La plupart de ces instruments régionaux ne contiennent pas d'articles consacrés en particulier à l'interdiction des discours de haine, mais, de manière plus générale, ils autorisent les États à limiter la liberté d'expression – les dispositions s'appliquant à certains cas spécifiques. Nous examinerons dans ce chapitre dans quelle mesure le droit à la liberté d'expression et ses restrictions sont définis au niveau régional et de quelle façon les documents régionaux complètent les autres textes en vigueur qui permettent la définition et la limitation du discours de haine.

La **Convention américaine relative aux droits de l'homme** décrit les restrictions à la liberté d'expression dans des termes proches de ceux du Pacte international relatif aux droits civils et politiques à l'article 19 (3). La Convention a en outre ajouté une clause de limitation spécifique interdisant la censure préalable ; cependant, afin de mieux protéger les enfants, elle autorise la censure préalable dans le cadre de la « protection morale des enfants et des adolescents ».<sup>37</sup> L'Organisation des États américains a également adopté une autre déclaration de principes relative à la liberté d'expression, qui inclut une clause spécifique prévoyant que « l'assujettissement de l'expression à des conditions prédéterminées, telles la véracité, l'opportunité et l'impartialité, est incompatible avec le droit à la liberté d'expression reconnu dans les instruments internationaux. »<sup>38</sup> La Cour interaméricaine a recommandé que « l'abus du droit à la liberté d'information ne peut donc pas être contrôlé par des mesures préventives mais uniquement par l'imposition ultérieure de sanctions à l'encontre des coupables de ces abus. »<sup>39</sup> La Cour impose aux États qui souhaitent mettre en œuvre des restrictions à la liberté d'expression de satisfaire aux exigences suivantes : « a) l'existence de motifs propres à engager la responsabilité ; b) la définition expresse et précise de ces motifs par la loi ; c) la légitimité des objectifs visés ; d) une preuve établissant que ces motifs propres à engager la responsabilité sont 'nécessaires pour atteindre' les objectifs susmentionnés. »<sup>40</sup> Enfin, le système interaméricain dispose d'un Rapporteur spécial pour la liberté d'expression, qui a mené une étude approfondie sur les discours de haine. Il en a conclu que le système interaméricain des droits de l'homme différerait de l'approche des Nations Unies et de celle de l'Europe sur un point essentiel : le système interaméricain ne traite que des discours de haine qui conduisent réellement à la violence, et ce sont uniquement ces types de discours qui peuvent faire l'objet de restrictions.<sup>41</sup>

La **Charte africaine des droits de l'homme et des peuples** adopte une approche différente dans l'article 9 (2), en autorisant des restrictions de certains droits dans la mesure où elles entrent « dans le cadre des lois et règlements ». Ce concept a fait l'objet de critiques et il existe une vaste quantité de documents juridiques sur les clauses dites « de récupération » et leur interprétation (Viljoen 2012). Les critiques portent principalement sur le fait que les pays peuvent manipuler leur propre législation et affaiblir ainsi l'essence même du droit à la liberté d'expression. Il importe toutefois d'ajouter que la Déclaration de principes sur la liberté d'expression en Afrique énonce un niveau d'exigence plus élevé pour les restrictions à la liberté

d'expression. Elle affirme que le droit à la liberté d'expression « ne doit pas faire l'objet de restrictions pour des raisons d'ordre public ou de sécurité nationale sauf s'il existe un risque réel de dommages contre un intérêt légitime et qu'il y ait un lien étroit de causalité entre le risque de dommages et les propos exprimés ».<sup>42</sup>

En 1990, l'Organisation de la conférence islamique (renommée plus tard Organisation de la coopération islamique – OCI) a adopté la **Déclaration du Caire sur les droits de l'homme en Islam** qui, dans son préambule, énonce que les droits de l'homme doivent être « conformes à la Charria »<sup>43</sup> On peut estimer que cette clause risque d'avoir des répercussions sur le seuil des restrictions. Le principe selon lequel les droits de l'homme sont soumis à la Charria est la raison pour laquelle les États membres de l'OCI ont appelé à la pénalisation de tout discours qui, au-delà des cas de violence imminente, inclut « des actes de langage qui dénotent une intolérance ou une haine manifestes ».<sup>44</sup> Le droit à la liberté d'expression est exposé de la manière suivante à l'article 22 de la Déclaration du Caire sur les droits de l'homme en Islam :

- (a) Tout homme a le droit d'exprimer librement son opinion pourvu qu'elle ne soit pas en contradiction avec les principes de la Charria.
- (b) Tout homme a le droit d'ordonner le bien et de proscrire le mal, conformément aux préceptes de la Charria.
- (c) L'information est un impératif vital pour la société. Il est prohibé de l'utiliser ou de l'exploiter pour porter atteinte au sacré et à la dignité des prophètes ou à des fins pouvant nuire aux valeurs morales et susceptibles d'exposer la société à la désunion, à la désintégration ou à l'affaiblissement de la foi.
- (d) Il est interdit d'inciter à la haine ethnique ou sectaire ou de se livrer à un quelconque acte de nature à inciter à la discrimination raciale, sous toutes ses formes.<sup>45</sup>

La **Charte arabe des droits de l'homme**, adoptée par le Conseil de la Ligue des États arabes en 2004, inclut à l'article 32 des dispositions qui s'appliquent également à la communication sur Internet, garantissant « le droit à la liberté d'opinion et d'expression et le droit de rechercher, de recevoir et de répandre des informations par tout moyen, sans considération de frontières géographiques ».<sup>46</sup> La Charte autorise des restrictions d'une portée très large au paragraphe 2 : « Ces droits et libertés sont exercés dans le cadre des principes fondamentaux de la société ».<sup>47</sup> Cette position diffère de celle du Comité des droits de l'homme, dans son Observation générale n° 22, qui stipule que « la conception de la morale découle de nombreuses traditions sociales, philosophiques et religieuses ; en conséquence, les restrictions apportées à la liberté de manifester une religion ou une conviction pour protéger la morale doivent être fondées sur des principes qui ne procèdent pas d'une tradition unique ».<sup>48</sup>

La **Déclaration de l'ASEAN sur les droits de l'homme** inclut le droit à la liberté d'expression à l'article 23. L'article 7 de la Déclaration prévoit des restrictions générales, affirmant que « le respect des droits de l'homme doit être considéré à la fois sur le plan régional et national,

en tenant compte des différents contextes politiques, économiques, législatifs, sociaux, culturels, historiques et religieux. »<sup>49</sup> À cet égard, le Haut-Commissariat aux droits de l'homme a rappelé les dispositions de la Déclaration de Vienne qui affirme que, en dépit des différences de contextes, « les États ont le devoir, quels que soient leurs systèmes politiques, économiques et culturels, de promouvoir et de protéger tous les droits de l'homme et les libertés fondamentales ». <sup>50</sup>

Ce bref débat démontre que, comparés aux normes internationales, certains textes régionaux sont potentiellement plus restrictifs en matière de liberté d'expression. Dans d'autres textes régionaux, en revanche, les critères permettant d'évaluer quelles restrictions de la liberté d'expression peuvent être considérées comme légitimes sont encore plus étroits que dans les normes internationales.

La Charte des droits fondamentaux de l'Union européenne, dont l'article 11 proclame le droit à la liberté d'expression, contient une clause interdisant l'abus de droit. Selon cette clause, la Charte ne doit pas être interprétée comme impliquant « des limitations plus amples des droits et libertés que celles qui sont prévues par la présente Charte ». <sup>51</sup> On trouve un exemple de limitation qui implique des critères stricts de nécessité et de proportionnalité dans la disposition sur la liberté d'expression incluse dans la Convention européenne des droits de l'homme, qui souligne que l'exercice du droit à la liberté d'expression comporte des devoirs et des responsabilités. Par conséquent, il « peut être soumis à certaines formalités, conditions, restrictions ou sanctions prévues par la loi, qui constituent des mesures nécessaires, dans une société démocratique, à la sécurité nationale, à l'intégrité territoriale ou à la sûreté publique, à la défense de l'ordre et à la prévention du crime, à la protection de la santé ou de la morale, à la protection de la réputation ou des droits d'autrui, pour empêcher la divulgation d'informations confidentielles ou pour garantir l'autorité et l'impartialité du pouvoir judiciaire ». <sup>52</sup> La Cour européenne des droits de l'homme prend soin de bien faire la distinction entre les discours de haine et le droit de tout individu à exprimer librement ses opinions, même si d'autres personnes s'en offusquent. <sup>53</sup>

Il existe par ailleurs des instances régionales spécifiquement compétentes en matière de discours de haine sur Internet. Le Conseil de l'Europe a publié en 2000 une Recommandation de politique générale portant sur la lutte contre la diffusion de matériels racistes, xénophobes et antisémites par l'internet. <sup>54</sup> La création de la Convention sur la cybercriminalité par le Conseil de l'Europe en 2001, qui pose les règles de l'entraide judiciaire concernant les pouvoirs d'investigation, offre aux pays signataires un mécanisme permettant de traiter les données informatiques qui incluraient les discours de haine transnationaux en ligne. <sup>55</sup> En 2003, le Conseil de l'Europe a ajouté à la Convention sur la cybercriminalité un Protocole additionnel relatif à l'incrimination d'actes de nature raciste et xénophobe commis par le biais de systèmes informatiques. La convention et son protocole *étaient ouverts à la signature et à la ratification* par les pays situés à l'extérieur de l'Europe, et certains, tels le Canada et l'Afrique du Sud, ont ratifié cette convention. Le Protocole impose l'obligation aux États membres d'ériger en infraction pénale « l'insulte en public, par le biais d'un système informatique, (i)

d'une personne en raison de son appartenance à un groupe qui se caractérise par la race, la couleur, l'ascendance ou l'origine nationale ou ethnique, ou la religion dans la mesure où cette dernière sert de prétexte à l'un ou l'autre de ces éléments, ou (ii) d'un groupe de personnes qui se distingue par une de ces caractéristiques ».<sup>56</sup>

### *Résumé des dispositions internationales visant à limiter la liberté d'expression*

L'article 19 (3) du Pacte international relatif aux droits civils et politiques soumet les restrictions de la liberté d'expression à des critères stricts, qui peuvent être appliqués dans les cas où les systèmes juridiques régionaux relatifs aux droits de l'homme ont un spectre très large de restrictions. Dans le Pacte, la condition requise « prévu par la loi », par exemple, signifie que les restrictions à la liberté d'expression et d'information doivent être définies dans la législation formelle ou dans une norme équivalente non écrite du droit coutumier et doivent fournir des précisions suffisantes sur la licéité de l'ingérence des organes chargés de l'application de la loi ». (Nowak 1993). De surcroît, la question de la finalité est pertinente. Le Pacte relatif aux droits civils et politiques stipule que la liberté d'expression ne peut être soumise à des restrictions que si celles-ci sont « nécessaires au respect des droits ou de la réputation d'autrui » ou bien « à la sauvegarde de la sécurité nationale, de l'ordre public, de la santé ou de la moralité publiques ».<sup>57</sup> Comme l'explique le Comité des droits de l'homme dans son Observation générale n° 34 :

Des restrictions pour des motifs qui ne sont pas spécifiés dans le paragraphe 3 ne sont pas permises, même au cas où de tels motifs justifieraient des restrictions à d'autres droits protégés par le Pacte. Les restrictions doivent être appliquées exclusivement aux fins pour lesquelles elles ont été prescrites et doivent être en rapport direct avec l'objectif spécifique qui les inspire.<sup>58</sup>

On constate, dans le droit international, une tendance à souligner la réciprocité entre la liberté d'expression et la protection contre les discours de haine. Des documents récents, tels que l'Observation générale n° 34 du Comité des droits de l'homme et le Plan d'action de Rabat, ont mis l'accent à plusieurs reprises sur cette question. Ce dernier en donne un aperçu :

En vertu des normes internationales relatives aux droits de l'homme, qui doivent orienter la législation au niveau national, les propos étiquetés en tant que « discours de haine » peuvent faire l'objet de restrictions conformément aux articles 18 et 19 du Pacte international relatifs aux droits civils et politiques, si ces restrictions sont nécessaires au respect des droits d'autrui, à la sauvegarde de l'ordre public, voire même parfois à la sécurité nationale. Les États ont également l'obligation d' « interdire » tout propos qui constitue une « incitation » à la discrimination, à l'hostilité ou à la violence (conformément à l'article 20.2 du Pacte international relatifs aux droits civils et politiques et, sous différentes conditions,

conformément à l'article 4 de la Convention internationale sur l'élimination de toutes les formes de discrimination raciale).<sup>59</sup>

En résumé, assurer un équilibre entre la liberté d'expression et les restrictions apportées aux discours de haine est une question extrêmement complexe, à la fois sur le plan des lois internationales et sur celui de leurs équivalents régionaux. Ceci explique la diversité des conceptions juridiques des propos haineux à travers le monde et les difficultés que pose l'interprétation de la loi dans chaque cas donné. Ce qui est clair, c'est que toute restriction juridique doit toujours être considérée comme adjacente au droit, plus vaste, à la liberté d'expression et que « le rapport entre le droit et la restriction, entre la règle et l'exception, ne doit pas être inversé ». <sup>60</sup>

### *Espaces d'expression privés et discours de haine*

Les instruments juridiques internationaux et régionaux examinés jusqu'à présent offrent aux États un cadre permettant de combattre les discours de haine, conformément à leur devoir de promouvoir et de protéger les droits fondamentaux, et notamment de maintenir un équilibre entre le droit à la liberté d'expression et les droits à la dignité, à l'égalité et à la sécurité. Néanmoins, lorsqu'il s'agit de réprimer les discours de haine *sur Internet*, les États ne sont pas toujours les acteurs les plus efficaces. Les intermédiaires de l'Internet, tels que les plates-formes des réseaux sociaux, les fournisseurs de service Internet ou les moteurs de recherche, stipulent dans leurs conditions générales d'utilisation qu'ils sont susceptibles d'intervenir en permettant, restreignant ou canalisant la création de contenus spécifiques et l'accès à ces contenus. Une grande quantité d'interactions en ligne ont lieu sur les plates-formes des réseaux sociaux, interactions qui transcendent les juridictions nationales ; or ces plates-formes ont, elles aussi, élaboré leurs propres définitions des discours de haine et leurs propres mesures pour les combattre. Lorsqu'un utilisateur enfreint les conditions générales d'utilisation, le contenu de ce qu'il ou elle a posté peut être supprimé de la plate-forme ou bien la visibilité de ses messages peut être restreinte à une certaine catégorie d'utilisateurs (par exemple ceux qui vivent à l'extérieur d'un certain pays).

Les principes dont s'inspirent les conditions générales d'utilisation et les mécanismes que chaque entreprise élabore afin d'en garantir la mise en œuvre ont d'importantes conséquences sur la capacité des internautes à s'exprimer en ligne et sur leur protection contre les discours de haine.

La plupart des intermédiaires doivent engager des négociations avec les gouvernements nationaux, dont l'étendue varie en fonction du type d'intermédiaire, du lieu où se situe le siège de l'entreprise et du régime juridique applicable. Comme l'explique Tsesis, « si les transmissions sur Internet sont envoyées et reçues à certains endroits, des forums spécifiques conservent leur compétence pour poursuivre les activités illégales dans le cyberspace »

(Tsesis 2001). Les fournisseurs d'accès à Internet sont les plus directement concernés par la législation nationale parce qu'ils doivent être domiciliés dans un pays donné afin de pouvoir fonctionner. Les moteurs de recherche, quant à eux, qui peuvent modifier les résultats des recherches pour des raisons commerciales ou d'autorégulation, ont de plus en plus tendance à s'adapter au régime de responsabilité s'appliquant aux intermédiaires en vigueur dans les juridictions de leur siège social et également à celui en vigueur dans d'autres juridictions, celles des pays auxquels ils fournissent leurs services, soit en supprimant des liens vers des contenus de manière proactive ou bien à la demande des autorités (MacKinnon et al. 2015).

Tous les intermédiaires de l'Internet sont des entreprises privées, mais ils ont cependant le devoir de respecter les droits de l'homme. C'est ce qui est énoncé dans les Principes directeurs relatifs aux entreprises et aux droits de l'homme, élaborés par le Haut-Commissariat des Nations Unies aux droits de l'homme. Ce document met l'accent sur la responsabilité des entreprises dans la défense des droits de l'homme. Le principe 11 stipule : « Les entreprises devraient respecter les droits de l'homme. Cela signifie qu'elles devraient éviter de porter atteinte aux droits de l'homme d'autrui et remédier aux incidences négatives sur les droits de l'homme dans lesquelles elles ont une part. ».<sup>61</sup>

Pour résoudre ces questions, les intermédiaires de l'Internet, comme le font déjà d'autres entreprises, devraient « évaluer les incidences effectives et potentielles sur les droits de l'homme, regrouper les constatations et leur donner une suite, suivre les mesures prises et faire savoir comment il est remédié à ces incidences. »<sup>62</sup> Les Principes directeurs des Nations Unies indiquent également que, dans les cas où il y a violation des droits de l'homme, les entreprises devraient « prévoir des mesures de réparation ou collaborer à leur mise en œuvre suivant des procédures légitimes. ».<sup>63</sup> Dans le cas des intermédiaires de l'Internet, cela signifie qu'ils doivent s'assurer que des mesures sont mises en place pour fournir des réponses adéquates aux discours de haine.

Ces principes sont cependant encore trop rarement mis en œuvre dans la pratique quotidienne. Plusieurs facteurs y font obstacle. Tout d'abord, le fait qu'une société du secteur privé a le droit de fixer des conditions générales d'utilisation qui peuvent être plus restrictives en matière de liberté d'expression que les restrictions qu'un État est censé autoriser conformément aux normes internationales, telles que le Pacte international relatif aux droits civils et politiques. Cela s'apparente à certains égards à la situation de la presse, où chaque média a le droit d'établir sa propre politique éditoriale, alors que les médias sociaux reposent aussi clairement sur le discours des utilisateurs, contrairement aux plates-formes de médias d'information, où le discours provient des personnes employées par la plate-forme elle-même. Autre obstacle : comment les entreprises, dans la mesure où elles se conforment aux normes internationales relatives aux droits de l'homme, décident-elles de maintenir un équilibre entre la liberté d'expression et le respect de la vie privée, de l'égalité ou de la dignité ? Enfin, les entreprises ont du mal à prendre des décisions quand les lois nationales ne sont pas conformes aux normes internationales relatives aux droits de l'homme, notamment en matière de restrictions légitimes à la liberté d'expression.

Natasha Lomas a écrit, non sans polémique, sur le cas du harcèlement en ligne : « Twitter (qui signifie ‘gazouillement’) se borne à hausser les épaules face au problème du harcèlement en ligne, prend le temps de lisser ses plumes puis répond en gazouillant que ‘les tweets doivent suivre leur cours’ » (à moins bien sûr que l’entreprise ne soit contrainte par la loi à agir contre ce problème) ». <sup>64</sup> Face à de tels sentiments, les intermédiaires de l’Internet, en réaction à ce type de pression, se sont montrés disposés à reconsidérer certaines de leurs politiques et à accroître leur transparence. <sup>65</sup> D’une manière générale, la situation est dynamique et continue d’évoluer.

### *La définition et la réglementation des discours de haine au sein des espaces d’expression privés*

Les intermédiaires de l’Internet ont élaboré des définitions très diverses des discours de haine et mis au point des directives tout aussi diverses visant à les réglementer. Certaines entreprises n’utilisent pas le terme « discours de haine », mais disposent d’une liste descriptive de termes connexes. *Yahoo!*, dans ses conditions générales d’utilisation, interdit la diffusion de « tout contenu qui serait illégal, nuisible, menaçant, abusif, harcelant, retors, diffamatoire, vulgaire, obscène, calomnieux, constituant une violation de la vie privée d’un tiers, haineux ou répréhensible sur le plan racial, ethnique ou autre ». <sup>66</sup> De même, Twitter ne mentionne pas explicitement l’interdiction des discours de haine, mais avertit ses utilisateurs qu’ils « peuvent être exposés à des Contenus qui pourraient être offensants, blessants, inexacts ou inappropriés, ou, dans certains cas, à des messages mal titrés ou trompeurs. » On peut lire également dans les conditions d’utilisation de Twitter : « En aucun cas, Twitter ne pourra être tenue responsable de quelque manière que ce soit d’une perte ou d’un dommage quelconque, de quelque nature que ce soit, résultant de l’utilisation des Contenus, y compris, de manière non exhaustive, en cas d’erreur ou d’omission dans les Contenus, que ces Contenus soient affichés, transmis par courrier électronique, transmis ou rendus disponibles d’une autre manière au moyen des Services ou diffusés autrement. » <sup>67</sup> Ce texte est complété par les Règles de Twitter, à l’attention des utilisateurs, qui contient des restrictions telles que « Vous ne devez pas diffuser ou promouvoir de menaces violentes à l’égard d’autres personnes. ». <sup>68</sup> Comme nous l’expliquons plus en détail ci-dessous et au chapitre 5, Twitter a réagi aux demandes des gouvernements et des organisations de la société civile relatives aux discours de haine qui exigeaient d’agir sur les contenus.

D’autres entreprises font explicitement référence aux discours de haine. *YouTube*, par exemple, dans son « Règlement de la communauté », cherche à trouver un équilibre entre le respect de la liberté d’expression et les restrictions de certains types de contenus : « Nos produits sont des plates-formes sur lesquelles chacun peut s’exprimer librement. Cependant, nous n’approuvons pas les contenus qui tolèrent ou justifient la violence envers une personne ou un groupe en raison de son origine ethnique, sa religion, son handicap, son sexe, son âge, sa nationalité, son statut d’ancien combattant ou son orientation/identité sexuelle. » <sup>69</sup> Cette

définition est donc plus large que celle mentionnée dans le Pacte international relatif aux droits civils et politiques, qui interdit uniquement des discours de haine qui constituent un appel à la haine et une incitation à la discrimination, à l'hostilité ou à la violence. Voilà un cas de figure où une entreprise privée est plus restrictive que la législation internationale, et même que certaines lois régionales ou nationales sur les discours de haine.

*Facebook*, dans ses conditions d'utilisation, interdit tout contenu nuisible, menaçant ou qui risquerait d'inciter à la haine ou à la violence, en précisant : « Facebook supprime tout discours de haine, notamment tout contenu qui s'en prend directement à des personnes en raison de leur race, leur origine ethnique, leur appartenance nationale, leur religion, leur orientation sexuelle, leur sexe, leur identité sexuelle, leur handicap ou leur maladie ». <sup>70</sup> Un peu plus loin, on peut lire : « Nous autorisons l'humour, la satire ou la critique sociale en rapport avec ces sujets, et nous pensons que si les internautes utilisent leur véritable identité, ils adoptent un comportement plus responsable lorsqu'ils partagent ce type de commentaires. C'est pour cette raison que nous demandons que les propriétaires de pages Web associent leur nom et leur profil Facebook à tout contenu qui manquerait de respect à l'égard d'autrui, même si ce contenu n'enfreint pas nos règles. Comme toujours, nous incitons nos utilisateurs, lorsqu'ils partagent ce type de contenus, à prendre en compte le public destinataire ». <sup>71</sup>

*Microsoft* a établi des règles spécifiques concernant les discours de haine pour ses diverses applications. Dans le cas des téléphones portables, l'entreprise interdit les applications qui « incluent tout contenu prônant la discrimination, la haine ou la violence, fondées sur des considérations relatives à la race, l'origine ethnique, la nationalité, la langue, le sexe, l'âge, le handicap, la religion, l'orientation sexuelle, le statut d'ancien combattant ou l'appartenance à un quelconque groupe social ». <sup>72</sup> L'entreprise a également fixé un règlement pour les jeux en ligne, qui interdit tout contenu suggérant « des discours haineux, des sujets religieux faisant l'objet de controverses et des événements historiques ou d'actualité délicats ». <sup>73</sup> Il s'agit, là encore, d'un exemple où une entreprise privée est plus restrictive que la législation régionale ou internationale sur les discours de haine : « Les sujets religieux faisant l'objet de controverses et les événements historiques ou d'actualité délicats » ne sont pas nécessairement interdits dans le droit international, ni ne sont considérés automatiquement comme discriminatoires. Néanmoins, afin de promouvoir ce qu'elle considère comme une communauté virtuelle plus sécurisée, l'entreprise Microsoft a choisi d'imposer une réglementation restrictive sur certains des produits qu'elle offre. En revanche, ces conditions d'utilisation peuvent s'avérer plus permissives que les restrictions légales imposées par certaines juridictions.

## *Définition et mise en œuvre*

Généralement, seule une petite minorité d'utilisateurs lit les conditions générales d'utilisation<sup>74</sup> et il existe différents niveaux de « qualité » parmi les divers types de règlements.<sup>75</sup> Le défi majeur n'est probablement pas tant de savoir comment les intermédiaires de l'Internet définissent les discours de haine que comment ils appliquent leurs définitions. On se heurte ici au problème de la responsabilité de ces intermédiaires. Beaucoup d'entre eux soutiennent qu'ils ne génèrent pas ni ne contrôlent les contenus en ligne, et que par conséquent leur responsabilité devrait être limitée.<sup>76</sup> Cette interprétation leur permet de s'exempter du pré-filtrage ou de la modération des contenus et n'engage leur responsabilité qu'après publication, si leur attention est attirée par des contenus qui enfreignent la loi et/ou leurs conditions d'utilisation. Il existe dans le monde entier différents régimes de responsabilité, qui ont des impacts différents, mais à terme, il n'y aura probablement qu'une juridiction qui pourra ordonner légalement qu'une entreprise limite certains propos sur Internet.

La notion de responsabilité limitée différencie les intermédiaires de l'Internet des médias en ligne. Cependant, ce qui fait débat, c'est dans quelle mesure les médias devraient bénéficier d'une responsabilité limitée pour des commentaires générés par les utilisateurs sur leurs sites. Leurs pratiques et leurs conditions générales d'utilisation relatives à la modération des contenus, de même que leurs systèmes d'autorégulation tels que les conseils de presse pourraient en tout cas être d'importance pour les intermédiaires de l'Internet. Dans un cas où l'on a examiné des propos haineux xénophobes en Afrique du Sud, on a conclu en se demandant si les médias interactifs en ligne jouaient un rôle différent des médias traditionnels concernant toute décision relative à la restriction de la liberté d'expression.<sup>77</sup>

Quant aux fournisseurs de services Internet, leur responsabilité devant une juridiction donnée est relativement claire. Tout comme les autres intermédiaires de l'Internet, ils peuvent définir leurs propres paramètres lorsqu'ils offrent un service, mais comme ils sont tenus au principe de la territorialité, ils ont tendance à respecter la législation du pays dans lequel ils offrent leur service (Ryngaert 2008). Ce qui les rend plus réactifs que d'autres intermédiaires aux demandes extérieures de retrait de contenus particuliers (Goldsmith & Wu 2006; Kohl 2002).

La question est plus complexe dans le cas des plates-formes des réseaux sociaux, dont la portée est mondiale (par exemple Facebook ou Twitter). Étant donné le volume énorme de données qu'elles traitent, les plates-formes des réseaux sociaux se fondent principalement sur les notifications d'utilisateurs qui signalent des contenus qu'ils considèrent comme déplacés, choquants ou dangereux. Les plates-formes décident alors, la plupart du temps conformément à leurs conditions d'utilisation, si les contenus signalés doivent être supprimés ou non ou bien si elles doivent prendre d'autres mesures afin de restreindre l'accès à ces contenus ou d'empêcher leurs auteurs d'utiliser les services de la plate-forme. En l'absence d'un pouvoir juridictionnel couvrant plusieurs pays et ayant autorité sur l'entreprise, et compte tenu des

compétences limitées d'une seule juridiction (excepté celle du lieu où est domiciliée la plateforme), beaucoup d'intermédiaires ont tendance à fonctionner conformément à leurs propres conditions générales d'utilisation à l'échelle mondiale.

Hormis d'anciennes directives divulguées par les employés d'entreprises vers lesquelles les plates-formes de réseaux sociaux externalisaient en partie la réglementation des contenus,<sup>78</sup> on connaît peu de choses sur la façon dont les conditions d'utilisation se traduisent dans la pratique : qu'est-ce qui est conservé, filtré ou supprimé ? Certains ont suggéré que Facebook élaborait un ensemble de normes objectives visant à agir sur des propos qui, selon l'entreprise, risquent de susciter des actes violents.<sup>79</sup> Mais lorsqu'on l'a interrogée pour cette étude, Monika Bickert, responsable du contrôle des contenus sur Facebook, a indiqué que Facebook s'efforçait d'éviter une approche théorique et préférait, autant que possible, tenir compte du contexte.<sup>80</sup>

Certaines entreprises, au fil du temps, sont devenues plus attentives aux réclamations des utilisateurs. Ainsi, en 2012, Facebook a mis en place un système permettant de suivre le signalement de tout utilisateur qui attire l'attention sur un contenu jugé inapproprié, et ce jusqu'à ce que le problème soit réglé.<sup>81</sup> L'entreprise a également fourni aux utilisateurs des outils permettant d'améliorer la signalisation : l'utilisateur peut en effet avertir tout d'abord l'auteur d'un contenu particulier en privé avant de demander formellement à Facebook de supprimer ce contenu. Ces nouvelles possibilités offertes aux internautes sont des mesures intéressantes qui viennent s'ajouter aux autres méthodes déjà existantes visant à combattre les propos haineux, même si nous manquons pour le moment de preuves attestant de leur efficacité et du degré de satisfaction des utilisateurs à leur égard.

Autre exemple, YouTube aurait conféré le droit de signaler certains contenus à des agences britanniques qui cherchaient à lutter contre la propagande exercée par l'État islamique (EI) sur la plateforme.<sup>82</sup> En Malaisie, Google aurait supprimé un blog sur sa plateforme de blogueurs au motif que son contenu « dépassait les bornes car il contenait des propos incitant à la haine ou à la violence envers un groupe en raison de sa race, son origine ethnique, sa religion, son handicap, son sexe, son âge, son statut d'ancien combattant ou son orientation/identité sexuelle », règles que l'entreprise applique également à YouTube.<sup>83</sup>

D'une manière générale, comme le suggèrent quelques cas spécifiques analysés au chapitre suivant, tandis que les intermédiaires de l'Internet ont tendance à invoquer leurs conditions générales d'utilisation pour décider de la suppression ou du maintien d'un type particulier de contenu, ils font également preuve d'une certaine flexibilité en répondant aux exigences d'un large éventail d'acteurs. Twitter, par exemple, sous la pression populaire à la suite d'une affaire d'abus à l'égard d'une militante féministe au Royaume-Uni<sup>84</sup>, a fini par créer un bouton qui sert à signaler des tweets offensants, après s'être longuement opposé au préalable à l'ajout de cette fonctionnalité.<sup>85</sup>

## Conclusion

Dans ce chapitre, nous avons présenté une vue d'ensemble des législations internationales et régionales et évoqué la situation des entreprises transnationales intermédiaires de l'Internet. Différentes définitions des discours de haine sont apparues au sein de cette mosaïque disparate que forment les instruments internationaux, eux-mêmes appliqués de diverses manières par les gouvernements et les entreprises privées. Alors que les acteurs impliqués devraient tous privilégier les normes fixées par les traités universels, la réalité concrète est nettement plus diverse et complexifiée par la relative autonomie des intermédiaires de l'Internet et leur rôle essentiel dans les communications au sein du cyberspace.

Le présent rapport se tourne maintenant vers l'examen des réactions sociales aux discours perçus comme haineux sur les plates-formes en ligne.

# 5. ANALYSE DES RÉPONSES DE LA SOCIÉTÉ

Après des événements dramatiques, on en appelle souvent à des mesures plus restrictives et intrusives visant à mettre un frein à la capacité d'Internet de propager les incitations à la haine et à la violence, comme si les liens entre violence en ligne et violence réelle étaient bien connus. Or, comme le montre l'exemple qui suit, les apparences sont souvent trompeuses.

*Stormfront* est considéré comme le premier site d'incitation à la haine (Meddaugh et Kay, 2009). Lancé en mars 1995 par un ancien chef du Ku Klux Klan, il est rapidement devenu un espace populaire de débat d'idées liées au néo-nazisme, au nationalisme et au séparatisme blancs, d'abord aux États-Unis puis dans le monde entier (Bowman-Grieve, 2009). Le forum, où sont proférées toutes sortes d'ignominies, notamment des appels à une guerre sainte raciale et des incitations à recourir à la violence pour résister à l'immigration (Bowman-Grieve, 2009), est considéré comme un espace de recrutement de militants et, dans une certaine mesure, de coordination d'actions violentes (Chang *et al.* 2014)1,18]], "season": "06:55:30"}]], "schema": "https://github.com/citation-style-language/schema/raw/master/csl-citation.json"} . Les quelques études qui ont cherché à savoir qui étaient vraiment les utilisateurs de *Stormfront* brossent cependant un tableau plus complexe. Certains militants d'extrême droite connus, loin de considérer le forum comme un espace de coordination d'actions, l'accusent de ne représenter qu'un rassemblement de « guerriers du clavier ». L'un d'eux, par exemple, cité par De Koster et Houtman, explique qu'à la lecture de plusieurs articles un peu partout sur le forum, ce qui le frappe est le fait que l'on parle beaucoup mais que l'on agit peu. La section « militantisme et politique » lui semble elle-même parfaitement ridicule, sans parler des réunions, auxquelles n'assistent que quatre personnes tout au plus (De Koster et Houtman, 2008). Ce que rétorquent les membres actifs du site en réponse à ces accusations est encore plus révélateur. L'un d'eux explique par exemple qu'il a le droit d'exprimer son opinion sans être obligé de passer à l'acte. Il ne participe à aucune manifestation et n'est adhérent d'aucun parti politique ; si cela fait de lui un guerrier du clavier, poursuit-il, peu importe, cela lui convient très bien et il n'en a pas honte (De Koster et Houtman, 2008). Si De Koster et Houtman n'ont fait porter leur enquête que sur une section nationale de *Stormfront* et sur un échantillon non représentatif d'utilisateurs, des déclarations comme celles-ci devraient pour le moins inciter à prendre avec circonspection les hypothèses de lien entre parole et action, même dans les espaces dont la fonction principale est d'offrir un lieu d'expression aux idées extrémistes.

Dans ce contexte, l'analyse présentée dans les chapitres qui suivent cherche à offrir un panorama nuancé des différents moyens d'action envisageables pour prévenir les discours de haine et la violence. Le premier chapitre met l'accent sur les efforts destinés à répondre à

la menace du discours de haine sur Internet, et à mettre au point des systèmes et méthodes d'alerte précoce permettant de faire la distinction entre les différentes typologies des actes de langage. Le deuxième chapitre porte sur les actions coordonnées menées par des membres de la société civile en vue de créer des coalitions nationales et internationales dans le but de contrer les nouvelles menaces reliant haine et violence. Le troisième chapitre présente des exemples où des groupes de pression ont interagi au travers des plateformes en ligne, notamment les réseaux sociaux, pour lutter plus vigoureusement contre le discours de haine sur Internet. Le quatrième et dernier chapitre met l'accent sur l'éducation de base aux médias et à l'information, et sur la façon dont cette éducation a été intégrée par les établissements d'enseignements, les autorités locales et autres institutions en vue de préparer les utilisateurs à interpréter les messages de haine et à y réagir.

## Suivi du discours de haine et discussions à son sujet

Le climat est généralement plus favorable au développement du discours de haine dans des situations où les enjeux politiques sont importants, par exemple en période électorale. Les rumeurs et accusations qui circulent à ce moment-là peuvent créer les conditions propices aux appels à la violence. On s'intéressera, dans le présent chapitre, aux questions d'ordre général que soulèvent les mesures concrètes mises au point pour éviter la prolifération des propos haineux sur Internet en pareilles circonstances. Une initiative en particulier se prête bien à des observations d'ordre général : le projet de recherche UMATI, lancé en septembre 2012 (avant les élections kényanes de mars 2013), qui a suivi le discours en ligne au Kenya en vue d'évaluer la fréquence et la virulence des propos diffamatoires. Les expériences ainsi menées ont donné aux différentes parties prenantes l'occasion d'analyser quels étaient les problèmes soulevés et les personnes ciblées, et de réfléchir collectivement au potentiel de certains actes de langage de conduire à la violence. On voit toute l'importance de cette initiative à l'observation faite par la Rapporteuse spéciale sur les questions relatives aux minorités : « Peu de pays recueillent des données sur les crimes de haine, leurs causes et les victimes, qui permettraient aux responsables politiques de mieux protéger les groupes de population à risque » (HRC, 2015).

En 2007, le Kenya a tenu les élections les plus contestées et les plus violentes depuis son retour au multipartisme en 1991, avec plus de 1000 morts et 600 000 personnes déplacées<sup>86</sup>. De nombreux observateurs ont été surpris qu'un pays qui avait été salué pour sa scène politique dynamique et sa prospérité économique puisse assister à de tels niveaux de conflits. D'autres ont invoqué des tensions très ancrées (Kanyinga, 2009). Ce furent les premières élections où les nouvelles technologies de la communication et de l'information ont fait partie intégrante de la course électorale (Cheeseman, 2008). Les médias sociaux, les courriels et les SMS ont été utilisés à des niveaux sans précédent pour rassembler les sympathisants et diffuser l'information, mais aussi pour faire courir des rumeurs. Dans les lieux les plus divisés, comme les bidonvilles, les rumeurs se sont largement répandues, des individus appartenant

à différents groupes politiques et ethniques expliquant comment leurs opposants planifiaient des attaques, des assassinats, et des expulsions d'individus et de communautés (Osborn, 2008). De faux documents ont été fabriqués et diffusés en ligne pour semer le doute sur les candidats à la présidentielle. Les technologies de la communication et de l'information ont continué de jouer un rôle pendant la période post-élections, anciens et nouveaux médias étant utilisés pour diffuser des appels à la violence et coordonner des actions (Sommerville, 2011). Face à la violence, le Kenya a mis en place une commission nationale pour la cohésion et l'intégration, qui a travaillé de concert avec les médias et les forces de l'ordre pour lutter contre les problèmes de tensions ethniques.

Intéressé par ce contexte, un groupe de chercheurs et d'entrepreneurs s'est réuni en amont des prochaines élections, celles de 2013, et a lancé UMATI (qui veut dire « foule » en Kiswahili), projet visant à suivre les occurrences de propos diffamatoires sur Internet. L'objectif global d'UMATI était de détecter les signes de tensions qui montaient parmi les citoyens kényans afin de montrer les principales défaillances des différentes phases du processus électoral, et de tirer le signal d'alarme avant qu'il ne soit trop tard.

Les élections se sont déroulées en mars 2013, et le projet a duré neuf mois, entre septembre 2012 et mai 2013. Il a suivi les contenus publiés par les kényans sur les blogs, forums et journaux en ligne ainsi que sur Facebook et Twitter, en anglais, ainsi que dans les principales langues parlées au Kenya, à savoir le kikuyu, le luhya, le kalenjin, le Luo, le kiswahili, le sheng (argot parlé principalement dans les zones urbaines) et le somali. S'appuyant sur la définition du discours dangereux élaborée par Susan Benesch (2012), qui le considère comme la sous-catégorie de discours de haine présentant le plus fort potentiel de catalyse de la violence, l'équipe d'UMATI a défini des critères pratiques pour faire le tri parmi les différents actes de langage et évaluer ce potentiel. Elle a évalué les questions abordées en fonction de l'influence du locuteur sur la communauté en ligne, du contenu des déclarations, et du contexte social et historique dans lequel s'inscrivait le discours. L'application de cet ensemble de critères a permis de classer les actes de langage en trois catégories : propos blessants, discours modérément dangereux et discours dangereux. Le suivi quotidien, la possibilité de replacer les différents actes de langage dans un contexte global et de prendre en compte d'autres variables, notamment les cibles des propos diffamatoires et le lien ou non des actes de langage avec des événements spécifiques, ont permis de suivre l'évolution du discours de haine dans le temps et de comprendre de manière plus nuancée les risques réels et perçus.

Les conclusions du projet UMATI par rapport aux cas de violence ou à leur absence pendant les élections kényanes de 2013 montrent bien, plus largement, combien il est complexe d'établir un lien entre des propos tenus sur Internet et un passage à l'acte. Contrairement au contexte électoral précédent, les élections de 2013 se sont en fait globalement déroulées dans le calme. Non pas que le discours de haine ait été moins corrosif ou généralisé. Malgré l'absence de point de référence qui permettrait d'effectuer des comparaisons précises, en 2013, le projet UMATI a tout de même repéré des cas graves, retentissants et persistants de propos diffamatoires et d'appels à la violence (enquête iHub, 2013). Ces actes de langage,

néanmoins, ne se sont pas directement traduits par de la violence sur le terrain. Comme l'équipe l'a suggéré, d'autres facteurs ont probablement pesé plus lourd dans la balance de la violence ou du pacifisme. Les nombreux appels à la paix, émanants de différents secteurs de la société, notamment des médias, groupes religieux, et hommes politiques des différents bords de l'échiquier politique ont créé un climat où les actes de violence étaient durement condamnés (enquête iHub, 2013).

Le projet UMATI a aussi offert l'occasion de comparer la façon dont le public percevait le discours de haine, par rapport aux universitaires et cercles politiques. Dans le cadre d'une enquête menée parmi des kényans, le projet a montré que la majorité de ceux qui ont participé à la recherche considéraient les insultes personnelles, la propagande et les commentaires négatifs sur les politiciens comme un discours de haine. Leur notion de discours de haine était, de même, plus large que la définition donnée dans la Constitution kényane de 2010, qui interdit, à l'article 33, « la propagande en faveur de la guerre, l'incitation à la violence, le discours de haine ou l'apologie de la haine qui constitue une incitation au racisme, un dénigrement des autres ou une incitation à leur faire du mal ». Comme Nanjira Sambuli, chef du projet UMATI, l'a expliqué, connaître la notion de discours de haine telle qu'elle est perçue par les kényans permet non pas seulement de débattre de la signification de ce discours, mais aussi de le replacer dans un contexte plus large. Comme elle l'a dit :

« Les médias nous ont demandé de débattre du discours de haine, et nous avons dû expliquer la différence entre les propos qui peuvent causer du tort et les simples insultes à l'encontre des chefs politiques. Mais cela a aussi été l'occasion d'avoir des discussions plus générales sur la liberté d'expression. Ainsi, nous avons été invités à parler du discours de haine, mais nous avons fini par parler de liberté<sup>87</sup> ».

Enfin, le projet a montré comment les différentes plateformes de réseaux sociaux pouvaient offrir divers moyens de diffusion du discours de haine et de lutte contre celui-ci. Seulement 3 % des propos diffamatoires recensés par UMATI avaient été publiés sur Twitter, contre 90 % sur Facebook. Le rapport final d'UMATI donne des éléments d'explication :

« L'architecture de Facebook permet de lancer des conversations sur un sujet particulier et plus encore, de poursuivre sur ce sujet. Les fils de discussion, groupes et pages, qui ont tous des durées de vie différentes, le permettent. À l'inverse, sur Twitter, les sujets abordés ont une durée de vie proportionnelle à leur popularité. Avec l'utilisation des hashtags, les sujets les plus débattus sont les sujets « tendance », qui gagnent en exposition sur Twitter. En revanche, dès qu'un sujet plus populaire se fait jour, il y a des chances qu'il prenne la place du sujet tendance du moment, écourtant sa durée de vie. Contrairement à ce qui se passe sur Twitter, les groupes et pages Facebook permettent à des sujets d'exister quel que soit le niveau d'activité associé » (enquête iHub, 2013).

L'architecture de Facebook permet aussi aux utilisateurs d'adopter différents comportements d'un endroit à l'autre. Un utilisateur peut avoir un profil personnel « propre », tout en postant des messages de haines sur des groupes et pages spécifiques. Sur Twitter, au contraire,

toutes les publications d'un utilisateur sont conservées sous un seul répertoire d'information, et peuvent être vues par toutes les personnes qui suivent cet utilisateur.

Pour ce qui est de la remise en cause du discours de haine, le projet a montré comment les différentes plateformes permettaient des réponses différentes et plus ou moins efficaces. Souvent, les tweets considérés comme inacceptables étaient dénigrés et leurs auteurs publiquement tournés en ridicule. Dans certains cas, le contrevenant était même forcé à revenir sur ses propos à cause de la réaction du public, et/ou de fermer tout bonnement son compte Twitter. Comme le rapport l'a conclu « l'architecture particulière des fils de conversation sur Twitter facilite [ce type de réponse], étant donné que toutes les publications sont enregistrées sur un seul fil et sont visibles par tous » (enquête iHub Research, 2013). On est moins susceptible de trouver ce type de réaction sur Facebook, l'architecture de la plateforme cloisonnant davantage les conversations et les rendant moins accessibles à un vaste public.

Cette étude ne s'intéresse que peu au rôle des médias d'information vis-à-vis du discours de haine, étant donné que la dénonciation de ce type de discours n'est généralement pas une incitation à la discrimination, l'hostilité ou la violence, ou une apologie de celles-ci, mais bien un service d'intérêt public. Néanmoins, les institutions médiatiques éprouvent tout de même fréquemment le besoin de répondre aux propos tenus par les utilisateurs sur leurs plateformes en ligne – par exemple, dans la section commentaires des articles, ou là où s'affichent les sms et tweets publiés en direct. Différents systèmes sont en place et différentes pratiques ont été analysées en profondeur dans le cadre de deux études. L'une, menée par l'Institut albanais des médias, est un compte-rendu des mesures juridiques et institutionnelles en Europe du Sud-Est, qui offre aussi des conseils de bonnes pratiques pour une autorégulation optimale. La deuxième étude, intitulée « Online comment moderation: emerging best practices », réalisée par l'Association mondiale des journaux et des éditeurs, analyse les pratiques de 104 organismes de presse de 63 pays (Goodman et Cherubini, 2013). Traiter le flux dynamique des messages des usagers sans restreindre la liberté d'expression représente en effet une gageure qui exige d'élaborer des politiques basées sur la définition de ce que chaque institution considère comme des propos diffamatoires afin que des réponses adaptées puissent être envisagées – notamment des réponses qui encouragent à argumenter contre le discours de haine et à lancer un débat. Des procédures types sont nécessaires – comme dans le cas des émissions de radio qui prennent les appels des auditeurs. Tout ceci requiert qu'un système de surveillance soit mis en place par chaque institution médiatique, même s'il s'agit d'un système très rudimentaire où les lecteurs marquent et signalent les incidents pour que les éditeurs de la plateforme étudient ensuite les choses de plus près. Les pratiques de surveillance du discours de haine en ligne dans les médias d'information et de discussions sur celui-ci pourraient être mises en commun avec profit avec les sociétés intermédiaires Internet, malgré le positionnement différent de ces deux types d'entités.

## Mobiliser la société civile

L'expérience du Myanmar offre un exemple de solutions positives que la société civile peut mettre en place pour sensibiliser le public et contrer les voix de la haine. Le pays est en transition vers une plus grande ouverture et l'accès à Internet a connu une croissance sans précédent. Dans ce contexte, néanmoins, les médias sociaux ont souvent été utilisés par certains pour diffuser des appels à la violence<sup>88</sup>. En 2014, la Rapporteuse spéciale du Conseil des droits de l'homme sur les questions relatives aux minorités a exprimé son inquiétude face à la propagation de la désinformation, du discours de haine et de l'incitation à la violence, à la discrimination et à l'hostilité dans les médias et sur Internet, en particulier à l'égard d'une communauté minoritaire (HRC, 2015). La tension croissante observée en ligne est allée de pair avec des cas de violence réelle qui ont fait des centaines de morts et des milliers de déplacés, bien que, comme il est montré dans cette étude, il serait simpliste de chercher un lien direct de cause à effet entre le discours en ligne et le passage à l'acte dans la réalité<sup>89</sup>.

Après avoir adopté une nouvelle constitution en 2008 et tenu des élections en 2010, le Myanmar s'est lancé sur une nouvelle voie, vers une ouverture et une inclusion sociale plus grandes. Le gouvernement, dirigé par le Président Thein Sein, ancien commandant militaire, a permis des réformes dans des secteurs clés, notamment les médias, où des espaces de débat sans précédent ont pu se développer. En 2013, seulement 1,2 % de la population avait accès à Internet, et un peu plus de 12 % à la téléphonie mobile (en 2009, ce pourcentage était inférieur à 1 %). Cependant, les deux entreprises qui ont remporté les contrats de développement de l'infrastructure pour les technologies de l'information et de la communication au Myanmar, les sociétés qatari Ooredoo et norvégienne Telenor, se sont engagées à atteindre une couverture pour le réseau mobile de 90 % en cinq ans<sup>90</sup>.

L'une des difficultés pour y parvenir concerne les minorités ethniques et religieuses. En 2013, 43 personnes ont été tuées lors d'affrontements qui avaient éclaté après un différend dans l'État de Rakhine, dans la partie occidentale du pays<sup>91</sup>. Une année plus tôt, plus de 200 personnes avaient été tuées et des milliers déplacées en raison de violences ethniques déclenchées par un cas supposé de viol<sup>92</sup>. Dans ce contexte, l'émergence rapide de nouveaux espaces en ligne, bien qu'ils ne concernent qu'une partie de la population, a fait écho sous une forme nouvelle à certaines de ces tensions fortement enracinées. Faire face à l'intolérance et au discours de haine sur Internet est une question nouvelle.

Facebook est rapidement devenue la plateforme favorite des citoyens faisant leurs premiers pas en ligne. Dans cet environnement, certains individus et groupes ont prôné un usage plus agressif de ce média, surtout lorsqu'ils se sentaient protégés par le sentiment d'agir pour le bien et pour la défense de ce qu'ils prétendaient être l'intérêt national. Des figures politiques ont aussi utilisé les médias en ligne pour des causes particulières. Sur les médias sociaux, des termes péjoratifs ont été utilisés pour faire référence aux minorités<sup>93</sup>.

Face à cette situation complexe, divers acteurs ont commencé à se mobiliser, cherchant à proposer des solutions susceptibles d'éviter la montée de la violence. Facebook s'est efforcé de jouer un rôle plus actif dans la surveillance des utilisations de sa plateforme au Myanmar, en formant des partenariats avec des organisations locales et en faisant traduire en langue birmane les lignes directrices élaborées pour signaler les problèmes<sup>94</sup>. Le ministre de l'information du Myanmar, Ye Htut, s'est engagé à prendre des mesures supplémentaires pour lutter contre le discours de haine en ligne et a fait part du souhait de son pays de tisser des liens plus étroits avec les États-Unis en vue de trouver des moyens efficaces de limiter le discours de haine en ligne<sup>95</sup>. Les solutions créatives trouvées par la société civile locale sont analysées ci-après.

La société civile locale s'est élevée avec vigueur contre le discours de haine en ligne en condamnant ouvertement la propagation, tout en appelant à trouver d'autres recours que la censure. Parmi les solutions les plus innovantes, on citera le Panzagar, qui en birman signifie littéralement « discours fleuri » - une campagne lancée par le blogueur et militant Nay Phone Latt pour s'opposer ouvertement au discours de haine. Le but de l'initiative était d'offrir un joyeux exemple de la façon dont les personnes peuvent interagir, aussi bien en ligne qu'en dehors. Les fleurs sont très symboliques au Myanmar. Comme l'a expliqué Nay Phone Latt, « si le public comprend le message, il s'opposera à ceux qui utilisent un discours de haine dangereux. Nous voulons aussi progressivement convaincre les groupes extrémistes qui propagent le discours de haine d'arrêter de le faire »<sup>96</sup>. La campagne a encouragé les utilisateurs de Facebook à actualiser leur profil en présentant une photo d'eux une fleur à la bouche.

La campagne a reçu beaucoup d'attention, tant au niveau national qu'international, mais comme l'ont reconnu certains militants interrogés pour cette étude, le message véhiculé doit s'implanter parmi ceux qui vivent dans les zones rurales et parmi les moins éduqués. Des coalitions positives doivent être formées, et les chefs religieux largement respectés doivent être raliés à la cause. Par ailleurs, on se rend bien compte qu'il ne suffit pas d'encourager à « le dire avec des fleurs », mais qu'il faut aussi dénoncer la violence. Les militants sont conscients de la nécessité de clarifier les limites de ce qui peut ou ne peut pas être dit, ainsi que le rôle de l'État dans la résolution de ce problème.

Si des initiatives telles que Panzagar ont permis de rassembler différents groupes autour d'une même cause, les groupes de la société civile ne sont pas forcément unanimes quant aux solutions à adopter face au phénomène du discours de haine. Myat Ko Ko, par exemple, a expliqué combien les opinions pouvaient être différentes, voire contrastées, entre ces différents groupes. Certains sont contre l'adoption de lois plus punitives ; d'autres y sont favorables. Parallèlement, poursuit-il, « ils ont peur que le discours de haine puisse compromettre une transition pacifique. Ils sont conscients que le pays est fragile. Si la solution vient de la société civile, celle-ci y sera attachée. C'est elle qui en prendra le contrôle. Si les mesures sont prises au niveau législatif, c'est l'État qui assurera le contrôle »<sup>97</sup>.

Les militants locaux se sont concentrés sur des solutions locales, plutôt que d'essayer de mobiliser la société civile dans son ensemble sur ces questions – contrairement à d'autres campagnes en ligne qui ont réussi à attirer l'attention de la communauté internationale sur des problèmes relativement négligés. Des initiatives telles que celle mise en avant par la *Save Darfur Coalition* dans le cadre de la guerre civile au Soudan, ou par l'organisation *Invisible Children* avec la campagne *Kony2012* qui dénonçait les atrocités commises par l'Armée de résistance du Seigneur, sont des exemples connus. Comme les observateurs de ces campagnes l'ont souligné, des solutions d'ampleur internationale comme celles-ci peuvent avoir des répercussions négatives sur la capacité de trouver des solutions au plan local (Schomerus, 2012).

Le cas du Myanmar est un exemple de la façon dont les organisations de la société civile peuvent se mobiliser de manière proactive pour créer des coalitions locales capables de répondre aux menaces qui se font jour. Harry Myo Lin, militant du Panzagar, explique : « À l'heure actuelle, nous voulons nous concentrer sur une collaboration avec les écoles monastiques, mais aussi avec les religieux musulmans. C'est un processus qui prendra beaucoup de temps, mais nous ne voulons pas nous contenter de nous focaliser sur le discours de haine dans les médias »<sup>98</sup>.

Comme les militants interrogés dans cette étude l'ont reconnu, l'équilibre entre les solutions locales, la sensibilisation du public au niveau international, la production de résultats intéressants au plan local, et la nécessité d'éviter de perturber un processus de transition délicat est difficile à maintenir. Néanmoins, les efforts qu'ils déploient montrent que la mobilisation contre la haine en ligne peut permettre d'utiliser Internet comme arme face aux conflits trouvant des répercussions sur le Web.

## Faire pression sur les entreprises privées

Diverses organisations qui ont lutté contre le discours de haine sous d'autres formes, ou qui ont défendu les droits de groupes spécifiques par le passé, se sont vues jouer un rôle de plus en plus important sur Internet. Ce phénomène est particulièrement frappant dans les pays développés, où la pénétration d'Internet est élevée, et où les entreprises privées sont les intermédiaires clés. Le présent chapitre s'intéresse aux campagnes et initiatives menées aux États-Unis, en Australie et au Royaume-Uni, où les questions de haine en ligne sont apparues en lien avec la religion, la race et le genre. Il met en évidence la signification plus large de ces initiatives par rapport aux cas de discours de haine perçus comme tels.

## Différentes approches stratégiques

Les organisations telles que l'*Anti-Defamation League* (ADL) et le *Women, Action and the Media* (WAM!) (États-Unis) ; le *Online Hate Prevention Institute* (Australie) ; le *Sentinel Group for Genocide Prevention* (Canada) ; et le *TellMAMA - Measuring Anti-Muslim Attacks* (Royaume-Uni) se sont de plus en plus investis dans la lutte contre le discours de haine en ligne, en ciblant les intermédiaires Internet, en demandant à ces derniers d'assumer une plus grande responsabilité dans la modération des contenus, et en essayant de sensibiliser les usagers.

Christopher Wolf, l'un des membres du conseil de l'ADL, a déclaré : « Il semble que les personnes exposées au discours de haine en ligne aujourd'hui ne font que hausser les épaules et passer outre, parce qu'ils s'y sont habitués et sont comme anesthésiés. Nous ne sommes pas assez éduqués à l'utilisation d'Internet ; les gens se contentent de prendre les appareils et de les utiliser sans être clairement guidés, et je crois que nous pouvons mieux faire que cela en termes d'éducation »<sup>99</sup>. La nécessité d'améliorer l'éducation et l'accès à l'information pour lutter contre le discours de haine en ligne a aussi été soulevée par le directeur des opérations du *Sentinel Project for Genocide Prevention*, Drew Boyd, qui a conclu que « L'ignorance est le fil rouge du discours de haine, tant en ligne qu'en dehors. Le manque d'information, qu'il soit intentionnel ou non, conduit les gens à adopter une vision du monde étroite qui s'accommode de la haine de l'autre. Il est nécessaire de faire un effort collaboratif pour diffuser des informations factuelles et d'autres visions du monde auprès des réseaux fermés où circulent de fausses informations ou des informations inexacts »<sup>100</sup>.

Outre la sensibilisation des utilisateurs, ces groupes œuvrent à faire pression sur les intermédiaires Internet pour qu'ils prennent des mesures plus strictes contre le discours de haine en ligne. On peut découper leur action selon deux stratégies souvent complémentaires.

Dans certains cas, les organisations s'emploient principalement à faire pression sur les entreprises privées en prenant des situations spécifiques et en lançant des négociations. Dans ce cadre, les organisations en question peuvent choisir de faire valoir leur point de vue par l'intermédiaire de campagnes en ligne, de plaintes et de blocus organisés, de lettres ouvertes, de pétitions en ligne et d'appels vigoureux à la mobilisation de leurs sympathisants, aussi bien en ligne qu'en dehors. Cependant, ce sont surtout les organisations qui se font les chefs de file d'une cause spécifique. Cela a été le cas, par exemple, concernant la demande qui a été faite à Google de modifier ses critères de recherche de façon que *Jew Watch* - site Web antisémite faisant l'apologie du négationnisme, n'apparaisse pas parmi les premiers résultats lorsque l'on recherche le mot « juif » (Foxman & Wolf 2013) -, ou dans le cadre de la campagne menée par WAM!, décrite ci-dessous.

Un second type d'initiatives mis en avant parmi certaines de ces organisations s'emploie à recueillir les plaintes d'utilisateurs concernant certains types de contenus. Cette activité semble particulièrement intéressante si on la met en relation avec la procédure adoptée par

les intermédiaires Internet pour résoudre les cas de discours de haine. Si des entreprises telles que Google, par exemple, ont commencé à publier des rapports de transparence répertoriant les demandes faites par les gouvernements concernant la publication ou la suppression de données, informations et contenus, elles ne publient pas de telles informations pour les demandes faites par des utilisateurs à titre individuel. Lorsque des particuliers signalent un contenu comme étant inapproprié, ils peuvent être tenus informés de l'état d'avancement du traitement de leur plainte (sur Facebook, par exemple, un tableau de bord est prévu à cet effet), mais cette procédure reste largement cachée des autres utilisateurs et organisations, ce qui limite la possibilité de mieux comprendre ce que les personnes jugent comme étant des propos blessants, irrespectueux, insultants ou haineux.

Diverses initiatives font appel à l'externalisation ouverte (*crowdsourcing*) pour les demandes de sanctions face à certains types de messages, par exemple *HateBase*, qui émane du *Sentinel Group for Genocide Prevention and Mobocracy* ; la *Islamophobic incidents reporting platform*, plateforme de *TellMAMA* ; et *Fight Against Hate*, de l'*Online Hate Prevention Institute*. Ces initiatives vont plus loin dans la lutte contre le discours de haine en ligne et servent d'outils innovants pour ce qui est de garder trace des propos diffamatoires sur les réseaux sociaux et de la façon dont ils sont répliqués par les différentes entreprises privées.

*HateBase* s'emploie à établir une cartographie du discours de haine déployé dans les messages accessibles publiquement sur les plateformes de réseaux sociaux afin d'obtenir une carte géographique des contenus haineux diffusés en ligne, ce qui permet à la fois d'avoir une vision globale et un aperçu plus précis des termes spécifiques qui sont utilisés et des tendances et cibles les plus courantes de ces propos diffamatoires. Cette base de données comporte en outre une fonction complémentaire permettant de signaler individuellement des cas de discours de haine en ligne, qui est utilisée pour améliorer la précision et l'ampleur de l'analyse, les utilisateurs vérifiant les exemples signalés et confirmant leur caractère haineux au sein d'une communauté donnée<sup>101</sup>. De même, *Fight Against Hate* permet de signaler le discours de haine en ligne proféré sur différents réseaux sociaux sur une plateforme unique, ce qui aide aussi les utilisateurs à garder trace du nombre de personnes ayant signalé des contenus haineux, du lieu où ces contenus ont été publiés, du temps qu'il a fallu aux entreprises privées pour réagir, et de la modération effective (ou non) de ces contenus<sup>102</sup>. Enfin, *TellMAMA*, organisation basée au Royaume-Uni, offre une fonction similaire permettant de signaler des cas observés sur différents sites sur une seule plateforme, bien qu'elle se concentre exclusivement sur les contenus anti-musulmans. Elle facilite aussi la constitution d'un dossier sur les incidents à caractère racial ou religieux pour analyse ultérieure. Les cas signalés sur la plateforme sont traités par l'organisation, qui contacte ensuite les victimes et les aide à suivre la procédure appropriée pour porter certains incidents à la connaissance des autorités compétentes chargées de faire appliquer la loi<sup>103</sup>. Les informations enregistrées sont aussi utilisées pour repérer les tendances du discours de haine, tenu en ligne et en dehors, à l'égard des musulmans au Royaume-Uni<sup>104</sup>.

Concernant l'importance qu'il y a à générer des données empiriques, le PDG de l'*Online Hate Prevention Institute*, Andre Oboler, a indiqué que de telles plateformes offraient la possibilité de faire des demandes qui soient visibles des autres utilisateurs enregistrés, ce qui permet de garder trace de la date où un contenu a été signalé pour la première fois, du nombre de personnes qui l'ont signalé et du temps qu'il a fallu en moyenne pour qu'il soit retiré.

Par ces moyens, ainsi que d'autres, ces organisations peuvent s'inscrire au sein d'ensembles plus vastes d'acteurs participant à un débat sur la nécessité de trouver un équilibre entre liberté d'expression et respect de la dignité humaine. L'exemple ci-dessous l'illustre bien ; une page Facebook exprimant de la haine à l'égard des aborigènes d'Australie a fini par être retirée par Facebook même si elle n'enfreignait pas ses conditions d'utilisation, car elle a été jugée insultante par tout un éventail d'acteurs, notamment la société civile et des groupes de pression, des chargés de réglementation et des particuliers.

### *Faire campagne sur le long terme*

Le cas ci-après illustre comment une vaste controverse au niveau des simples citoyens, au sujet de propos diffamatoires publiés en ligne, peut arriver aux oreilles des organisations et autorités gouvernementales compétentes et les conduire à participer à leur tour activement au débat en ligne et à faire pression sur les entreprises privées pour résoudre la question. En 2012, une page Facebook ridiculisant les aborigènes d'Australie, intitulée « Aboriginal Memes » a causé un tollé local en ligne qui s'est manifesté par un flot organisé pour signaler le contenu abusif, une vaste couverture médiatique, une campagne sociale et une pétition en ligne assortie d'une lettre ouverte exigeant que Facebook supprime le contenu visé. Le terme de meme (même Internet) fait référence dans ce cas à une forme d'expression visuelle faisant passer des messages courts au moyen d'une série d'images, assorties d'inscriptions intégrées dans le corps de ces images.

Le vaste soutien manifesté sur Internet en faveur de la lutte contre la page Facebook « Aboriginal Meme » a pu être observé à travers différentes plateformes de réseaux sociaux et d'information, suscitant aussi un intérêt accru parmi les chaînes d'information étrangères. En réponse à l'agitation des médias, Facebook a publié une déclaration officielle sur les médias australiens : « Nous sommes conscients des inquiétudes soulevées auprès du public par les pages controversées de mêmes Internet qui ont été créées par des Australiens sur Facebook. Nous pensons que le partage d'informations, et l'ouverture qui en résulte, invite à la conversation, au débat, et à une meilleure compréhension. D'un autre côté, nous sommes conscients que certains contenus partagés peuvent être controversés, blessants ou même illégaux dans certains pays »<sup>105</sup>. En réponse à la déclaration officielle de Facebook, le Commissaire australien aux droits de l'homme interrogé à la télévision publique a affirmé qu'il désapprouvait la page controversée ainsi que le fait que Facebook applique le premier

amendement de la constitution américaine sur une question qui concernait un contrevenant d'origine australienne et des victimes d'origine australienne<sup>106</sup>.

La pétition en ligne a été élaborée pour réagir de plus belle face au refus de Facebook de supprimer le contenu, refus qui était manifesté par l'envoi automatique du message suivant à plusieurs personnes ayant signalé le contenu abusif : « Après examen de votre demande, nous n'avons pu confirmer que la page que vous aviez signalée contrevenait à la Déclaration des droits et responsabilités de Facebook »<sup>107</sup>. La lettre ouverte accompagnant la pétition expliquait que le contenu était considéré comme insultant en raison d'attaques répétées à l'encontre d'un groupe spécifique, pour motifs racistes, et exigeait que Facebook prenne des mesures en retirant les pages concernées et pages similaires visant les Australiens autochtones. Le 11 août 2012, la créatrice de la pétition, Jacinta O'Keefe, a publié une note dans laquelle elle a remercié les participants et annoncé la victoire de la pétition lorsqu'elle s'est aperçu que Facebook avait retiré le contenu en question ainsi qu'une autre page visant les Australiens autochtones<sup>108</sup>. Néanmoins, les pages n'avaient été supprimées que temporairement pour examen de leur contenu, et après des discussions avec le Commissaire à la lutte contre la discrimination raciale et l'*Online Hate Prevention Institute*, Facebook est parvenu à la conclusion que le contenu ne contrevenait pas à ses conditions d'utilisation et a autorisé les pages à continuer d'exister, sous réserve que le mot « controversé » soit ajouté à leur titre, de façon à indiquer clairement qu'elles comportaient des contenus controversés<sup>109</sup>.

Une deuxième phase a été déclenchée dans la régulation des contenus après qu'un utilisateur de Facebook a commencé à cibler des militants anti-haine en les attaquant personnellement avec des propos diffamatoires dans le cadre de l'affaire de la page *Aboriginal Memes*. Facebook a réagi en retrouvant et en bannissant les nombreux comptes factices qu'il avait créés, tout en l'autorisant à garder un compte en fonctionnement. Enfin, au cours d'un troisième temps, Facebook a limité l'accès à la page controversée en Australie après que le Commissaire à la lutte contre la discrimination raciale et l'Autorité australienne des communications et des médias ont publiquement exprimé leurs inquiétudes. La page Facebook bannie reste cependant fonctionnelle et accessible hors d'Australie, et continue de propager les contenus haineux publiés sur d'autres pages accessibles en Australie. Pour essayer d'empêcher certains utilisateurs précis de diffuser davantage les images controversées, ces derniers se sont vu interdire l'utilisation de Facebook pendant 24h – mesure qui a été jugée inefficace pour les dissuader d'arrêter leur activité (Oboler, 2012).

### *Faire campagne contre les publicitaires*

Dans ce cas, les organisations impliquées se sont saisies d'une controverse en ligne de longue date et ont agi au-delà de leur rôle d'intermédiaires chargés de relayer les plaintes, en faisant elles-mêmes activement et fortement pression sur les entreprises, exigeant une modération

plus rigoureuse des contenus, ainsi que des mesures d'autoréglementation supplémentaires et pérennes. En 2013, le groupe *Women, Action and the Media (WAM!)*<sup>110</sup> et le *Everyday Sexism Project* (projet sur le sexisme dans la vie quotidienne), au Royaume-Uni<sup>111</sup>, a adopté une stratégie de relations publiques agressive en lançant une campagne commune montrant les pages de publicités d'entreprises de renom qui diffusaient des infographies sexistes. En réponse à cette campagne, Nissan et la compagnie d'assurance Nationwide ont retiré leurs publicités de Facebook. Voyant leur succès, les organisateurs, soutenus par des sympathisants et militants en ligne, ont commencé à adresser des plaintes écrites et des photos de diverses publicités sur des pages haineuses à d'autres grandes entreprises comme Dove et American Express sur leurs plateformes des médias sociaux, les exhortant à suivre l'exemple. À l'issue de cette campagne, 15 grandes entreprises ont décidé de retirer leurs publicités de Facebook<sup>112</sup>.

La campagne a aussi comporté une lettre ouverte officielle, rédigée par les deux groupes susmentionnés, faisant la liste des pages encourageant le viol et la violence à l'égard des femmes, exigeant que celles-ci soient supprimées et que Facebook révise sa politique de réglementation des contenus. Outre la lettre ouverte, une pétition en ligne sur *change.org* a recueilli plus de 225 000 signatures et contribué à sensibiliser les utilisateurs sur le Web<sup>113</sup>. Les militants soutenant la campagne ont décidé d'aller plus loin et ont organisé une manifestation de grande ampleur devant le lieu de réunion des actionnaires de Facebook, publiant le nom de toutes les entreprises de renom qui utilisaient la plateforme pour y faire de la publicité en ligne, appelant les personnes à leur envoyer des lettres de plainte, et exhortant les publicitaires à retirer leurs publicités de Facebook. En outre, des militants ont aussi engagé des journalistes économiques sur leurs pages sur les médias sociaux, leur demandant d'analyser les pertes budgétaires que Facebook serait susceptible de subir compte tenu du nombre croissant d'entreprises qui se retireraient. La campagne en ligne qui utilisait le hashtag *#FBrape* contre Facebook et les entreprises faisant de la publicité sur la plateforme a eu pour effet que Facebook a contacté les organisations concernées pour leur faire une demande de coopération. La campagne *#FBrape* n'a reçu une certaine attention des médias qu'une fois qu'elle a réussi à faire pression sur l'entreprise pour qu'elle se lance dans une lutte active contre les contenus haineux ciblant les femmes, surtout parce qu'il s'agissait d'un coup d'éclat qui visait des entreprises précises et leurs campagnes de publicité et non pas simplement Facebook directement<sup>114</sup>.

La réaction de Facebook, cependant, n'a dans un premier temps pas été de coopérer, l'entreprise maintenant que les pages recensées dans la lettre ouverte ne violaient pas ses conditions d'utilisation. Néanmoins, peu après le début de la campagne, quand les entreprises ont commencé à se retirer, le contenu diffamatoire a rapidement été retiré. Facebook a alors publié une déclaration officielle, indiquant son intention d'éclaircir certains points de ses conditions d'utilisation et politiques de régulation des contenus, et de promouvoir la coopération avec les organisations œuvrant à la défense de la liberté d'expression tout en s'efforçant d'éviter que le discours de haine en ligne vise des groupes et individus spécifiques. À la fin de la lettre, Facebook a ajouté : « ces derniers jours, il est devenu évident que nos systèmes de repérage et de suppression du discours de haine n'ont pas été aussi efficaces

que nous l'aurions souhaité, en particulier autour des questions de haine sexiste. Dans certains cas, les contenus ne sont pas retirés aussi vite que nous l'aimerions. Dans d'autres cas, des contenus qui auraient dû être supprimés ne l'ont pas été ou ont été évalués sur la base de critères obsolètes »<sup>115</sup>. L'entreprise a ensuite déclaré qu'elle avait l'intention de réviser et d'actualiser ses lignes directrices en matière de modération du discours de haine, d'offrir à ses modérateurs de contenus une formation de meilleure qualité, et de renforcer sa collaboration avec les organisations concernées afin de faciliter une action commune et rapide en vue de mieux faire barrage aux contenus haineux et aussi d'agir pour que les diffuseurs de ces contenus répondent de leurs actes<sup>116</sup>.

Dans un autre cas en lien avec celui-ci, Twitter a aussi pris position contre le harcèlement des femmes en collaboration avec le *WAM!*, en lançant un projet pilote conjoint sous la forme d'une plateforme qui permettait de faire modérer le contenu signalé dans les 24h. Le fait que les victimes signalent les cas de propos abusifs à l'égard des femmes sert un double objectif : permettre au *WAM!* de récolter des données sur les contenus insultants axés sur le harcèlement sexiste en ligne, afin d'étudier le phénomène en profondeur ; et aider Twitter à améliorer ses mécanismes de réglementation vis-à-vis de la discrimination et des abus sexistes<sup>117</sup>. Cet outil demande aux femmes de nommer les utilisateurs individuels qui les harcèlent ou les tweets précis qu'elles trouvent insultants, de définir le type de harcèlement dont il s'agit, et de répondre à des questions d'ordre général sur le nombre de fois où elles ont été harcelées, sur quelles plateformes et si le harcèlement en question émanait d'un ou plusieurs utilisateurs. Une fois la plainte déposée, les demandes sont étudiées par le *WAM!* puis transmises à Twitter pour une enquête plus approfondie et pour modération<sup>118</sup>.

Twitter a publié une déclaration sur les efforts qu'il menait pour lutter contre le discours de haine en ligne, indiquant : « Nous essayons toujours d'améliorer la façon dont nous traitons les problèmes de violence, et le *WAM!* est l'une des nombreuses organisations avec lesquelles nous travaillons à travers le monde sur les meilleures pratiques pour la sécurité des usagers »<sup>119</sup>. La Directrice exécutive du *WAM!*, Jaclyn Friedman, a souligné l'importance de l'initiative : « Nous sommes ravis de travailler avec Twitter pour rendre leur plateforme sûre pour les femmes. Le fait que les femmes soient très majoritairement ciblées en ligne fait qu'elle ne s'expriment plus dans les débats publics. Nous sommes vraiment heureux que Twitter reconnaisse que la meilleure façon d'assurer la liberté d'expression pour tous les utilisateurs sur sa plateforme est de veiller à ce que tous les utilisateurs soient libres de parler sans faire l'objet de harcèlement, d'abus ni de menaces »<sup>120</sup>. Le programme pilote a fonctionné pendant trois semaines, au cours desquelles 700 plaintes auraient été déposées et plus de 100 personnes auraient trouvé de l'aide pour obtenir une réaction plus rapide de la part de Twitter. L'organisation prévoit de produire un rapport sur les données recueillies pendant l'expérience, en vue de parvenir à une meilleure compréhension du discours de haine à l'égard des femmes en ligne<sup>121</sup>.

## *Limites à la lutte contre le discours de haine en ligne et réactions des intermédiaires Internet*

Il semblerait que la lutte contre le discours perçu comme discours de haine en ligne commence à préoccuper un certain nombre d'acteurs, des gouvernements aux entreprises privées en passant par les fournisseurs de services Internet, ainsi qu'un nombre croissant d'organisations actives et de victimes. De nombreux particuliers et communautés en ligne luttent également quotidiennement aux côtés des organisations plus formelles contre les contenus diffamatoires sur Internet. Toutefois, cette lutte nécessite une action à grande échelle afin de faire en sorte que les discours de haine en ligne puisse être repérés et évincés de manière efficace, dans leur contexte et à long terme, ce qui exige de donner aux utilisateurs les outils pour qu'ils puissent repérer et combattre eux-mêmes le discours de haine de manière autonome sans faire obstacle à l'exercice légitime de la liberté d'expression, créant ainsi des espaces d'expression plus ouverts.

Les intermédiaires Internet et les plateformes de réseaux sociaux en particulier ont fait des progrès dans leurs interventions en cas de discours de haine présumé ; ils interagissent scrupuleusement avec les utilisateurs qui signalent des contenus et rendent leurs procédures de réglementation plus transparentes. Dans un entretien, Monika Bickert et Ciara Lyden<sup>122</sup>, de Facebook, ont indiqué qu'elles s'appuyaient sur le travail de plusieurs équipes pour s'occuper des différents types de contenus dans les différentes langues, afin de traiter les cas signalés le plus rapidement et le plus efficacement possible. En outre, Facebook a lancé un tableau de bord permettant de signaler des contenus, grâce auquel les utilisateurs peuvent suivre la procédure de traitement de leurs demandes, la communication individuelle avec chaque utilisateur s'en trouvant améliorée. Adoptant des mécanismes similaires pour traiter du discours de haine, Twitter a lancé un bouton permettant de signaler des contenus en 2013<sup>123</sup>, en réponse à une pétition en ligne lancée par un particulier<sup>124</sup>.

En résumé, de plus en plus, les intermédiaires Internet travaillent en étroite coopération avec différentes organisations afin d'apporter des solutions rapides et efficaces au discours de haine sur leurs plateformes, mais ils soulignent aussi qu'ils accordent autant d'attention aux plaintes déposées par des particuliers et qu'ils les traitent aussi rigoureusement qu'ils le font avec les pétitions et autres formes d'actions collectives. Enfin, ces entreprises privées commencent aussi à publier des rapports de transparence afin d'informer les usagers de tout changement dans leurs politiques et paramètres de confidentialité, bien que peu d'entre elles fournissent des informations sur le rapport quantitatif entre les contenus signalés par des particuliers et les demandes officielles faites par des autorités gouvernementales<sup>125</sup>.

## Lutter contre le discours de haine en ligne par l'éducation de base aux médias et à l'information

Alors que les chapitres précédents ont abordé la plupart des solutions apportées en réaction à la prolifération des discours de haine sur Internet, le présent chapitre offre un aperçu des tentatives de solutions plus structurelles, à travers l'éducation. Il livre une analyse d'une série d'initiatives ciblant les jeunes ou menées en partenariat avec des établissements d'enseignement pour faire prendre conscience aux jeunes des enjeux et des réponses possibles face au discours perçu comme discours de haine sur Internet.

### *Éducation à la citoyenneté et citoyenneté numérique*

L'éducation à la citoyenneté vise à préparer les individus à être des citoyens informés et responsables, à travers l'étude des droits, libertés et responsabilités qui sont les leurs ; elle a été employée de diverses manières dans les sociétés sortant d'un conflit violent (Osler et Starksey, 2005). Un de ses principaux objectifs est la sensibilisation aux droits politiques, sociaux et culturels des individus et des groupes, y compris à la liberté d'expression et aux responsabilités et conséquences sociales qui en découlent. Dans certains cas, l'acquisition de capacités d'argumentation efficace et d'expression respectueuse des croyances et opinions personnelles a été incluse parmi les résultats d'apprentissage visés dans les programmes d'éducation à la citoyenneté.

Le lien entre l'éducation à la citoyenneté et le discours de haine est double : il s'agit d'une part d'acquérir les connaissances et compétences nécessaires pour repérer les discours de haine, et d'autre part de donner aux individus les moyens de contrecarrer les messages haineux. L'un des défis actuels de l'éducation à la citoyenneté est d'adapter ses objectifs et stratégies au monde numérique, en proposant les connaissances et compétences, non seulement en matière d'argumentation, mais aussi de technologie, dont le citoyen peut avoir besoin pour lutter contre le discours de haine sur Internet. Un nouveau concept de citoyenneté numérique est proposé par certains des organismes figurant dans cette étude ; il intègre les objectifs fondamentaux de l'éducation de base au bon usage des médias et de l'information en vue de développer les compétences techniques et l'esprit critique des usagers et des producteurs des médias, en lien avec les questions éthiques et civiques d'ordre plus général.

On citera à cet égard l'éducation à la citoyenneté mondiale (ECM), qui est l'un des axes stratégique de travail du programme Éducation (2014-2017) de l'UNESCO<sup>126</sup> et l'une des trois priorités de l'Initiative mondiale pour l'éducation avant tout (GEFI) du Secrétaire général de l'ONU, lancée en septembre 2012. L'éducation à la citoyenneté mondiale vise à doter les apprenants de tous âges de valeurs, connaissances et compétences fondées sur les droits de

l'homme, la justice sociale, la diversité, l'égalité des genres et la durabilité de l'environnement, et favorisant le respect de ces éléments. L'ECM offre des compétences aux apprenants ainsi que des occasions d'exercer leurs droits et de s'acquitter de leur devoir de promouvoir un monde et un avenir meilleurs pour tous.

Dans cette perspective plus large, l'UNESCO encourage aussi l'éducation de base aux médias et à l'information. Il s'agit d'un concept global qui regroupe diverses compétences de base (à la fois en ligne et hors ligne). Il comporte le développement des compétences et aptitudes techniques nécessaires à l'utilisation des technologies numériques, ainsi que les connaissances et capacités nécessaires à la recherche, l'analyse, l'évaluation et l'interprétation de textes spécifiques issus des médias, la création de messages médiatiques, et la prise de conscience de l'influence des médias aux plans social et politique (Hoechsmann et Poyntz, 2012). Ces dernières années, ceux qui mettent l'accent sur l'éducation de base aux médias ont commencé à insister sur les répercussions sociales de l'utilisation des technologies, les implications éthiques de celle-ci et les droits et devoirs civiques qui en découlent. Aujourd'hui, l'éducation de base à l'information ne peut contourner les questions de droit à la libre expression et à la vie privée, de citoyenneté critique ou d'autonomisation des citoyens en vue de leur participation à la vie politique (Mossberger *et al.*, 2008). Il devient essentiel d'acquérir des compétences multiples et complémentaires.

L'apparition des nouvelles technologies et des médias sociaux a joué un rôle important dans ce changement de paradigme. Les particuliers ne sont plus de simples consommateurs de messages médiatiques mais sont aujourd'hui des producteurs, des créateurs et des détenteurs d'informations, ce qui se traduit par de nouveaux modèles de participation qui interagissent avec les modèles traditionnels, comme le vote ou l'adhésion à un parti politique. Les stratégies d'enseignement évoluent en conséquence ; il ne s'agit plus simplement de favoriser l'esprit critique dans la lecture des messages médiatiques mais aussi d'apprendre à créer de manière autonome du contenu médiatique (Hoechsmann et Poyntz, 2012). La notion d'éducation de base aux médias et à l'information continue d'évoluer, prenant désormais en compte la dynamique de l'Internet. Elle commence à intégrer les questions d'identité, d'éthique et de droits dans le cyberspace (voir la Déclaration de Paris)<sup>127</sup>.

Certaines de ces compétences peuvent être particulièrement importantes lorsqu'il s'agit de repérer le discours de haine sur Internet et d'y répondre, et le présent chapitre analyse une série d'initiatives visant à fournir des informations ainsi que des outils pratiques aux utilisateurs d'Internet pour qu'ils deviennent des citoyens numériques actifs. Les projets et organisations impliqués dans ce type de solutions sont :

- « No place for hate », projet de l'Anti-Defamation League (ADL), États-Unis ;

---

- « In other words », projet de la Province de Mantoue et de la Commission européenne ;

---

- « Faire face à la haine sur Internet », projet d'HabiloMédias, Canada ;

---

- « Mouvement contre le discours de haine », projet du Département de la jeunesse du Conseil de l'Europe ;<sup>128</sup>
- « Online hate », projet de l'Online Hate Prevention Institute, Australie.

Une mise en regard comparative de tous ces projets et de leurs supports sur les questions de discours de haine en ligne a été réalisée. En outre, des entretiens avec des représentants des organisations ou personnes chargées des programmes éducatifs ont été menés.

### *L'éducation comme outil de lutte contre le discours de haine*

Bien que les initiatives et organisations présentées possèdent des caractéristiques spécifiques et poursuivent des objectifs qui leur sont propres, toutes soulignent l'importance de l'éducation de base aux médias et à l'information et des stratégies éducatives comme moyens de contrecarrer le discours de haine. Elles soulignent qu'une approche axée sur l'éducation représente une solution plus structurelle et plus viable face au discours de haine, si l'on considère, par comparaison, combien il est complexe de décider de supprimer ou de censurer un contenu en ligne, ou combien il peut être long et coûteux d'intenter des actions en justice si l'on veut obtenir des résultats tangibles. Beaucoup sont d'avis que l'ensemble de compétences relevant de l'éducation de base aux médias et à l'information peut autonomiser les individus et les doter des connaissances et aptitudes dont ils ont besoin pour répondre au discours perçu comme discours de haine de manière plus immédiate. Ces aptitudes peuvent aussi être particulièrement importantes, compte tenu de l'accent que les plateformes de réseaux sociaux mettent sur le fait de signaler à titre individuel les cas de propos insultants, d'incitations à la haine ou de harcèlement.

Les personnes impliquées dans ces initiatives, cependant, reconnaissent l'importance de pouvoir se référer à un cadre juridique dans leur action. Comme l'a affirmé Laura Geraghty, du Mouvement contre le discours de haine : « L'éducation est essentielle pour prévenir le discours de haine en ligne. Il faut sensibiliser et autonomiser les personnes pour qu'elles agissent sur Internet de manière responsable. Néanmoins, on aura toujours besoin d'un fondement et d'instruments juridiques pour juger les crimes de haine, y compris le discours de haine en ligne, sans quoi les actions préventives ne seront d'aucune utilité<sup>129</sup> ».

La plupart des initiatives prévoient une éducation aux instruments et procédures juridiques à saisir pour poursuivre les auteurs de discours de haine sur Internet, et beaucoup encouragent à voir les aspects juridiques et éducatifs comme complémentaires. L'éducation aux médias peut représenter une stratégie de plus en plus importante pour lutter contre le discours de haine en ligne, mais étant donné l'ampleur du problème, elle doit être envisagée en complément d'autres types de solutions.

## *Développement de l'esprit critique pour contrer le discours de haine en ligne*

L'un des dénominateurs communs des initiatives analysées est l'accent mis sur le développement des facultés de pensée critique et sur un usage des médias sociaux intégrant une réflexion éthique, comme acquis de base de l'éducation aux médias et à l'information pour la lutte contre le discours de haine sur Internet. On part du principe que ces acquis dans le domaine des médias et de l'information peuvent : renforcer les capacités individuelles de repérer et de contester le discours de haine en ligne, et de comprendre certains de ses présupposés, partis pris et préjugés ; et encourager l'élaboration d'arguments contre ces derniers éléments.

Comme on peut s'y attendre, le repérage du discours de haine sur Internet n'est pas nécessairement aussi simple qu'il y paraît. Comme l'a expliqué un enseignant participant à un groupe de travail coordonné par HabiloMédias : « Il m'a même fallu quelques minutes avant que je comprenne que j'étais sur un site de sympathisants Nazis. C'était superbement écrit, de la façon la plus sournoise qui soit. Cela enrobait les vrais messages racistes et haineux dans de la prose. De la rhétorique, quoi. Du coup, j'ai montré ça aux enfants, et quand j'ai éteint ma lampe, la leur était toujours allumée. Ils ne savaient pas ce qu'il regardaient. Je leur ai demandé de regarder d'un peu plus près, quelques uns ont commencé à voir les choses ; d'autres, toujours pas. Et cela les a intéressés, car je voyais quelque chose qui leur échappait. C'était une façon pour eux de réaliser, d'être intéressés par l'idée que quelqu'un prônait la haine sans même qu'on s'en rende compte »<sup>130</sup>.

Les initiatives analysées sont destinées à des publics divers concernés et touchés par le discours de haine en ligne. Pour les organisations participantes étudiées ici, il est particulièrement important de centrer les efforts sur les groupes vulnérables et sur ceux susceptibles d'être visés par la haine ou d'être recrutés par des groupes prônant la haine. Les enfants et les jeunes sont l'un des principaux publics visés par ces initiatives. Comme l'a indiqué Matthew Johnson, Directeur de l'éducation chez HabiloMédias : « Expliquer aux jeunes que les médias sont des constructions qui représentent la réalité contribue à leur faire prendre conscience de la nécessité de comprendre qui se cache derrière ce qu'ils voient ; leur expliquer que ces médias diffusent des idéologies – sur les valeurs, le pouvoir, et l'autorité – qui ont des implications au plan social et politique les aide à comprendre pourquoi il faut tout de même réagir face aux contenus haineux que l'on trouve dans la musique, les jeux, ou d'autres médias, aussi « triviaux » soient-ils »<sup>131</sup>.

Les parents, les enseignants et la communauté scolaire sont aussi généralement considérés comme un public important de par leur rôle vis-à-vis de l'exposition des enfants aux contenus haineux et de la protection des enfants face à ces contenus. Les autres groupes visés sont les groupes susceptibles de façonner le paysage juridique et politique de la lutte contre le discours de haine sur Internet, notamment les responsables de l'élaboration des politiques et

les organisations non gouvernementales, ainsi que ceux qui peuvent avoir une grande influence sur la communauté en ligne en l'exposant à des propos haineux, en particulier les journalistes, blogueurs et militants. On trouvera au tableau 1 un résumé des différents publics visés par les initiatives analysées.

|                                       | Enfants | Jeunes | Enseignants | Parents | Responsables de l'élaboration des politiques | Bloggers | ONG | Grand public |
|---------------------------------------|---------|--------|-------------|---------|--|----------|-----|--------------|
| Anti-défamation league                | X       | X      | X           | X       | X  |          |     |              |
| In Other Words                        |         |        |             |         | X  | X        | X   | X            |
| Mouvement contre le discours de haine |         | X      |             |         |  | X        |     | X            |
| HabiloMédias                          | X       | X      | X           | X       |  |          |     |              |
| Online Hate Prevention                |         | X      |             |         |  | X        | X   | X            |

Tableau 1 : Public visé par chaque initiative éducative

Comme il fallait s'y attendre, c'est le contenu mais aussi les objectifs et buts de chaque projet qui sont étroitement liés aux intérêts et besoins du public visé par chaque initiative. Par exemple, HabiloMédias a développé un jeu vidéo en ligne destiné aux enfants de 12 à 14 ans, qui a été conçu pour accroître la capacité des élèves à reconnaître les partis pris, les préjugés et la propagande haineuse. Dans ce jeu vidéo, lorsqu'ils repèrent différents degrés de préjugés et de discrimination, sous forme de blagues, nouvelles ou vidéos, les enfants doivent déterminer en quoi ces messages sont susceptibles d'inciter à la haine, puis développer des stratégies pour faire face au problème, en choisissant soit d'ignorer ces messages, soit d'agir de front.

L'ADL a focalisé une grande partie de ses efforts de sensibilisation et d'éducation sur les enseignants et les parents, leur fournissant des informations essentielles sur la façon d'aborder les questions de haine et de violence avec les enfants, et d'encourager les jeunes à réagir de la manière la plus appropriée. La diffusion de fascicules, l'organisation de sessions de formation dans les écoles et d'ateliers au sein des communautés sont autant de stratégies mises en œuvre par ADL pour développer une culture ouverte et un climat scolaire respectueux.

Le Mouvement contre le discours de haine a organisé des sessions de formation à l'intention des blogueurs et des jeunes militants afin qu'ils puissent discuter dans un environnement accueillant de certaines de leurs expériences avec le discours de haine sur Internet, et partager les meilleures pratiques pour lutter contre ce discours. Par exemple, comment mettre en œuvre des campagnes d'affirmation qui dépeignent les groupes minoritaires de manière positive, comme mesure préventive contre la discrimination ; ou comment mener des campagnes obstructives pour signaler et restreindre les contenus discriminatoires et l'activité en ligne<sup>132</sup>. Ces sessions de formation visent à promouvoir une compréhension du discours de haine au

niveau des citoyens de base et à sensibiliser le public à l'influence que les blogueurs et les militants peuvent avoir pour venir à bout des contenus haineux.

En revanche, le projet « In Other Words » a cherché à influencer les responsables de l'élaboration des politiques et la société civile pour qu'ils surveillent différents types de médias. Il préconise de donner une représentation fidèle des minorités et groupes vulnérables dans les médias, encourageant la société à exercer une surveillance pour éviter la diffusion de stéréotypes, préjugés et autres types de discours discriminatoires.

### *Objectifs éducatifs de l'éducation de base aux médias et à l'information pour répondre du discours de haine*

Malgré les particularités du contenu de chacune des initiatives et des publics qu'elles visent, elles ont toutes trois objectifs éducatifs en commun : s'informer sur le discours de haine, l'analyser, et y répondre. Ces trois objectifs peuvent être considérés dans un ensemble qui comporte des objectifs progressifs associés à des objectifs spécifiques, chacun mettant l'accent sur différents aspects du problème et proposant des solutions spécifiques pour répondre au discours en ligne. On en trouvera un résumé au tableau 2.

| Information   | Analyse  | Action   |
|---|--|--|
| <ul style="list-style-type: none"> <li>- Sensibiliser au discours de la haine et à ses conséquences</li> <li>- Faire passer et diffuser les informations</li> <li>- Communiquer sur le cadre juridique pertinent</li> </ul> | <ul style="list-style-type: none"> <li>- Repérer et évaluer le discours de haine</li> <li>- Analyser les causes courantes et les présupposés et préjugés sous-jacents</li> <li>- Reconnaître les partis-pris</li> <li>- Signaler et mettre au jour le discours de haine</li> </ul> | <ul style="list-style-type: none"> <li>- Réagir face au discours de haine</li> <li>- Écrire contre le discours de haine</li> <li>- Changer la narration du discours de haine</li> <li>- Assurer un suivi des médias</li> </ul> |

Tableau 2 : Objectifs et buts éducatifs

Le premier objectif éducatif porte sur la diffusion de l'information sur le discours de haine, le deuxième sur l'analyse critique du phénomène, et le troisième encourage les individus à prendre des mesures spécifiques. Les initiatives permettant de répondre à l'objectif éducatif relatif à l'information consistent à sensibiliser au discours de haine, à ses différentes formes et à ses conséquences possibles. Elles donnent aussi des informations sur les cadres juridiques nationaux, régionaux et internationaux applicables. Elles peuvent prendre diverses formes : les vidéos « No Hate Ninja Project - A Story About Cats, Unicorns » et « Hate Speech » du Mouvement contre le discours de haine<sup>133</sup>, par exemple ; le tutoriel interactif en ligne « Faire face à la haine sur Internet » d'HabiloMédias<sup>134</sup>; ou la boîte à outils mise au point dans le cadre du projet « In Other Words »<sup>135</sup>.

Le deuxième objectif éducatif est plus complexe et porte sur l'analyse du discours de haine sur Internet. Cette analyse comporte des évaluations des différents types de discours de haine sur Internet, qui expriment notamment du racisme, du sexisme et de l'homophobie, et des multiples formes qu'il peut prendre. Un aspect important de l'analyse est l'examen critique du discours de haine, qui permet d'en déterminer les causes courantes et de comprendre quels sont les présupposés et préjugés qui le sous-tendent. Ce processus analytique permet aux individus de signaler et de mettre au jour le contenu haineux publié sur Internet. Des exemples de projets répondant à cet objectif éducatif sont le forum de discussion « No Hate » et la plateforme « Reporting hate speech ». Le forum de discussion, géré par le Mouvement contre la haine sur Internet, permet aux jeunes de débattre de ce qui constitue un contenu haineux et de montrer quelques exemples qu'ils ont déjà rencontrés de ce type de discours en ligne. L'extrait suivant provient d'un forum de discussion : « Le droit d'exprimer son opinion est un droit de l'homme essentiel, qui se fonde sur la liberté d'expression, qui nous est chère que nous estimons [...]. Cependant, il semblerait que la distinction n'ait pas bien été faite entre les mots utilisés pour lancer le débat et les mots utilisés pour inciter à la haine. Le phénomène de mobilisation de foules dans le but de perpétrer des actions violentes par l'expression d'une hostilité ouverte à l'égard de groupes ciblés a largement été exploité par des personnes en position d'être entendues »<sup>136</sup>.

Toujours sur la question de l'analyse, la plateforme conçue par l'*Online Hate Prevention Institute* permet à des particuliers de signaler et de surveiller le discours de haine sur Internet en mettant au jour ce qu'ils perçoivent comme un contenu offensant, en effectuant un suivi de certains sites Web, forums et groupes, et en examinant les supports incitants à la haine mis au jour par d'autres personnes.

Enfin, le troisième objectif éducatif déterminé dans ces initiatives porte sur la promotion de mesures pouvant être prises pour lutter contre les actes de langage haineux. Les ressources relevant de cet objectif éducatif visent à promouvoir des actions et réponses concrètes discours de haine sur Internet. Les actions proposées varient en fonction de l'orientation particulière du projet et de l'organisation, selon qu'elle est plus de nature plus ou moins combative ou dans la confrontation ; néanmoins, l'aspect le plus important reste de donner aux particuliers les moyens de répondre aux propos haineux et de lutter activement contre. On citera comme exemple des sessions de formation pour blogueurs, journalistes et militants organisées par le Mouvement contre le discours de haine ; les supports pédagogiques et plans d'enseignement mis au point par HabiloMédias ; et les politiques de surveillance des médias proposées par le projet « In Other Words ».

## *Évaluer l'éducation de base au bon usage des médias et de l'information et les initiatives éducatives*

Si certaines organisations et initiatives se concentrent sur le contenu du discours de haine sur Internet, d'autres mettent l'accent sur l'aspect personnel de ce discours, en attirant l'attention sur les victimes ou sur l'incidence générale sur la communauté. Quel que soit leur point de mire, la plupart des projets considèrent le développement de compétences numériques comme un élément essentiel pour prévenir, mettre au jour et combattre le discours de haine sur Internet. Les outils et stratégies analysés montrent toutes sortes de moyens de développer de telles compétences, des simples modes d'emploi les plus élémentaires aux formations les plus complexes et spécialisées. La vaste gamme de formats étudiée dans le cadre des différentes d'initiative (vidéos, blogs, sites Web, jeux vidéos, et médias sociaux) permettent d'atteindre et d'attirer des publics très divers.

Il n'existe cependant aucune évaluation exhaustive, et il est difficile de savoir dans quelle mesure ces initiatives réussissent à lutter contre le discours de haine ou à toucher les groupes qui sont les plus susceptibles de se livrer à ce genre de discours sur Internet. Par exemple, même si les initiatives et ressources d'HabiloMédias ont reçu de multiples récompenses et distinctions, rien n'indique clairement qui fait le meilleur usage de ses ressources, et il est difficile d'évaluer les résultats des programmes. Dans le cas du projet « In Other Words », les résultats escomptés prévoyaient la mise au point de matériel de diffusion, mais on ne dispose d'aucune information pour savoir comment ce matériel a été utilisé depuis sa publication ou quels publics ont été atteints. Par ailleurs, dans le cas du Mouvement contre le discours de haine, qui a mis au point différents supports et ressources (vidéos, manuels de formation, outils pédagogiques, et la plateforme en ligne servant à signaler les contenus haineux), aucune instruction claire ou précise n'explique comment évaluer l'incidence ou s'en rendre compte<sup>137</sup>. Si la plupart de ces initiatives sont louables et ont le potentiel d'offrir de puissants outils de lutte contre le discours de haine au niveau structurel, il faudra disposer de plus d'informations pour comprendre comment les individus intègrent les compétences nouvellement acquises dans leur vie quotidienne et quelle incidence cette intégration a effectivement sur leur activité en ligne.

# 6. CONCLUSION ET RECOMMANDATIONS

L'apparition et la diffusion du discours de haine sur Internet sont un phénomène qui ne cesse d'évoluer, et il est nécessaire d'entreprendre des efforts collectifs pour comprendre son importance et ses conséquences, ainsi que pour trouver des solutions efficaces. Les manifestations publiques d'indignation sont une réaction courante, certains personnages publics appelant à des sanctions renforcées à l'encontre des personnes qui propagent des messages de haine, et à des contrôles plus stricts sur la communication en ligne. Comme cette étude l'a suggéré, cependant, à se concentrer exclusivement sur les mesures de répression, on risquerait de passer à côté de la complexité d'un phénomène encore mal compris qui demande des réponses sur mesure et coordonnées de la part de tout un éventail d'acteurs de la société. Les espaces en ligne, compte tenu de leur potentiel de favoriser les échanges, et du fait qu'ils offrent des quantités sans précédent de données pouvant être analysées au moyen de toutes sortes de nouvelles techniques, offrent un observatoire unique des comportements humains. Les solutions concrètes doivent être fondées sur une meilleure compréhension de la manière dont les différentes formes d'expression se font jour, interagissent, et éventuellement se dissipent dans cet environnement.

La présente étude a proposé plusieurs exemples concrets montrant comment différentes situations ont amené à trouver des solutions sur mesure. L'apparition de chaque solution est liée à des circonstances uniques, mais leur analyse et leur diffusion offre une palette complète de méthodes adaptables à différents contextes par les diverses parties prenantes. Ce chapitre final résume les principales conclusions que l'on peut tirer de cette étude concernant les moyens de résoudre une partie des principales tensions que révèlent le discours de haine sur Internet.

## Définition

- Le problème de la définition de la notion de discours de haine tient à ce qu'elle se situe à la conjonction de multiples éléments contradictoires. Comme il a été souligné par certains participants à cette étude, et par des processus similaires au Plan d'action de Rabat, il serait utile d'élaborer des définitions qui soient largement partagées par un vaste éventail de parties prenantes, et de mettre au point des tests qui pourraient permettre de déterminer ce qui doit et ce qui ne doit pas être assimilé à des propos haineux. Ce type de processus permet des réponses coordonnées, bien qu'il faille concéder que même les institutions internationales

se sont généralement **gardées** de fournir des **définitions trop strictes**. Ce principe de précaution semble être partagé par certains des acteurs les plus importants du secteur privé qui façonnent la communication en ligne. Bien qu'on leur demande de faire preuve de transparence dans leurs procédures de modération de contenus, les plateformes de réseaux sociaux ont jusqu'à présent évité de proposer des règles et des procédures trop strictes pour déterminer quel type de contenu devait être supprimé. À l'inverse, certaines d'entre elles ont essayé de mutualiser le processus de modération, permettant aux utilisateurs de résoudre certaines controverses en échangeant entre eux grâce à la plateforme. Des nuances sont ainsi possibles et évitent que l'on ne s'enferme dans une approche trop mécanique.

- Des définitions plus étroites ont été proposées, notamment pour le terme de « **discours dangereux** » - qui fait référence à des actes de langage dont la probabilité de catalyser ou d'amplifier la violence d'un groupe à l'égard d'un autre est élevée-, ou celui de « **discours de peur** », qui désigne les propos instillant au sein d'un groupe la peur qu'un autre groupe fomente un recours à la violence. Des notions comme celle-ci trouvent écho dans les principes juridiques, mais plutôt que de cibler la poursuite en justice des contrevenants, elles visent à offrir des outils permettant de repérer et de décrire un type de discours particulier, avec la possibilité de signaler des cas critiques ou des zones de danger où il peut être nécessaire d'apporter des solutions collectives pour éviter la propagation de la violence. Pour ce qui est du discours de haine sur Internet, cependant, toutes les définitions abordent toujours la question sans réponse des **liens entre** les expressions de haine sur Internet et le **véritable dommage** causé (hostilité, discrimination, violence...). C'est un problème partagé par d'autres médias, mais certains des éléments qui caractérisent la communication en ligne (notamment l'anonymat réel ou perçu des utilisateurs, et la façon dont un message donné peut atteindre instantanément de larges publics) rendent les choses particulièrement complexes. Comme l'illustre l'exemple du Kenya évoqué plus haut, les incitations à la haine et les appels à la violence en ligne pendant les élections de 2013 ne se sont pas traduits par des violences réelles comme il semble que cela a été le cas lors de la précédente campagne électorale. D'autres facteurs, notamment peut-être les nombreux appels à la paix qui ont émané de différentes sphères de la société, semblent avoir pesé plus lourd dans la balance. Comme évoqué plus loin avec la question de la *compréhension* du discours de haine sur Internet, cela ne veut pas dire qu'il est impossible d'étudier et de comprendre progressivement les liens entre violence en ligne et violence réelle, mais nous ne disposons pas encore de recherches systématiques à ce sujet.
- L'accent mis sur le **potentiel** d'un acte de langage de conduire à la violence et de causer du tort devrait aussi être envisagé avec ses hypothèses intrinsèques, qui peuvent conduire à une vision étroite se cantonnant à des considérations d'**ordre**

**public.** Mettre l'accent sur la violence peut conduire à trouver des solutions qui privilégient l'État (en tant qu'acteur exerçant le contrôle légitime du recours à la violence), au détriment éventuel d'autres acteurs susceptibles de proposer des solutions différentes ou complémentaires. En outre, comme il a été souligné au début de ce rapport, la législation relative au discours de haine tend à être l'expression du groupe dominant qui contrôle la teneur de la loi. D'autres interprétations du discours de haine se sont plutôt concentrées sur le respect de la dignité humaine, donnant aux victimes des actes de langage le pouvoir d'exiger le respect et d'être défendues, les plaçant elles, plutôt que l'État ou un autre acteur, au centre des solutions concrètes. Cette approche n'est pas dénuée de problèmes et de contradictions, car un accent excessif sur la dignité peut conduire à une surenchère de relativisme - ou de soutien à des idées particularistes qui ne sont pas en accord avec les droits de l'homme. On suggère néanmoins que lorsqu'on s'attaque au discours de haine sur Internet, différentes perspectives devraient être prises en considération et comparées les unes aux autres pour ce qui est d'expliquer le phénomène et ses liens complexes avec la violence réelle, et d'offrir des solutions qui ne s'appliquent pas au détriment d'une approche globale.

- Paradoxalement, la complexité associée à la définition de la notion de discours de haine est aussi l'occasion de développer des **interprétations locales communes** des différentes normes internationales sur ce discours. Celui-ci opère comme une sorte de « **signifiant vide** » (Laclau et Mouffe, 1985). C'est un terme qui peut sembler aller de soi, mais pour lequel les gens donnent généralement des descriptions très diverses lorsqu'on leur pose la question. Cela peut poser problème, par exemple si l'on émet des accusations de messages haineux aux fins de discréditer un propos légitime ou de justifier la censure. Voilà des cas où la critique ou la ridiculisation d'individus, d'opinions ou de croyances peuvent être étiquetés comme discours de haine – bien qu'ils aillent bien au-delà des paramètres définis par le Pacte international relatif aux droits civils et politiques. Les caractéristiques du terme en tant que signifiant vide, cependant, peuvent aussi offrir des possibilités à différents acteurs de se réunir et de débattre de questions qu'il serait difficile d'aborder autrement.

## Juridiction

- La responsabilité du repérage du discours de haine sur Internet et de la riposte à celui-ci a pour une large part été placée sur les épaules des gouvernements. Les **intermédiaires Internet**, c'est-à-dire les services qui jouent un rôle de médiateur dans la communication en ligne, cependant, ont pesé de plus en plus lourd dans l'équation, aussi bien en autorisant qu'en limitant certains propos. Beaucoup d'entre

eux, en particulier les moteurs de recherche et les plateformes de réseaux sociaux, étendent leurs services d'un pays à l'autre et réglementent les interactions des utilisateurs en fonction de leurs propres conditions d'utilisation et définitions du discours de haine. Ils comptent largement sur l'intervention des utilisateurs pour leur signaler les contenus considérés comme répréhensibles, et lorsqu'un cas est porté à leur attention, la réponse par défaut est de se prononcer sur la base de leurs propres conditions d'utilisation. Cependant, les modalités de fonctionnement des intermédiaires Internet quant aux règles et réglementations nationales et internationales, aux groupes de pressions et aux utilisateurs individuels sont en constantes évolution. Comme cette étude l'a indiqué, dans de nombreux exemples récents, les intermédiaires Internet ont réagi sous la pression de différents acteurs.

- Les répercussions de la décision rendue en 2014 par la Cour européenne de justice contre Google, l'entreprise étant désormais obligée de retirer, sur demande, les résultats de recherche sur telle ou telle personne si elle juge que l'information n'est pas pertinente ou obsolète, montre qu'il est possible d'exercer un pouvoir juridictionnel là où pendant longtemps on semblait pouvoir échapper aux autorités gouvernementales. La territorialisation **formelle** des espaces en ligne, cependant, est particulièrement controversée. La raison principale en est qu'elle pourrait conduire à une fragmentation progressive de l'Internet, des États ou groupes d'États imposant leurs propres règles et détruisant le potentiel du Web d'échanges par-delà les frontières et de rapprochement entre les populations. On assisterait à un scénario où l'Internet serait vécu différemment d'un endroit à l'autre, et où la norme de libre circulation des informations serait assujettie à l'exceptionnalisme national ou régional. Les entreprises privées elles-mêmes, ainsi que de nombreux acteurs de la société civile, semblent particulièrement mal à l'aise lorsque des organismes privés sont mandatés pour agir comme des tribunaux et décider de ce qui doit ou non être proposé en ligne. Le débat est en cours pour déterminer dans quelle mesure on peut s'éloigner de l'autoréglementation volontaire, par laquelle les entreprises mettent leurs propres voies de recours à disposition des plaignants individuels, même si ces derniers conservent le droit de résoudre un différend particulier en passant par les tribunaux en cas d'échec.
- La plupart des intermédiaires Internet préfèrent adopter une position plus **informelle** pour répondre aux demandes émanant de gouvernements, de groupes, ou de particuliers. Facebook, par exemple, a activé une fonction de « social reporting » offrant aux utilisateurs la possibilité d'envoyer un message à une personne publiant des informations qu'ils n'apprécient pas mais qui ne violent pas les conditions d'utilisation de Facebook<sup>138</sup>. Une autre fonction liée au droit de reproduction permet à des particuliers, via l'intermédiaire Internet, d'exiger le retrait ou de contester l'emploi d'une expression spécifique<sup>139</sup>. Comme cette étude l'a montré, on a observé de nombreux cas où les plateformes de réseaux sociaux ont modifié ou amélioré

les mécanismes de surveillance et de modération de contenus. Notamment Twitter, qui a fini par accepter d'ajouter un bouton pour signaler des contenus directement depuis un tweet. Cette façon de faire a impliqué divers degrés de collaboration avec les gouvernements, mais dans ces cas le caractère informel a pu être mis en avant pour limiter la responsabilité et l'obligation de transparence tant des États que des entreprises privées. Si dans certains cas ce mode de fonctionnement informel est parfaitement adapté au caractère fluide du discours de haine sur Internet, il a le désavantage d'une application au cas par cas et sporadique. Parfois, la différence peut tenir à la capacité d'un groupe de pression de toucher la corde sensible, et non à l'importance ou à la recevabilité d'une cause spécifique en soi, ou à l'enfreinte des restrictions internationales en matière d'exercice légitime de la liberté d'expression.

## Compréhension

- Le caractère inadmissible des messages de haine donne apparemment de fortes raisons de limiter ceux-ci et de faire taire leurs auteurs en les bannissant d'une plateforme, voire d'Internet. Ces raisons, malgré le fait qu'elles puissent être disproportionnées et qu'elles ne répondent donc pas au critère de nécessité qui préside à la légitimité d'une restriction, ont tendance à prendre de l'ampleur après des événements dramatiques. Dans ces moments-là, les autorités peuvent appeler à prendre des mesures fortes visant à mettre un frein à la capacité d'Internet de propager la haine et la violence, même si les liens entre violence en ligne et violence réelle sont peut-être ténus. Dans ce contexte, les efforts déployés pour comprendre le phénomène du discours de haine avec pour objectif non pas simplement de le contrer et de l'éliminer, mais aussi de comprendre de quoi il est l'expression, sont particulièrement difficiles – ils restent néanmoins très importants. La présente étude a évoqué les recherches portant sur l'**identité** des personnes occupant les espaces extrémistes sur Internet, sur les **raisons** qui les poussent à dire ce qu'ils disent, sur la **façon** dont ils interprètent ce discours, présentant des conclusions souvent contre-intuitives<sup>140</sup>. De telles études sont encore rares, mais une **meilleure compréhension** des mécanismes qui peuvent conduire à certains types de langages peut offrir une vision bien plus riche qui peut aussi inspirer des solutions innovantes qui ne se fondent pas uniquement sur la répression et la réduction au silence. Par exemple, y a-t-il des liens entre les **inégalités économiques** et le **discours de haine** ? Comment certaines personnes peuvent-elles exploiter avec succès le discours de haine à des fins partisans, et pourquoi nombre de leurs victimes sont-elles la plupart du temps issues de milieux vulnérables ou défavorisés<sup>141</sup> ? Y a-t-il un rapport entre l'**accès à l'éducation** et le **discours de haine** ? Les réponses à des questions comme celles-ci peuvent offrir des solutions qui montrent l'utilité

de politiques concrètes de plus grande inclusion sociale, plutôt que de mesures ciblant exclusivement le discours de haine, qui n'est qu'un symptôme de maux plus profonds encore peu étudiés.

- Le discours de haine sur Internet recouvre un ensemble très large de phénomènes observables sur différentes plateformes technologiques. Les architectures sur lesquelles reposent ces plateformes, cependant, peuvent grandement varier et ont des répercussions importantes sur la manière dans le discours de haine se propage et peut être contrôlé. Comme le projet UMATI évoqué au chapitre 4 l'indique, l'architecture de Twitter a permis aux utilisateurs de rejeter facilement ceux qui publient des messages de haine et des appels à la violence. L'utilisation des hashtags, qui rassemblent différents utilisateurs autour d'un sujet commun, offre la possibilité de partager un tweet ou tout un compte, permettant un processus de modération collective. Dans le cas de Facebook, de tels mécanismes sont moins susceptibles de se mettre en place étant donné que la plateforme permet de développer plusieurs fils de conversations partiellement cachés les uns des autres. D'autres plateformes, comme Snapchat, limitent la possibilité de partager un message en dehors d'un cercle très restreint. Une compréhension plus fine de la façon dont chaque plateforme peut autoriser ou limiter la production et la diffusion de différents types de messages peut ainsi représenter un facteur très important dans l'élaboration de solutions appropriées.
- Les grandes plateformes de réseaux sociaux comme Facebook et Twitter ont principalement adopté une démarche réactive dans la gestion des messages de haine signalés par leurs utilisateurs, en déterminant si oui ou non ces messages contrevenaient à leurs conditions d'utilisation. Ces plateformes pourraient, cependant, jouer un rôle plus proactif. Elles ont accès à des quantités colossales de données qui peuvent être mises en relation, analysées, et associées à des événements du monde réel qui permettraient une compréhension plus nuancée des mécanismes qui caractérisent le discours de haine en ligne. De très nombreuses données sont déjà collectées et mises en relation à des fins commerciales. Des efforts similaires pourraient être faits dans le cadre de la responsabilité sociale des entreprises qui possèdent ces plateformes, ce qui contribuerait à accumuler des connaissances nouvelles qui pourraient être partagées avec un large éventail de parties prenantes.
- Des initiatives préconisant une meilleure éducation de base aux médias et à l'information ont commencé à voir le jour comme solution plus structurelle au problème du discours de haine sur Internet. Compte tenu de l'exposition croissante des jeunes aux médias sociaux, les informations sur la façon de repérer le discours de haine et d'y réagir pourrait prendre une importance croissante. Si certains établissements scolaires ont manifesté leur intérêt pour l'intégration progressive d'une éducation de base aux médias et à l'information dans leurs programmes, ces

initiatives restent éparpillées et n'atteignent souvent pas les plus vulnérables qui ont le plus besoin d'être mis en garde contre les risques du discours de haine sur Internet et en dehors. Il est particulièrement important que ces modules anti discours de haine soient intégrés dans les pays où le véritable risque de violence généralisée est le plus élevé. Il est aussi nécessaire d'intégrer à ces programmes des modules engageant une réflexion sur l'identité, de façon que les jeunes soient capables de détecter les tentatives de manipulation de leurs émotions en faveur de la haine, et qu'ils aient les moyens de faire valoir leur droit individuel d'avoir eux-mêmes le contrôle de ce qu'ils sont et de ce qu'ils veulent devenir. Des initiatives anticipées et préventives comme celles-ci devraient aussi être accompagnées de mesures destinées à évaluer l'incidence sur le comportement réel des élèves, en ligne et en dehors, ainsi que sur leur capacité à repérer les messages de haine et à y répondre.

## Intervention

- Le problème du discours de haine sur Internet exige des **solutions collectives**. Cette conclusion peut paraître évidente et s'appliquer à de nombreuses autres sphères de la vie sociale. Comme l'a montré cette étude, cependant, certains éléments propres au discours de haine sur Internet sont de nature à rendre très inefficaces les solutions où l'intégralité de la responsabilité est confiée à un acteur unique ou à un nombre restreint d'acteurs.
- L'**Internet s'étend au-delà des frontières**, et malgré des tentatives répétées par des États de réaffirmer leur souveraineté dans le domaine numérique, des problèmes complexes comme celui de la lutte contre le discours de haine sur Internet ne peuvent être résolus facilement, en comptant simplement sur le pouvoir étatique. Par exemple, repérer et poursuivre toutes les personnes qui publient des messages de haine ne serait pas réalisable pour la plupart des États.
- Comme l'a proposé la Rapporteuse spéciale sur les questions relatives aux minorités du Conseil de droits de l'homme des Nations Unies, les **États** pourraient œuvrer de concert avec les organisations et projets qui mènent des campagnes de lutte contre le discours de haine, y compris sur Internet, notamment en leur fournissant un soutien financier (HRC, 2015).
- Les **intermédiaires Internet**, de leur côté, ont intérêt à entretenir une relative indépendance et une **image « sans reproche »**. Ils ont cherché à atteindre cet objectif en montrant leur réactivité face aux pressions des groupes de la société civile, particuliers et gouvernements. Jusqu'à présent, ces négociations se sont toutefois déroulées au cas par cas, et n'ont pas abouti à l'élaboration de principes généraux applicables collectivement.

- Comme l'on suggéré certains individus interrogés dans le cadre de cette étude, de nombreux utilisateurs semblent avoir été **anesthésiés** par la récurrence du **discours de haine sur Internet**. Des initiatives plus structurelles sont nécessaires pour expliquer non seulement comment certaines occurrences peuvent être signalées, mais aussi pourquoi il est important de créer des espaces de dialogue sur le discours de haine. Il semble par ailleurs qu'il soit possible de consolider la position intermédiaire pour la maintenir éloignée des extrêmes haineux, grâce à l'action de militants s'engageant dans le débat avec des arguments pour contrer le discours de haine sur Internet<sup>142</sup>.

---

- La définition du discours de haine sur Internet continuera probablement encore longtemps d'échapper à un consensus universel, mais il reste possible de faire face au phénomène qu'elle recouvre sur une **base** qui fasse intervenir de multiples parties prenantes. Il est évident qu'aucun acteur unique ne peut résoudre le problème du discours de haine qui se développe sur Internet, quelle que soit la définition qu'on en donne. Les processus participatifs actuels visant à comprendre la nature de ce discours sont aussi importants que les solutions évolutives trouvées pour répondre au phénomène.

---

# Auteurs et remerciements

**Iginio Gagliardone** est chercheur dans le domaine des nouveaux médias et des droits de l'homme au centre d'études socio-juridiques et membre du Programme de droit et politique comparés des médias (Comparative Media Law and Policy - PCMLP) à l'Université d'Oxford. Ses recherches et publications portent sur la relation entre les médias et les changements politiques, en particulier en Afrique subsaharienne, et sur l'émergence de différents modèles de la société de l'information dans le monde. Il est aussi chercheur associé de l'*Oxford Internet Institute*, du Centre pour la gouvernance et les droits de l'homme de l'Université de Cambridge, et du Centre d'études sur la communication mondiale de l'école de communication d'Annenberg de l'Université de Pennsylvanie.

**Danit Gal** est spécialisée dans la sécurité informatique et la guerre de l'information. Elle a mené des recherches très poussées sur la propagande et les messages extrémistes en ligne en tant que boursière de l'*Argov Fellows Program in Leadership and Diplomacy* et de la fondation Ragonis au Centre interdisciplinaire d'Herzliya, en Israël.

**Thiago Alves Pinto** mène actuellement des recherches dans le cadre de son doctorat de philosophie, spécialité droit, de l'Université d'Oxford. Il a travaillé pour le Rapporteur spécial sur la liberté de religion ou de conviction et l'Organisation internationale pour les migrations, en Finlande, sur des projets de développement et de lutte contre la traite des êtres humains.

**Gabriela Martínez Sainz** est en doctorat de philosophie à l'Université de Cambridge. Elle travaille sur l'éducation aux droits de l'homme au Mexique, dans l'objectif d'intégrer la théorie dans les pratiques professionnelles en appliquant les connaissances acquises à l'élaboration de programmes éducatifs et de programmes de formation à l'intention des éducateurs dans le domaine des droits de l'homme. Elle est l'auteur de manuels scolaires d'éducation à la citoyenneté et à l'éthique destinés aux élèves de l'enseignement secondaire au Mexique.

Les auteurs sont profondément reconnaissants aux nombreuses personnes qui ont contribué à la réalisation de cette étude. Nous sommes redevables aux nombreuses personnes interrogées qui ont pris le temps de s'entretenir avec nous, parfois à plusieurs reprises, malgré leurs nombreux engagements. Nous avons eu la chance de pouvoir compter sur un impressionnant comité consultatif, dont les membres nous ont donné des orientations et des retours au fur et à mesure que nous élaborions ce rapport, notamment Monroe Price, Richard Danbury, Cherian George, Nazila Ghanea, Robin Mansell, Bitange Ndemo et Nicole Stremlau. Nous sommes également reconnaissants aux nombreuses personnes, au sein de l'UNESCO, qui nous ont fait part de leurs observations sur les versions antérieures de ce rapport, nous aidant à le rendre plus solide et à faire en sorte qu'il contribue au débat plus vaste sur la liberté en ligne.

# Liste des œuvres citées :

- Albanian Media Institute. 2014. *Hate speech in online media in South East Europe*. <http://www.institutemedia.org/Documents/PDF/Hate%20speech%20in%20online%20media%20in%20SEE.pdf>
- Aung, S. Y. 2014. « Hate Speech Pours Poison Into the Heart », <http://www.irrawaddy.org/interview/hate-speech-pours-poison-heart.html>.
- Australian Copyright Council. 2013. *Websites: Social Networks, Blogs and user generated content*. Information sheet G108v03, juin 2013.
- Benesch, S. 2012. *Dangerous Speech: A Proposal to Prevent Group Violence*, World Policy Institute, New York, n° 2, <http://www.worldpolicy.org/sites/default/files/Dangerous%20Speech%20Guidelines%20Benesch%20January%202012.pdf>;
- Benesch, S. 2012. *Words as Weapons*, World Policy Journal, vol. 29, n° 1, p. 7 à 12.
- Bleich, E. 2013. *Freedom of Expression versus Racist Hate Speech: Explaining Differences Between High Court Regulations in the USA and Europe*, *Journal of Ethnic and Migration Studies*, p.2 à 37.
- Bowman-Grieve, L. 2009. *Exploring 'Stormfront': A Virtual Community of the Radical Right*, *Studies in Conflict & Terrorism*, n° 11, p. 989 à 1007.
- Buckels, E.E., Trapnell, P.D. & Paulhus, D.L., 2014. *Trolls just want to have fun*. *Personality and Individual Differences*, vol. 67, p. 97 à 102.
- Buyse, A., 2014. *Words of Violence: "Fear Speech" or How Violent Conflict Escalation Relates to the Freedom of Expression*. *Human Rights Quarterly*, vol. 36, n° 4, p.779 à 797.
- Citron, K. D. et Norton, H. 2011. *Intermediaries and hate speech: Fostering digital citizenship for our information age*. *Boston University Law Review*, vol. 91, p. 1435 à 1484.
- Chang, J. et al., 2014. *The Causes and Consequences of Group Violence: From Bullies to Terrorists*, Lexington Books.
- Cheeseman, N. 2008. *The Kenyan Elections of 2007: An Introduction*, *Journal of Eastern African Studies* 2, n° 2, p. 166 à 184.
- Chorley, M. et Camber, R. 2014. *Anti-terror unit to push to remove hate video from YouTube as Cameron vows to drive out "extremist poison"*. <http://www.dailymail.co.uk/news/article-2665335/Video-hate-STILL-YouTube-Online-site-failing-stop-terrorists.html>.
- Cao, Q. Sirivianos, M. Yang, X. et Pregoiero, T. 2011. *Aiding the detection of fake accounts in large scale social online services*. [https://www.usenix.org/system/files/conference/nsdi12/nsdi12-final42\\_2.pdf](https://www.usenix.org/system/files/conference/nsdi12/nsdi12-final42_2.pdf)
- Defeis, E.F., 1992. *Freedom of speech and international norms: A response to hate speech*. *Stan. Journal of International Law.*, vol. 29, p. 57 à 74.
- De Koster, W. et Houtman, D. 2008. « "STORMFRONT IS LIKE A SECOND HOME TO ME" On Virtual Community Formation by Right-Wing Extremists », *Information, Communication & Society*, vol. 11, n° 8, p. 1169 à 1185.
- Duggan, M. 2014. *Online Harassment: Summary of findings*. <http://www.pewinternet.org/2014/10/22/online-harassment/>
- Epstein, K. 2014. « Twitter teams up with advocacy group to fight online harassment of women ». <http://www.theguardian.com/technology/2014/nov/08/twitter-harassment-women-wam>

- Foxman, A.H. & Wolf, C., 2013. *Viral hate: Containing its spread on the Internet*. Macmillan.
- Erjavec, K. & Kovačič, M.P., 2012. « You Don't Understand, This is a New War! » *Analysis of Hate Speech in News Web Sites' Comments*. *Mass Communication and Society*, vol. 15, n° 6, p.899 à 920.
- Herman, B. 2014. « WAM! and Twitter tackle problem of online harassment of women ». <http://www.ibtimes.com/wam-twitter-tackle-problem-online-harassment-women-1720876>
- Hindstrom, H. 2012. « State Media Issues Correction after Publishing Racial Slur », *DVB Multimedia Group*, 6 juin 2012, <https://www.dvb.no/news/state-media-issues-correction-after-publishing-racial-slur/22328>.
- Hochsmann, M. et Poyntz, S. R. 2012. *Media Literacies. A Critical Introduction*. Oxford: Wiley-Blackwell.
- Holland, H. 2014. « Facebook in Myanmar: Amplifying Hate Speech? », 14 juin 2014, <http://www.aljazeera.com/indepth/features/2014/06/facebook-myanmar-rohingya-amplifying-hate-speech-2014612112834290144.html>.
- Ghanea, N., 2008. Articles 19 et 20 du Pacte international relatif aux droits civils et politiques. In *Séminaire d'experts sur les relations entre les articles 19 et 20 du Pacte international*. Genève.
- Ghanea, N., 2013. Intersectionality and the Spectrum of Racist Hate Speech: Proposals to the UN Committee on the Elimination of Racial Discrimination. *Human Rights Quarterly*, vol. 35, n° 4, p.935 à 954.
- Goldsmith, J.L. & Wu, T., 2006. *Who controls the Internet? Illusions of a borderless world*, Oxford University Press New York. [http://jost.syr.edu/wp-content/uploads/who-controls-the-internet\\_illusions-of-a-borderless-world.pdf](http://jost.syr.edu/wp-content/uploads/who-controls-the-internet_illusions-of-a-borderless-world.pdf).
- Goodman, E et Cherubini, F. 2013. *Online comment moderation: emerging best practices. A guide to promoting robust and civil online conversation*. <http://www.wan-iffra.org/reports/2013/10/04/online-comment-moderation-emerging-best-practices>.
- Herring, S. et al., 2002. Searching for safety online: Managing "trolling" in a feminist forum. *The Information Society*, vol. 18, n° 5, p.371 à 384.
- Hochsmann, M. et Poyntz, S. R. 2012. *Media Literacies. A Critical Introduction*. Oxford: Wiley-Blackwell.
- Howell, J. et Lind, J. 2009. *Counter-Terrorism, Aid and Civil Society: Before and after the War on Terror*. New York: Palgrave-Macmillan, 2009.
- HRC. 2015. *Rapport de la Rapporteuse spéciale sur les questions relatives aux minorités, Rita Izsák*. A/HRC/28/64
- In Other Words Project. 2013. *Toolbox*. Disponible en ligne à l'adresse suivante : <http://www.inotherwords-project.eu/sites/default/files/Toolbox.pdf>
- iHub Research. 2013. *Umati Final Report* (Enquête iHub de 2013, Rapport final du projet UMATI). [http://www.research.ihub.co.ke/uploads/2013/june/1372415606\\_936.pdf](http://www.research.ihub.co.ke/uploads/2013/june/1372415606_936.pdf).
- Kanyinga, K. 2009. The Legacy of the White Highlands: Land Rights, Ethnicity and the Post-2007 Election Violence in Kenya, *Journal of Contemporary African Studies*, vol. 27, n° 3, p. 325 à 344.
- Kelion, L. 2013. *Facebook lets beheading clips return to social network*. <http://www.bbc.co.uk/news/technology-24608499> (consulté le 15 décembre 2014.)
- Kellow, C.L. & Steeves, H.L., 1998. The role of radio in the Rwandan genocide. *Journal of Communication*, vol. 48, n° 3, p.107 à 128.
- Kiersons, S. 2013. « The Colonial Origins of Hate Speech in Burma », 28 octobre 2013, <https://thesentinelproject.org/2013/10/28/the-colonial-origins-of-hate-speech-in-burma/>.

- Kinetz, E. 2013. « New Numerology of Hate Grows in Burma », 29 avril 2013, <http://www.irrawaddy.org/religion/new-numerology-of-hate-grows-in-burma.html>.
- Kohl, U., 2002. Eggs, Jurisdiction, and the Internet. *International and comparative law quarterly*, vol. 51, n° 3, p.556 à 582.
- Ko Ko, M. 2014. Entretien du 3 novembre 2014. (Justice Base)
- Laclau, E. et Mouffe, C. 1985. *Hegemony and Socialist Strategy. Towards a Radical Democratic Politics*. Londres, Verso, 1985.
- Leo, L.A., Gaer, F.D. & Cassidy, E.K., 2011. Protecting Religions from Defamation: A Threat to Universal Human Rights Standards. *Harv. JL & Pub. Pol'y*, vol. 34, p.769 à 795.
- MacKinnon, R. et al., 2015. *Fostering Freedom Online: The Role of Internet Intermediaries*, Paris, UNESCO.
- Mamdani, M. 2010. *Saviors and Survivors: Darfur, Politics, and the War on Terror*. Random House LLC.
- Meddaugh, P. M. et Kay, J. 2009. Hate Speech or "Reasonable Racism"? The Other in Stormfront, *Journal of Mass Media Ethics*, vol. 24, n° 4, p. 251 à 268. HabiMédiAs.NDA.
- HabiMédiAs. Faire face à la haine sur Internet. Disponible en ligne à l'adresse suivante : <http://habilomedias.ca/tutoriel/faire-face-haine-sur-internet>
- HabiMédiAs. *Prix et mentions*. Disponible en ligne à l'adresse suivante : <http://habilomedias.ca/qui-nous-sommes/prix-et-mentions>
- Morsink, J., 1999. *The universal declaration of human rights: Origins, drafting, and intent*. University of Pennsylvania Press.
2012. « Contents removed from racist Facebook page ». Disponible en ligne à l'adresse suivante : <http://www.smh.com.au/technology/technology-news/contents-removed-from-racist-facebook--page-20120808-23tr1.html>
- Mossberger, K. Tolbert, C. J. et Moneal, R. S. 2008. *Digital Citizenship. The Internet, Society and Participation*. Londres, The MIT Press, 2008.
- Mouvement contre le discours de haine. 2013a. Outils et supports de campagne. Disponibles en ligne à l'adresse suivante : <http://nohate.ext.coe.int/Campaign-Tools-and-Materials>
- Mouvement contre le discours de haine. 2013b. No Hate Ninja Project - A Story About Cats, Unicorns and Hate Speech. Disponible en ligne à l'adresse suivante : <https://www.youtube.com/watch?v=kp7ww3KvccE>
- Mouvement contre le discours de haine. 2014. Cinquième réunion du groupe de suivi. Disponible en ligne à l'adresse suivante : <http://nohate.ext.coe.int/The-Campaign/Follow-Up-Group-of-the-Joint-Council-on-Youth2>
- Nowak, M., 1993. *UN covenant on civil and political rights: CCPR commentary*. NP Engel Kehl.
- Oboler, A. 2012. An Online Hate Prevention Institute report on Aboriginal Memes and Online Hate. <http://ohpi.org.au/reports/IR12-2-Aboriginal-Memes.pdf>
- Online Hate Prevention Institute. 2014. « Press release: Launch of online tool to combat hate ». <http://ohpi.org.au/press-release-launch-of-online-tool-to-combat-hate/>
- Osborn, M. 2008. Fuelling the Flames: Rumour and Politics in Kibera, *Journal of Eastern African Studies*, vol. 2, n° 2, p. 315 à 327.
- Osler, A. et Starkey, H. 2005. *Changing Citizenship*. Berkshire: Open University Press.

- Peterson, A. et Knowles, C. 2009. « Active Citizenship: A Preliminary Study into Student Teacher Understandings », *Educational Research* 51, n° 1 (mars 2009) : p. 39 à 59, doi:10.1080/00131880802704731.
- Post, R., Hare, I. & Weinstein, J., 2009. Hate speech. In *Extreme speech and democracy*, Oxford University Press, p. 123 à 138.
- Ribble, M. 2011. *Digital Citizenship in Schools: Nine Elements All Students Should Know*. ISTE, 2011.
- Rosen, *The harm in hate speech*, Harvard University Press.
- Rosenfeld, M., 2012. Hate Speech in Constitutional Jurisprudence. In *The Content and Context of Hate Speech*, Cambridge University Press, p. 242 à 289.
- Rowbottom, J., 2012. To Rant, Vent and Converse: Protecting Low Level Digital Speech. *The Cambridge Law Journal*, Vol, 71, n° 2, p.355 à 383.
- Ryngaert, C., 2008. *Jurisdiction in international law*. Oxford University Press.
- Sambuli, N. 2014. Entretien du 18 novembre 2014. (Chef du projet UMATI)
- Schissler, M. 2014a. « Echo Chambers in Myanmar: Social Media and the Ideological Justifications for Mass Violence » (présenté lors du débat intitulé « Communal Conflict in Myanmar: Characteristics, Causes, Consequences », Yangon, 2014). Causes, Consequences, Yangon, 2014
- Schissler, M. 2014b. « May Flowers », *New Mandala*, 17 mai 2014, <http://asiapacific.anu.edu.au/newmandala/2014/05/17/may-flowers/>.
- Schomerus, M., Allen, T. et Vlassenroot, K. 2012. KONY 2012 and the Prospects for Change, *Foreign Affairs*.
- Shin, J., 2008. Morality and Internet Behavior: A study of the Internet Troll and its relation with morality on the Internet. In *Society for Information Technology & Teacher Education International Conference*. p. 2834 à 2840.
- Somerville, K. 2011. Violence, Hate Speech and Inflammatory Broadcasting in Kenya: The Problems of Definition and Identification, *Ecquid Novi: African Journalism Studies*, vol. 32, n° 1, p. 82 à 101.
- Stremlau, N., 2011. The press and the political restructuring of Ethiopia. *Journal of Eastern African Studies*, vol. 5, n° 4, p.716 à 732.
- Stremlau, N., 2012. Somalia: Media law in the absence of a state. *International Journal of Media & Cultural Politics*, vol. 8, n° 2-3, p.159 à 174.
- Tan, M. 2014. « Myanmar's First Non-Government Phone Carrier, Ooredoo, Goes Live », *CNET*, 6 août 2014, <http://www.cnet.com/uk/news/myanmar-first-non-government-phone-carrier-ooredoo-goes-live/>.
- Tsesis, A., 2001. Hate in cyberspace: Regulating hate speech on the Internet. *San Diego L. Rev.*, vol. 38, p. 817 à 854.
- Thompson, A., 2007. *The media and the Rwanda genocide*. CRDI.
- Viljoen, F., 2012. *International human rights law in Africa*. Oxford University Press.
- Waldron, J et Romano, A. 2013. « What will it take for Facebook to care about violence against women? » <http://www.dailydot.com/business/wam-fbrape-violence-against-women-facebook/> (consulté le 22 décembre 2014.)
- Yanagizawa-Drott, D., 2010. Propaganda and conflict: Theory and evidence from the Rwandan genocide. Working Paper, Harvard University.
- Zhang, C. Sun, J. Zhu, X, et Fang, Y. 2010. Privacy and security for online social networks: Challenges and opportunities. *IEEE Network*, numéro de juillet/août.

# Notes

1. <http://www.unesco.org/new/fr/internetstudy>
2. Voir Conseil de l'Europe, "Mapping study on projects against hate speech online", 15 avril 2012. Voir également les entretiens suivants : Christine Chen, Responsable en chef chargée des relations publiques, Google, 2 mars 2015 ; Monika Bickert, Responsable mondiale de la politique de contrôle des contenus, Facebook, 14 janvier 2015 ;
3. Voir HateBase – Hate speech statistics, <http://www.hatebase.org/popular>
4. Entretien : Andre Oboler, Directeur général, Online Hate Prevention Institute, 31 octobre 2014.
5. Voir le chapitre 4, partie 1, pour plus de détails.
6. Il existe des cas de personnes ayant dû purger une peine d'emprisonnement pour des tweets écrits sous l'emprise de l'alcool ou pour avoir tweeté en plaisantant qu'ils prévoyaient de faire exploser un aéroport. (Rowbottom 2012).
7. Entretien : Drew Boyd, Directeur des opérations, The Sentinel Project for Genocide Prevention, 24 octobre 2014.
8. Entretien : Ian Brown, Université d'Oxford, 26 novembre 2014.
9. Il s'agit d'un argument également avancé par la Rapporteuse spéciale sur les droits des minorités, Conseil des droits de l'homme de l'ONU, (CDH, 2015).
10. Déclaration universelle des droits de l'homme, article 7.
11. Déclaration universelle des droits de l'homme, article 19.
12. Annotations au texte des *Projets de pactes internationaux relatifs aux droits de l'homme*, A/2929, 1er juillet 1955, para. 189.
13. Australie, Nouvelle-Zélande, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord. Voir également Gelber, Katharine. *Australia's Response to Articles 19 and 20 of the International Covenant on Civil and Political Rights*. 2011 Expert Workshops on the Prohibition of Incitement to National, Racial or Religious Hatred, Bangkok, 6-7 juillet 2011, p. 3.
14. Comité des droits de l'homme. Observation générale n° 11, Article 20 : Interdiction de toute propagande en faveur de la guerre et de tout appel à la haine nationale, raciale ou religieuse, 29 juillet 1983, paragraphe 2. En 2011, le Comité a précisé son avis concernant les relations entre les articles 19 et 20 lorsqu'il a réaffirmé que les dispositions des deux articles se complétaient et que l'article 20 pouvait « être considéré comme une *lex specialis* à l'égard de l'article 19 » ; Comité des droits de l'homme, Observation générale n° 34, Article 19 : *Liberté d'opinion et liberté d'expression*, CCPR/C/GC/34, 12 septembre 2011, paragraphes 48 à 52.
15. Article 19 (2) du Pacte international relatif aux droits civils et politiques (italiques ajoutées).
16. Il convient de noter que l'expression « de toute espèce » a été reproduite à l'article 13 (1) de la Convention américaine relative aux droits de l'homme, mais pas dans la Convention européenne des Droits de l'Homme ni dans la Charte africaine des droits de l'homme et des peuples.
17. Article 19 (2) du Pacte international relatif aux droits civils et politiques.
18. Comité des droits de l'homme, Observation générale n° 34, *supra* note 14, paragraphe 12.

19. *Ibid.*, paragraphe 15.
20. Article 19 (3) du Pacte international relatif aux droits civils et politiques.
21. Article 5 (1) du Pacte international relatif aux droits civils et politiques.
22. Conseil des droits de l'homme. Rapport du Rapporteur spécial sur la promotion et la protection du droit à la liberté d'opinion et d'expression, M. Frank La Rue, A/HRC/14/23, 20 avril 2010, paragraphe 77.
23. Comité des droits de l'homme, Observation générale n° 34, *supra* note 14, paragraphe 43.
24. Même le Comité des droits de l'homme, qui s'est prononcé sur des cas relatifs à l'article 20, se garde bien de donner une définition de l'incitation à la haine. Conseil des droits de l'homme. *Incitation à la haine raciale et religieuse et promotion de la tolérance* : rapport du Haut-Commissaire aux droits de l'homme, A/HRC/2/6, 20 septembre 2006, paragraphe 36.
25. Faurisson c. France, Opinion individuelle d'Elizabeth Evatt et David Kretzmer, cosignée par Eckart Klein (opinion concordante), paragraphe 4.
26. Conseil des droits de l'homme. Rapport du Haut-commissaire des Nations Unies aux droits de l'homme et suivi de la Conférence mondiale sur les droits de l'homme, Additif, Séminaire d'experts sur les relations entre les articles 19 et 20 du Pacte international relatif aux droits civils et politiques : « *La liberté d'expression et les appels à la haine religieuse qui constituent une incitation à la discrimination, à l'hostilité ou à la violence* », A/HRC/10/31/Add.3, 16 janvier 2009, paragraphe 1.
27. Rapport du Haut-Commissaire aux droits de l'homme, A/HRC/2/6, *supra* note Erreur ! Signet non défini., para. 73
28. « En Russie, par exemple, les ONG signalent régulièrement des cas de recours abusif à la législation contre l'extrémisme dans lesquels les autorités engagent des poursuites contre des groupes religieux [...] pour incitation à la haine ». Petrova, Dimitrina, *Incitement to National, Racial or Religious Hatred: Role of Civil Society and National Human Rights Institutions*. Atelier d'experts sur l'interdiction de l'incitation à la haine nationale, raciale ou religieuse, Vienne, 910 février 2011, paragraphe 3.
29. Principes de Camden sur la liberté d'expression et l'égalité, paragraphe 12 (3).
30. Conclusions et recommandations issues des quatre ateliers régionaux d'experts organisés par le HCDH en 2011, adoptées par les experts à Rabat (Maroc) le 5 octobre 2012. Disponibles en ligne à l'adresse suivante : [http://www.ohchr.org/Documents/Issues/Opinion/SeminarRabat/Rabat\\_draft\\_outcome.pdf](http://www.ohchr.org/Documents/Issues/Opinion/SeminarRabat/Rabat_draft_outcome.pdf)
31. Conclusions et recommandations issues des quatre ateliers régionaux d'experts organisés par le HCDH en 2011, adoptées par les experts à Rabat (Maroc) le 5 octobre 2012. Disponibles en ligne à l'adresse suivante : [http://www.ohchr.org/Documents/Issues/Opinion/SeminarRabat/Rabat\\_draft\\_outcome.pdf](http://www.ohchr.org/Documents/Issues/Opinion/SeminarRabat/Rabat_draft_outcome.pdf)
32. *Ibid.*
33. Rapport du Haut-Commissaire aux droits de l'homme, A/HRC/2/6, paragraphe 39.
34. Comité pour l'élimination de la discrimination raciale, Recommandation générale n° XXIX concernant la discrimination fondée sur l'ascendance (61e session, 2002), U.N. Doc. A/57/18 ; 111 (2002), reproduit dans *Récapitulation des observations générales ou recommandations générales adoptées par les organes créés en vertu d'instruments internationaux relatifs aux droits de l'homme*, U.N. Doc. HRI\GEN\1\Rev.6 ; 223 (2003), paragraphes r, s et t.
35. Article 2 de la Convention sur l'élimination de toutes les formes de discrimination à l'égard des femmes.
36. Recommandation générale n° 28 concernant les obligations fondamentales des États parties découlant de l'article 2 de la Convention sur l'élimination de toutes les formes de discrimination à l'égard des femmes, paragraphe 19.

37. Résolution 17/19 adoptée par le Conseil des droits de l'homme, *Droits de l'homme, orientation sexuelle et identité de genre*, A/HRC/RES/17/19, 14 juillet 2011, préambule.
38. Comité des droits de l'homme, Observation générale n° 28, Article 3 (Égalité des droits entre hommes et femmes), U.N. Doc. CCPR/C/21/Rev.1/Add.10 (2000), paragraphe 22.
39. Article 13 (4) de la Convention américaine relative aux droits de l'homme.
40. Commission interaméricaine des droits de l'homme. Déclaration de principes interaméricaine sur la liberté d'expression, 20 octobre 2000, paragraphe 7.
41. Commission interaméricaine des droits de l'homme, Avis consultatif OC-5/85, 13 novembre 1985, paragraphe 39
42. *Ibid.*, paragraphe 39.
43. *Ibid.*, paragraphe 39.
44. Commission africaine des droits de l'homme et des peuples. Déclaration de principes sur la liberté d'expression en Afrique, 32e session, Banjul, 17 - 23 octobre 2002.
45. Organisation de la Conférence islamique, Déclaration du Caire sur les droits de l'homme en Islam, préambule, 5 août 1990.
46. Organisation de la Conférence islamique, sixième rapport d'observation sur l'islamophobie, présenté au 40<sup>e</sup> Conseil des ministres des affaires étrangères, Conakry (République de Guinée), décembre 2013, p.31.
47. Article 22 de la Déclaration du Caire sur les droits de l'homme en Islam.
48. Ligue des États arabes, Charte arabe des droits de l'homme, 22 mai 2004, entrée en vigueur le 15 mars 2008, paragraphe 32 (1)
49. *Ibid.*, paragraphe 32 (2).
50. Comité des droits de l'homme, Observation générale n° 22, Article 18 (48e session, 1993). *Récapitulation des observations générales ou recommandations générales adoptées par les organes créés en vertu d'instruments internationaux relatifs aux droits de l'homme*, U.N. Doc. HRI/GEN/1/Rev.1 à 35 (1994).
51. Article 7 de la Déclaration des droits de l'homme de l'ASEAN.
52. Haut-Commissariat des Nations Unies aux droits de l'homme, '*ASEAN Human Rights Declaration Should Maintain International Standards*', Lettre ouverte du Comité de coordination des procédures spéciales du Conseil des droits de l'homme sur le projet de Déclaration des droits de l'homme de l'ASEA, 1er novembre 2012.
53. Article 54 de la Charte des droits fondamentaux de l'Union européenne.
54. Article 10 de la Convention européenne des Droits de l'Homme.
55. *Handyside c. Royaume-Uni*, 7 décembre 1976, paragraphe 49. D'autres affaires relatives aux discours de haine devant la Cour européenne des droits de l'homme sont disponibles à l'adresse suivante : [http://www.echr.coe.int/Documents/FS\\_Hate\\_speech\\_FRA.pdf](http://www.echr.coe.int/Documents/FS_Hate_speech_FRA.pdf)
56. Recommandation de politique générale n°6 de l'ECRI sur la lutte contre la diffusion de matériels racistes, xénophobes et antisémites par l'Internet, adoptée le 15 décembre 2000.
57. Conseil de l'Europe, Convention sur la cybercriminalité, 23 novembre 2001, paragraphes 31-34.
58. Conseil de l'Europe, *Protocole additionnel à la Convention sur la cybercriminalité, relatif à l'incrimination d'actes de nature raciste et xénophobe commis par le biais de systèmes informatiques*, 28 janvier 2003, article 5 paragraphe 1.

59. Article 19 (3), alinéas (a) et (b), du Pacte international relatif aux droits civils et politiques (italiques ajoutées).
60. Comité des droits de l'homme, Observation générale n° 34, *supra* note Erreur ! Signet non défini., paragraphe 22.
61. Plan d'action de Rabat, paragraphe 14.
62. Comité des droits de l'homme, Observation générale n° 34, *supra*.
63. [http://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR\\_FR.pdf](http://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_FR.pdf), principe 11
64. [http://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR\\_FR.pdf](http://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_FR.pdf), principe 17.
65. [http://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR\\_FR.pdf](http://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_FR.pdf), principe 22.
66. Natasha Lomas, *#Gamergate Shows Tech Needs Far Better Algorithms*, Techcrunch, 18 octobre 2014. <http://techcrunch.com/2014/10/18/gamergate-tactics/>
67. En mars 2015, par exemple, Facebook a mis à jour ses standards de la communauté afin de donner davantage de précisions concernant certains types de contenus qui font l'objet d'une modération, notamment les messages émanant d'organisations considérées comme « dangereuses ». Voir Leo Kelion, *"Facebook revamps its takedown guidelines"*, BBC, 16 mars 2015, <http://m.bbc.co.uk/news/technology-31890521>.
68. <http://help.yahoo.com/l/us/yahoo/smallbusiness/bizmail/spam/spam-44.html>
69. <https://twitter.com/tos>
70. <https://support.twitter.com/entries/18311>
71. [https://www.youtube.com/t/community\\_guidelines](https://www.youtube.com/t/community_guidelines)
72. <https://www.facebook.com/communitystandards>
73. <https://www.facebook.com/communitystandards>
74. [http://msdn.microsoft.com/en-us/library/windows/apps/hh184842\(v=vs.105\).aspx](http://msdn.microsoft.com/en-us/library/windows/apps/hh184842(v=vs.105).aspx)
75. <http://www.xbox.com/en-GB/legal/codeofconduct>
76. Robert Glancy, *"Will you read this article about terms and conditions? You really should do"*, The Guardian, 24 avril 2014. <http://www.theguardian.com/commentisfree/2014/apr/24/terms-and-conditions-online-small-print-information>
77. L'une de ces initiatives, baptisée *"Terms of Service. Didn't Read"*, évalue les conditions d'utilisation de certains des principaux services et applications Internet. Cette initiative est accessible à l'adresse suivante : <https://tosdr.org>.
78. Voir LICRA et UEJF c. Yahoo! Inc. et Yahoo Fr., T.G.I. Paris, 22 mai 2000 ; Voir également Ministerio Publico Federal c. Google Brasil Internet Ltda, 17a Vara Cível São Paulo, Proc. No. 2006.61.00.018332-8, 30 août 2006.
79. Guy Berger, *"Barbarians and gatekeepers"*, présentation dans le cadre du Colloque international sur la liberté d'expression, UNESCO, Paris, 26 janvier 2011. [http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/Events/International\\_Symposium\\_on\\_Freedom\\_of\\_Expression/presentations/guy\\_berger\\_symposium\\_foe\\_26\\_january.pdf](http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/Events/International_Symposium_on_Freedom_of_Expression/presentations/guy_berger_symposium_foe_26_january.pdf). Voir également Lwanga Mwilu, *"Framing the foreigner : a close reading of readers' comments on Thought leader blogs on xenophobia published between May and June"*, 2008. [http://contentpro.seals.ac.za/iii/cpro/DigitalItemViewPage\\_external?sp=1002927](http://contentpro.seals.ac.za/iii/cpro/DigitalItemViewPage_external?sp=1002927)

80. Adrian Chen, "Inside Facebook's outsourced anti-porn and gore brigade, where 'camel toes' are more offensive than 'crushed heads'", Gawker, 16 février 2012, <http://gawker.com/5885714/inside-facebooks-outsourced-anti-porn-and-gore-brigade-where-camel-toes-are-more-offensive-than-crushed-heads>.
81. Jeffrey Rosen, <http://www.newrepublic.com/article/113045/free-speech-Internet-silicon-valley-making-rules>
82. Entretien : Monika Bickert, Responsable mondiale de la politique de contrôle des contenus, Facebook, 14 janvier 2015.
83. Josh Constine, "Track Your Facebook Abuse, Bullying, Spam Reports With Transparent New Support Dashboard", TechCrunch, 26 avril 2012. <http://techcrunch.com/2012/04/26/facebook-support-dashboard>
84. Shiv Malik, Sandra Lavelle, Elena Cresci et Aisha Gani, "Isis in duel with Twitter and YouTube to spread extremist propaganda", *The Guardian*, 24 septembre 2014.
85. <http://www.theguardian.com/world/2014/sep/24/isis-twitter-youtube-message-social-media-jihadi>
86. "Google removes Singapore hate-speech blog targeting Filipinos", The Malay Mail Online, 20 juin 2014. <http://www.themalaymailonline.com/tech-gadgets/article/google-removes-singapore-hate-speech-blog-targeting-philipinos>
87. Doug Gross, "Twitter faces new pressure to limit hate speech", CNN, 31 juillet 2013. [www.cnn.com/2013/07/30/tech/social-media/twitter-hate-speech](http://www.cnn.com/2013/07/30/tech/social-media/twitter-hate-speech)
88. Kim Graham, "Petition: Add A Report Abuse Button To Tweets", Change.org, <http://www.change.org/p/twitter-add-a-report-abuse-button-to-tweets>
89. Humanitarian Policy Group, « Crisis in Kenya: land, displacement and the search for 'durable solutions' », 12 avril 2008.
90. Entretien : Nanjira Sambuli, chef de projet, UMATI, 26 novembre 2014.
91. Hereward Holland, « Facebook in Myanmar: Amplifying Hate Speech? », Al Jazeera, 14 juin 2014, <http://www.aljazeera.com/indepth/features/2014/06/facebook-myanmar-rohingya-amplifying-hate-speech-2014612112834290144.html>
92. « Why Is There Communal Violence in Myanmar? », BBC, 3 juillet 2014. <http://www.bbc.co.uk/news/world-asia-18395788>.
93. Michael Tan, « Myanmar's First Non-Government Phone Carrier, Ooredoo, Goes Live », CNET, 6 août 2014, <http://www.cnet.com/uk/news/myanmar-first-non-government-phone-carrier-ooredoo-goes-live/>.
94. « Why Is There Communal Violence in Myanmar? », BBC, 3 juillet 2014. <http://www.bbc.co.uk/news/world-asia-18395788>.
95. Andrew Marshall, « Special Report: Plight of Muslim Minority Threatens Myanmar Spring », Reuters, 15 juin 2012, <http://www.reuters.com/article/2012/06/15/us-myanmar-rohingya-idUSBRE85E06A20120615>.
96. Erika Kinetz, « New Numerology of Hate Grows in Burma », Irrawaddy, 29 avril 2013 <http://www.irrawaddy.org/religion/new-numerology-of-hate-grows-in-burma.html>. Hereward Holland, « Facebook in Myanmar: Amplifying Hate Speech? », Al Jazeera, 14 juin 2014 ; Steven Kiersons, « The Colonial Origins of Hate Speech in Burma », The Sentinel Project, 28 octobre 2013, <https://thesentinelproject.org/2013/10/28/the-colonial-origins-of-hate-speech-in-burma/>, <http://www.aljazeera.com/indepth/features/2014/06/facebook-myanmar-rohingya-amplifying-hate-speech-2014612112834290144.html>.

97. Tim McLaughlin, « Facebook takes steps to combat hate speech », The Myanmar Times, 25 juillet 2014. <http://www.mmtimes.com/index.php/national-news/11114-facebook-standards-marked-for-translation.html>
98. Ibrahim Sah, « Burma, US pledge to fight hate speech », The Burma Times, 4 octobre 2014. <http://burmatimes.net/burma-us-pledge-to-fight-hate-speech>.
99. San Yamin Aung, « Burmese Online Activist Discusses Campaign Against Hate Speech », Irrawaddy, <http://www.irrawaddy.org/interview/hate-speech-pours-poison-heart.html>.
100. Entretien : Myat Ko Ko, Justice Base, 24 novembre 2014.
101. Entretien : Harry Myo Lin, Panzagar, 12 décembre 2014.
102. Entretien : Christopher Wolf, ADL, 13 novembre 2014.
103. Entretien : Drew Boyd, Directeur des opérations, Sentinel Project for Genocide Prevention, 24 octobre 2014.
104. Entretien : Drew Boyd, Directeur des opérations, Sentinel Project for Genocide Prevention, 24 octobre 2014.
105. Entretien : Andre Oboler, PDG, Online Hate Prevention Institute, 31 octobre 2014.
106. Entretien : Imran Awan, Université de Birmingham, 21 novembre 2014.
107. Tell Mama, 2014. Pour plus d'informations, voir : <http://tellmamauk.org>
108. Facebook, 2013. "Controversial, harmful and hateful speech on Facebook". <https://www.facebook.com/notes/facebook-safety/controversial-harmful-and-hateful-speech-on-facebook/574430655911054>
109. BBC, 2012. « Facebook removes 'racist' page in Australia ». <http://www.bbc.co.uk/news/world-asia-19191595>
110. Toor, A., 2012. « Facebook faces hate speech criticism over 'Aboriginal memes' page ». <http://www.theverge.com/2012/8/8/3227329/facebook-hate-speech-aborigine-memes-australia>
111. Jacinta O'Keefe, 2012. « Petition: Immediately remove the racist page called 'Aboriginal Memes' » Change.org. <https://www.change.org/p/facebook-headquarters-immediately-remove-the-racist-page-called-aboriginal-memes>
112. Entretien : Andre Oboler, PDG, Online Hate Prevention Institute, 31 octobre 2014.
113. Pour plus d'informations sur le groupe Women, Action and the Media, consulter son site Internet officiel à l'adresse suivante : <http://www.womenactionmedia.org/why-wam/what-we-do/>
114. Pour plus d'informations sur le projet Everyday Sexism, consulter son site Internet officiel à l'adresse suivante : <http://www.everydaysexism.com/index.php/about>
115. Pour obtenir la liste de ces 15 entreprises, voir : WAM! (2013), *Campaign wins and updates: It worked!* Disponible en ligne à l'adresse suivante : <http://www.womenactionmedia.org/facebookaction/campaign-wins-updates/>
116. John Raines, 2013. « Petition: Demand Facebook remove pages that promote sexual violence ». Change.org. <https://www.change.org/p/demand-facebook-remove-pages-that-promote-sexual-violence>
117. Garossino, S., 2013. « How Facebook learned rape is bad for business ». [http://www.huffingtonpost.ca/sandy-garossino/wam-facebooks-pr-disaster\\_b\\_3357187.html](http://www.huffingtonpost.ca/sandy-garossino/wam-facebooks-pr-disaster_b_3357187.html)
118. Facebook, 2013. « Controversial, harmful and hateful, speech on Facebook », International Business Times, 7 novembre 2014, <https://www.facebook.com/notes/facebook-safety/controversial-harmful-and-hateful-speech-on-facebook/574430655911054>

119. Ibid.
120. Barbara Herman, « WAM! And Twitter tackle problem of online harassment of women ». <http://www.ibtimes.com/wam-twitter-tackle-problem-online-harassment-women-1720876>
121. Arce, N., 2014. « *Twitter, WAM! partnership now makes it possible to report gender-based harassment on Twitter* ». <http://www.techtimes.com/articles/19692/20141107/twitter-wam-partnership-now-makes-it-possible-to-report-gender-based-harassment-on-twitter.htm>
122. Kayla Epstein, « Twitter teams up with advocacy group to fight online harassment of women », The Guardian, 10 novembre 2014. <http://www.theguardian.com/technology/2014/nov/08/twitter-harassment-women-wam>
123. WAM!, « To combat the harassment of women online, Women, Action and the Media (WAM!) announce a new partnership with Twitter ». <http://www.womenactionmedia.org/cms/assets/uploads/2014/11/Twitterprojectpressrelease-1.pdf>
124. WAM!, « WAM Twitter harassment reporting tool ». <https://womenactionmedia.wufoo.com/forms/ztaetj1jrhv10/>
125. Entretien : Monika Bickert et Ciara Lyden, Politique de contrôle des contenus, Facebook, 14 janvier 2015.
126. Twitter, *We hear you*. <https://blog.twitter.com/en-gb/2013/we-hear-you>
127. Kim Graham, « Petition: Add A Report Abuse Button To Tweets », Change.org, <http://www.change.org/p/twitter-add-a-report-abuse-button-to-tweets>
128. Google, « Transparence des informations ». <http://www.google.com/transparencyreport/>
129. <http://www.unesco.org/new/fr/global-citizenship-education>
130. La Déclaration de Paris appelle à un intérêt renouvelé pour l'éducation aux médias et à l'information (MIL) à l'ère numérique. <http://www.unesco.org/new/fr/communication-and-information/resources/news-and-in-focus-articles/in-focus-articles/2014/paris-declaration-on-media-and-information-literacy-adopted/>
131. Le Mouvement contre le discours de haine est une campagne régionale regroupant 50 pays, et ce bien au-delà du continent européen. Bien que la campagne ait des buts communs et mette au point des stratégies conjointes, les projets et initiatives spécifiques menés dans chaque pays sont placés relèvent de coordonnateurs nationaux et dépendent des capacités et des ressources de chaque pays.
132. Entretien : Laura Geraghty, Mouvement contre le discours de haine, 25 novembre 2014.
133. HabiloMédias, « Young Canadians in a wired world. Teachers' perspectives », 2012.
134. Entretien : Matthew Johnson, Directeur de l'éducation, HabiloMédias, 15 novembre 2014.
135. Mouvement contre le discours de haine, outils et supports de campagne. Disponibles en ligne à l'adresse suivante : <http://nohate.ext.coe.int/Campaign-Tools-and-Materials>
136. Mouvement contre le discours de haine, No Hate Ninja Project - A Story About Cats, Unicorns and Hate Speech. Disponible en ligne à l'adresse suivante : <https://www.youtube.com/watch?v=kp7ww3KvccE>
137. HabiloMédias, Faire face à la haine sur Internet. Disponible en ligne à l'adresse suivante : <http://habilomedias.ca/tutoriel/faire-face-haine-sur-internet>
138. Projet *In Other Words*, Toolbox. Disponible en ligne à l'adresse suivante : <http://www.inotherwords-project.eu/sites/default/files/Toolbox.pdf>
139. Aileen Donegan, « Debate 2 : "Hate speech is more than free speech" », forum de discussion du Mouvement contre le discours de haine, 17 octobre 2013, <http://forum.nohatespeechmovement.org/discussion/6/debate-2-hate-speech-is-more-than-free-speech/p1>.

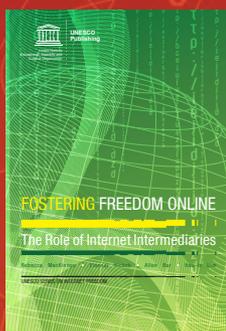
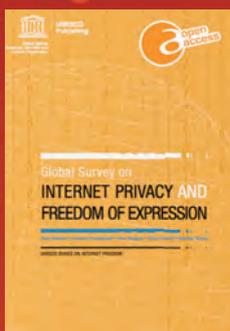
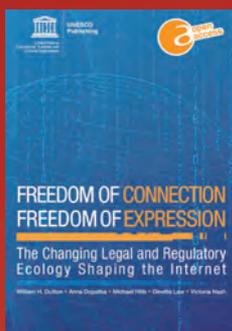
140. Mouvement contre le discours de haine, cinquième réunion du groupe de suivi. Disponible en ligne à l'adresse suivante : <http://nohate.ext.coe.int/The-Campaign/Follow-Up-Group-of-the-Joint-Council-on-Youth2>
141. <https://en-gb.facebook.com/help/128548343894719>
142. Article 19. Les intermédiaires Internet : Dilemme de la responsabilité, question-réponse. <http://www.article19.org/resources.php/resource/37243/fr/internet-intermediaries:-dilemma-of-liability-q-and-a>
143. Voir par exemple l'étude de Stormfront par De Koster et Houtman (2008) mentionnée au chapitre 5.
144. Voir par exemple l'article de Rowbottom (2012) et le « Jane Austin Row » mentionné au chapitre 4.
145. Voir Susan Benesch, 2014 : Troll Wrastling for Beginners: Data-Driven Methods to Decrease Hatred Online, <http://cyber.law.harvard.edu/events/luncheon/2014/03/benesch> ; et Ethan Zuckerman, 2014 : Susan Benesch on dangerous speech and counterspeech, <http://www.ethanzuckerman.com/blog/2014/03/25/susan-benesch-on-dangerous-speech-and-counterspeech/>

Les possibilités qu'offre Internet éclipsent largement les difficultés qui en résultent. Tout en gardant cela à l'esprit, nous devons néanmoins nous attaquer à certains de ces problèmes, et notamment celui des discours de haine en ligne. Mais en quoi consistent exactement les propos haineux sur Internet et comment les combattre efficacement ?

En matière de liberté d'expression, sur Internet ou hors Internet, l'UNESCO préconise que la libre circulation de l'information soit toujours la norme. Le contre-discours est généralement préférable à la répression de la liberté de parole. Et toute mesure de restriction de la liberté d'expression doit être soigneusement considérée afin de garantir que ce type d'action demeure tout à fait exceptionnel et que le débat légitime et approfondi ne soit pas entravé.

Getachew Engida  
Sous-Directeur général de l'UNESCO

## COLLECTION UNESCO SUR LA LIBERTÉ DE L'INTERNET



Organisation  
des Nations Unies  
pour l'éducation,  
la science et la culture

Secteur  
de la communication  
et de l'information



9 789232 000712