

Discours de haine sur l'internet

Angeliki Monnier et Annabelle Seoane

Référence électronique

Angeliki Monnier et Annabelle Seoane, Discours de haine sur l'internet. *Publictionnaire. Dictionnaire encyclopédique et critique des publics*. Mis en ligne le 12 juin 2019. Accès : <https://publictionnaire.huma-num.fr/notice/discours-de-haine-sur-linternet/>

Le Publictionnaire. Dictionnaire encyclopédique et critique des publics est un dictionnaire collaboratif en ligne sous la responsabilité du Centre de recherche sur les médiations (Crem, Université de Lorraine) ayant pour ambition de clarifier la terminologie et le profit heuristique des concepts relatifs à la notion de public et aux méthodes d'analyse des publics pour en proposer une cartographie critique et encyclopédique.

Accès : <https://publictionnaire.huma-num.fr>

Cette notice est mise à disposition selon les termes de la licence Creative Commons Attribution - Pas d'utilisation commerciale - Pas de modification 3.0 France. Pour voir une copie de cette licence, visitez <http://creativecommons.org/licenses/by-ncnd/3.0/fr/> ou écrivez à Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.



Discours de haine sur l'internet

L'expression *discours de haine*, le plus souvent utilisée dans son équivalent anglais de *hate speech*, s'est propagée depuis les années 1980. Elle est utilisée dans un ensemble d'études émanant d'institutions et associations, européennes et extra-européennes, visant à sensibiliser les publics contre l'expansion du phénomène des discours de haine au sein des sociétés contemporaines, notamment sur l'internet et les plateformes de réseaux socio-numériques (e.g. Bricks, 2016 ; Selma, 2019). Cette expression est également l'objet d'un nombre croissant de travaux en sciences humaines et sociales, dans le monde anglo-saxon (e.g. Siapera, Viejo-Otero, Moreo, 2017), mais aussi en France (e.g. Glucksmann, 2004 ; L'Heuillet, 2017). L'expansion de ces derniers s'appréhende notamment dans le sillage du phénomène que les chercheurs ont désigné comme le « tournant affectif » des années 1990, qui porte un regard renouvelé sur la manière dont l'affect façonne les expériences et les pratiques humaines (Slaby, Von Scheve, 2019). C'est dans le cadre de ce contexte épistémologique et disciplinaire que le discours de haine nous intéresse ici.

La haine à l'ère des publics affectifs de l'internet

L'internet s'est avéré un terrain particulièrement fertile pour l'émergence et l'expansion de discours haineux. Le psychiatre et psychanalyste Serge Tisseron (2011) pointe le rôle des dispositifs : l'immédiateté faciliterait la parole affective, l'anonymat réduirait les inhibitions ; le tout accentué par le « libertarianisme informationnel » (*informational libertarianism*) et l'« expansionnalité » (*spreadability*) des contenus, propres à l'internet (Siapera, Moreo, Zhou, 2018). Les dispositifs « à forte dominante captatrice » que seraient les réseaux socio-numériques renforceraient une mise en scène discursive « à visée pathémique » de la réalité (Charaudeau, 2000), c'est-à-dire favorisant le *pathos* dans la description et l'appréhension du monde. Le web participatif (Alloing, Pierre, 2017 ; Courbet, Fourquet-Courbet, Marchioli, 2015) engendrerait l'émergence de « publics affectifs » (Papacharissi, 2015) ou de « sociétés affectives » (Von Scheve, Slaby, 2019). François Jost (2018), chercheur en sciences de l'information et de la communication, y voit d'ailleurs surtout une nouvelle forme de la lutte des classes, visant les élites politiques, médiatiques, etc., qui sont souvent cibles de propos haineux.

Dans son usage courant, *hate speech* fait alors référence à un ensemble de manifestations affectives assez hétérogènes, allant de propos méprisants et/ou hostiles à l'égard de groupes et d'individus à des incitations à des actes criminels. En anglais, la construction syntaxique de l'expression semble intégrer trois facettes qui, en revanche, sont distinguables en français : celle du « discours de haine », issu d'un sentiment (préssumé) de haine ; du « discours haineux », comme porteur de haine ; d'un discours « provoquant la haine », en tant qu'incitation directe à la violence vis-à-vis de l'autre ou la consolidation de stigmates (Monnier, Seoane, Gardenier, 2019). Du reste, la définition du politiste Raphael Cohen-Almagor (2011 : 1-2) illustre bien cette acception « large » et tridimensionnelle du terme :

« Le discours de haine est défini comme un discours malveillant, motivé par des préjugés, visant une personne ou un groupe pour leurs caractéristiques innées réelles

ou perçues. Il exprime des attitudes discriminatoires, intimidantes, désapprobatrices, antagonistes et/ou préjudiciables à l'égard de ces caractéristiques, notamment le sexe, la race, la religion, l'appartenance ethnique, la couleur, l'origine nationale, un handicap ou l'orientation sexuelle. Le discours de haine a pour but de blesser, déshumaniser, harceler, intimider, affaiblir, dégrader et victimiser les groupes ciblés, et de fomenter l'insensibilité et la brutalité vis-à-vis de ces derniers. »

Haine ou liberté d'expression ?

À l'heure actuelle, le discours de haine n'a pas de définition précise en termes de droits humains internationaux et les politiques nationales divergent, reflétant des traditions propres à chaque pays. Exemple souvent cité, les États-Unis appliquent face aux propos haineux une attitude plus « tolérante » par rapport à celle des pays européens, protégeant davantage le principe de la liberté de parole (*freedom of speech*), inscrite dans le premier amendement de la Constitution étatsunienne (Siapera, Moreo, Zhou, 2018 ; Walker, 1994). Néanmoins, du côté européen, selon le Comité des ministres du Conseil de l'Europe (2015), le discours de haine couvre toute forme d'expression qui justifie ou répand la haine raciale, la xénophobie, l'antisémitisme ou toute forme de haine fondée sur l'intolérance, y incite ou en fait l'apologie. Certains groupes sont explicitement « protégés » par la loi en raison de leurs traits « identitaires » spécifiques : ethniques, religieux, sexuels, etc.

Modérer les discours de haine en ligne ?

Le foisonnement de ce phénomène a accentué la nécessité d'une régulation des affects en ligne ; mais certains contestent la légitimité d'une telle démarche, évoquant les aspects « cathartiques » de l'expression libre des affects (Kappelhoff, Lehmann, 2019). Bien que parfois contestées, les pratiques de modération des réseaux socio-numériques se confrontent aux difficultés d'évaluer systématiquement la portée et le sens des pratiques langagières de leurs usagers (Siapera, Viejo-Otero, Moreo, 2017 ; Gillespie, 2018). En effet, ces difficultés sont de plusieurs ordres : difficulté pour catégoriser les contenus considérables comme « haineux », difficulté technique pour repérer ces contenus, difficulté éthique ou juridique pour les cadrer. C'est pourquoi il est indispensable d'interroger les contextes juridique, historique, socio-culturel, etc., qui définissent le sens du discours de haine et les modalités de sa sanction.

Dans ce contexte, les initiatives se multiplient pour mettre en place un protocole d'identification de ces contenus en ligne afin de lutter contre le discours de haine. Des chercheurs en informatique ou en linguistique informatique se penchent ainsi sur la mise en place d'outils de détection automatique (pour une synthèse voir Fortuna, Nunes, 2018) par le biais de travaux dans lesquels prévaut une approche par le lexique : l'analyse et la lutte contre le discours de haine sur l'internet passent souvent par le recours à des listes préétablies de mots « haineux », l'accent étant mis sur le sens de ces mots, leur fréquence, et les configurations syntaxiques au sein desquelles ils apparaissent. Cependant, l'absence d'une « norme de référence » (« *gold standard* ») dans la définition du discours haineux rend son repérage difficile. Où commence la « haine » ? Est-ce qu'une trace langagière suffit pour désigner une supposée émotion ? La démarche s'avère davantage compliquée, d'autant plus que le contenu en ligne à analyser est vaste et disparate : il intègre les formes de « haine » les

plus subtiles, implicites (sarcasmes, euphémismes, stéréotypes, allusions à des connaissances partagées, etc.), les « stratégies de masquage » mises en place par les internautes (jeux avec l'orthographe, crypto-langages réservés aux initiés, etc.). Le discours haineux répond donc à des pratiques communicationnelles indexées sur deux paramètres énonciatifs : d'abord le locuteur et son destinataire (qui s'adresse à qui ?), ensuite le contexte situationnel, socioculturel et historique des échanges (sur quels forums ? après quels événements ? sur la base de quelles valeurs ?).

De la parole haineuse au phénomène social de la haine

Cette variété d'expressions de contenus haineux interroge l'activité langagière qu'elle soutient : s'agit-il d'une construction ininterrompue, orientée idéologiquement ou bien d'expressions d'affects toutes personnelles ? Ceci incite à questionner les mots eux-mêmes : si le terme *speech*, au singulier, tel qu'il apparaît dans l'expression, pose une unicité référentielle, *discours* quant à lui ne marque pas dans sa forme la différence entre singulier et pluriel. Il y a superposition de deux sens dans la même forme : au singulier, le terme réfère à une construction continue, relativement homogène, et polyphonique ; au pluriel, il désigne des réalisations langagières éparses et potentiellement attribuables à des sources énonciatives diverses. Pour cette raison, une approche multidimensionnelle semble la plus apte à mettre en évidence les processus interactionnels, connectifs et pluriels au sein desquels émergent les contenus haineux. Il s'agit bien sûr de se pencher sur les messages dont l'intensité s'avère scalaire, allant de l'offense à l'agression verbale la plus violente, celle qui incite à la mort. Il est également nécessaire de questionner le rôle des dispositifs qui favorisent un certain type de contenus (algorithmes, phrases courtes, discours provocateurs, etc., voir Gomez-Mejia, 2017). Il faut aussi s'arrêter sur les profils des internautes (genre, localisation géographique) et leurs usages (recours à l'anonymat, etc.), ainsi que sur les logiques de mise en scène et de visibilité qui sous-tendent leurs pratiques (authenticité de soi, appartenance, etc., Proulx *et al.*, 2011). Émerge alors l'image d'un phénomène social de « haine » en ligne (Monnier, Seoane, 2019), au sein d'un univers d'hostilité ambiante (Siapera, 2019) qui banalise le conflit et l'expression des états d'âme. En ce sens, le discours de haine a une fonction performative (Lünenborg, 2019) car il contribue, à son tour aussi, à la construction symbolique – ou du moins à la consolidation – des publics affectifs.

Angeliki Monnier et Annabelle Seoane bénéficient d'une aide de l'État, gérée par l'Agence nationale de la recherche, au titre du projet M-Phasis (*Migration and Patterns of Hate Speech in Social Media*), portant la référence ANR-18-FRAL-0005.

Bibliographie

Alloing C., Pierre J., 2017, *Le Web affectif. Une économie numérique des émotions*, Paris, INA Éd.

Bricks (Building Respect on the Internet by Combating hate Speech), 2016, #Silence Hate, University of Milano Bicocca. Accès: https://www.bricks-project.eu/wp/wp-content/uploads/2016/10/relazione_bricks_eng2-1.pdf.

Charaudeau P., 2000, « La pathémisation à la télévision comme stratégie d'authenticité », pp. 125-155, in : Plantin C., Doury M., Traverso V., dirs, *Les Émotions dans les interactions*, Lyon, Presses universitaires de Lyon.

Cohen-Almagor R., 2011, « Fighting Hate and Bigotry on the Internet », *Policy and Internet*, 3 (3), pp. 1-26.

Conseil de l'Europe, 2015, *Discours de haine*, Fiche thématique. Accès : https://www.echr.coe.int/Documents/FS_Hate_speech_FRA.pdf.

Courbet D., Fourquet-Courbet M.-P., Marchioli A., 2015, « Les médias sociaux, régulateurs d'émotions collectives », *Hermès. La Revue*, 1 (71), pp. 287-292. Accès : <https://www.cairn.info/revue-hermes-la-revue-2015-1-page-287.htm>.

Fortuna P., Nunes S., 2018, « A Survey on Automatic Detection of Hate Speech in Text », *ACM Computing Surveys*, 51 (4), pp. 1-30.

Gillespie T., 2018, *Custodians of the Internet. Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*, New Haven, Yale University Press.

Glucksmann A., 2004, *Le Discours de la haine*, Paris, Plon.

Gomez-Mejia G., 2017, « Troll », *Publictionnaire. Dictionnaire encyclopédique et critique des publics*. Accès : <http://publictionnaire.huma-num.fr/notice/troll/>.

Jost F., 2018, *La Méchanceté en actes à l'ère numérique*, Paris, CNRS Éd.

Journal officiel de la République française, 1972, *Loi n° 72-546 du 1^{er} juillet 1972 relative à la lutte contre le racisme*. Accès : https://www.legifrance.gouv.fr/jo_pdf.do?id=JORFTEXT000000864827&pageCourante=06803.

Kappelhoff H., Lehmann H., 2019, « The Temporal Composition of Affects in Audiovisual Media », pp. 130-139, in : Kahl A., ed., *Analyzing Affective Societies. Methods and Methodologies*, Berlin, Routledge.

L'Heuillet H., 2017, *Tu haïras ton prochain comme toi-même. Les tentations radicales de la jeunesse*, Paris, A. Michel.

Lünenborg M., 2019, « Affective Publics », pp. 319-329, in : Slaby J., Von Scheve C., eds, *Affective Societies. Key Concepts*, London/New York, Routledge.

Monnier A., Seoane A., 2019, « Suspendre un énoncé, sous-tendre un jugement : les points de suspension dans des discours haineux sur Internet », Colloque *Faire silence : matérialités, expériences, pouvoirs*, EHESS Marseille, 22 mai.

Monnier A., Seoane A., Gardenier M., 2019, « Réflexions méthodologiques sur le discours haineux anti-migrants », Journées d'études franco-italiennes *Médias et émotions. Catégories d'analyse, problématiques, concepts*, Université de Bordeaux, 11 avr.

Papacharissi Z., 2015, *Affective Publics. Sentiment, Technology, and Politics*, New York, Oxford University Press.

Proulx S. et al., 2011, « Paradoxical Empowerment of Producers in the Context of Informational Capitalism », *New Review of Hypermedia and Multimedia*, 17 (1), pp. 9-29.

Selma, 2019, *Hacking Online Hate: Building an Evidence Base for Educators*, European SchoolNet. Accès : www.hackinghate.eu.

Siapera E., 2019, « Organised and Ambient Digital Racism: Multidirectional Flows in the Irish Digital Sphere », *Open Library of Humanities*, 5 (1). Accès : <https://olh.openlibhums.org/articles/10.16995/olh.405/>.

Siapera E., Moreo E., Zhou J., 2018, *HateTrack. Tracking and Monitoring Racist Speech Online, Project Final Report*, Irish Research Council, DCU School of Communications, Institute for Future Media and Journalism (FuJo), Anti-Bullying, Centre (ABC), and Insight Centre for Computer Analytics, School of Computing. Accès : <https://www.ihrec.ie/app/uploads/2018/11/HateTrack-Tracking-and-Monitoring-Racist-Hate-Speech-Online.pdf>.

Siapera E., Viejo-Otero P., Moreo E., 2017, « Hate Speech. Genealogies, Tensions and Contentions », *Proceedings of the Association of Internet Researchers*, pp. 19-21, Estonia, University of Tartu.

Slaby J., Von Scheve C., eds, 2019, *Affective Societies. Key Concepts*, London/New York, Routledge.

Tisseron S., 2011, « Les nouveaux réseaux sociaux : visibilité et invisibilité sur le net », pp. 117-129, in : Aubert N., Haroche C., dirs, *Les Tyrannies de la visibilité. Être visible pour exister ?*, Toulouse, Éd. Érès.

Von Scheve C., Slaby J., 2019, « Emotion, Emotion Concept », pp. 42-51, in : Slaby J., Von Scheve C., eds, *Affective Societies. Key Concepts*, London/New York, Routledge.

Walker S., 1994, *Hate Speech. The History of an American Controversy*, Lincoln, University of Nebraska Press.