Mediascapes journal 16/2020

I discorsi d'odio online in una prospettiva comunicativa: ì un'agenda per la ricerca*

Sara Bentivegna
Sapienza Università di Roma**

Rossella Rega
Università degli Studi di Siena***

The phenomenon of hate speech has become a predominant component of public debate and a point of great interest in academic research. Plenty of studies investigating the various expressions of hate speech, including online hate speech, have been conducted. Despite the efforts made, however, the outlines of this phenomenon remain extremely uncertain. In fact, limitations in the intention to reduce hate speech to a purely legal approach is emerging. Even though the latter was originally the dominant approach, it appears to be inadequate for capturing a fluid and ever-changing phenomenon such as the hate speech today. Despite the similarities with offline hatred, online hatred has widened to include the most diverse expressions, messages, and hostility practices, which can be either organized or individual, and often implicit. All these features proper of the online hate speech make it difficult to establish, according to the specific circumstances, whether these are forms openly aimed at inciting hatred and violence against target groups likely to harm victims. Nonetheless, distinguishing the different roles of the actors involved into the spread of hate speech (e.g. perpetrators of online speech, more or less aware spreaders of hate, victims) can be a source of uncertainty. Building on this foundation, the article aims to investigate hate speech of today's world through a multidimensional framework fit into the context of hybrid media. Moving from Lasswell (1948)'s Communication Model, this contribution illustrates the importance of observing all the dimensions involved in hate speech, including the communicator, the message, the receiver, considering as well the digital media's context and the related specific properties, and the consequences arising from the interaction between the different dimensions. The idea behind this proposal is that in order to understand both mechanisms and hallmarks of ordinary online hate practices, it is necessary to move away from specific and fragmented approaches. These, indeed, (selectively) analyse specific portions of the phenomenon while losing sight of other dimensions' roles and their action and feedback relations. The transformative nature of hate speech requires the adoption of an innovative approach, which cannot be limited to the compilation of a list of examples and cases to be banned. On the contrary, it should aim for tuning in to cultural and digital practices used by users in the production of hate speech. Because, if it is true that online hate does not differ from hate tout court, it is also true that progressive communicative shifts might occur and lead to the transformation of forms of discrimination into something apparently harmless such as "just funny stuff".

Keywords: Hate Speech, discorsi d'odio online, social media, web affordance, pratiche comunicative.

* Articolo proposto il 28/05/2020. Articolo accettato il 20/09/2020
** sara.bentivegna@uniroma1.it
*** rossella.rega@gmail.com

L’hate speech, da straordinario a ordinario

Sono trascorsi non più di quattro anni da quando Joanne K. Rowling commentò l’elezione di Donald Trump alla carica di presidente degli Stati Uniti con la pubblicazione di un tweet del seguente tenore: «We don’t let hate speech become normalised» (9-11-2016; https://twitter.com/jk_rowling/status/796252371739430913). Con quel grido di allarme, la scrittrice inglese voleva richiamare l’attenzione sulla diffusione dei discorsi d’odio a ogni livello delle società contemporanee, addirittura a quello della presidenza degli Stati Uniti.

Non è stato certo un monito isolato quello dell’inventrice di Harry Potter. La sensazione che l’ostilità e i discorsi d’odio si siano normalizzati, cioè che siano entrati a pieno titolo nel dibattito pubblico e che si rintraccino in numerose e quotidiane esperienze comunicative, circola ormai tanto sulle pagine dei quotidiani quanto negli articoli di settore sulle trasformazioni delle pratiche di comunicazione definite ora “incivili”, ora “aggressive”, ora “violente”. Il timore che soprattutto gli spazi online possano trasformarsi in ambienti inospitali e ostili che impediscono la libera espressione e il confronto pubblico, al punto da costituire un vero pericolo per il funzionamento delle democrazie contemporanee, è diventato il *leitmotiv* che segna molte riflessioni sulle caratteristiche dell’attuale ecosistema mediale. Si tratta di un rischio che viene segnalato da più parti al punto da essere rappresentato come “una minaccia ambientale alla pace sociale, una sorta di veleno ad azione lenta, che si accumula qua e là, parola per parola, così che alla fine diventa più difficile e meno naturale anche per i membri *di buon cuore* della società fare la loro parte nel mantenere questo bene pubblico" (Waldron 2012, p. 4).

Un cambiamento significativo, se pensiamo che fino a qualche tempo fa i media digitali erano per lo più intesi come strumenti di liberazione che, soprattutto nell’ambito della comunicazione politica, avrebbero dato spazio e voce anche a soggetti più marginali, contribuendo ad arricchire il flusso di informazioni e scambi solitamente controllato da giornalisti e rappresentanti politici. Nel giro di qualche anno, però, si è visto che la diffusione di massa di piattaforme digitali e disintermediate se da un lato ha reso più orizzontale e aperto il dibattito pubblico, dall’altro ha fatto emergere in modo imprevisto diversi elementi di criticità. Non è dunque un caso se, contestualmente al dilagare del clima di odio generalizzato e degli episodi dilaganti di intolleranza, razzismo e antisemitismo, si stiano anche moltiplicando gli studi e le ricerche che cercano di cogliere le radici e i contorni di un fenomeno tutt’altro che privo di ambiguità. Collocandosi in un complesso intreccio tra libertà di espressione, diritti individuali, diritti di gruppo e delle minoranze, e concetti relativi alla dignità, libertà e uguaglianza (Gagliardone 2019), il fenomeno dell’incitamento all’odio non è riconducibile, infatti, a un approccio esclusivamente di ordine giuridico che, sebbene inizialmente sia stato quello dominante, con il tempo ha mostrato chiaramente tutti i suoi limiti. In particolare, ha mostrato e continua a mostrare la sua inadeguatezza a cogliere un fenomeno fluido e in continua evoluzione. Ciò è emerso con particolare evidenza in concomitanza con la diffusione di massa dei social media, quando si sono sviluppate proposte d’interpretazione volte a

superare il frame esclusivamente giuridico-normativo e a interrogarsi sulle implicazioni più ampie sottostanti alle pratiche di produzione, circolazione e consumo dei messaggi d'odio.

Il tema dell'odio online, infatti, sebbene nella sostanza non si discosti dall'odio offline, ridefinisce il proprio spazio fino a ricomprendere espressioni, messaggi e pratiche di ostilità le più diverse, spesso indirette o implicite, rispetto alle quali risulta molto complesso stabilire di volta in volta se si tratti di forme indirizzate esplicitamente a incitare odio e violenza contro gruppi mirati e suscettibili di danneggiare le vittime. Nel contesto più magmatico della comunicazione online appare chiaro che, oltre alla mutevolezza delle forme di hate speech (da adesso HS[1]), occorra anche tenere conto del numero più ampio degli attori implicati nei processi di diffusione dei messaggi d'odio, dell'incertezza e talvolta indistinzione dei ruoli (produttori di HS, diffusori più o meno consapevoli, vittime) e dell'istantaneità delle dinamiche di propagazione e di contagio. Tutti elementi che contribuiscono nell'insieme a rendere più complessa la possibilità di caratterizzare e descrivere i discorsi d'odio in modo certo e univoco. Senza dimenticare, poi, la complessità derivante dal fatto che l'HS può essere considerato, a nostro parere, un "moving concept" che si definisce e ridefinisce in relazione al contesto culturale, politico, sociale e comunicativo nel quale si colloca.

Partendo da queste premesse e volendo evitare l'ennesimo tentativo di definizione dell'hate speech, questo contributo si propone di offrire un framework multidimensionale attraverso cui esaminare i discorsi d'odio online e orientarne lo studio e l'interpretazione. Con l'obiettivo di uscire dal perimetro delle condanne morali e aprire una riflessione sul senso delle pratiche d'odio ordinarie che affollano il web e sulle motivazioni alla base della loro diffusione, ci interessa in questa sede offrire uno strumento attraverso il quale guidare l'analisi dell'HS, provando a comprenderne i meccanismi e i tratti distintivi. Perché, ad esempio, i linguaggi d'odio, l'ostilità e la contrapposizione sono oggi così presenti e visibili negli ambienti online? Quali caratteristiche condividono i discorsi d'odio che innescano meccanismi di contagio orizzontale e producono esperienze di rabbia e ostilità condivise? In quali contesti o situazioni possono diventare più corrosivi e pericolosi? Perché piattaforme come Twitter e Facebook dovrebbero porre più attenzione ai messaggi problematici provenienti da fonti politico-istituzionali rispetto a quelli diffusi da altri soggetti?

Questi sono soltanto alcuni degli interrogativi che vorremmo porre al centro della riflessione e ai quali riteniamo si possa utilmente rispondere partendo da un principio base per chi studia sociologia della comunicazione, ovvero che il potere di influenza di una *comunicazione* dipende soltanto in parte dalle caratteristiche del messaggio, il quale acquisisce un significato decisivo in determinate *circostanze* e *contesti.* Sulla base di questo presupposto, proponiamo di considerare l'online HS nei termini di un processo comunicativo di natura relazionale per comprendere il quale occorre tenere conto delle diverse dimensioni implicate nel processo: le caratteristiche del comunicatore, le specificità dell'ambiente digitale nel quale si muove, il messaggio nei suoi aspetti sia formali che sostanziali, i destinatari di questo messaggio intesi nella doppia veste di pubblici e performer e, infine, gli effetti provocati dal messaggio. Dimensioni tutte parimente coinvolte e in un rapporto costante di azione e retroazione, ma che considerate

in maniera disgiunta o disorganica, difficilmente possono restituire il senso complessivo di un atto di comunicazione così come di un discorso d'odio. Si tratta, in altre parole, di esaminare l'HS attraverso il modello della comunicazione adottato inizialmente da Lasswell (1948), che, concepito secondo una prospettiva più dinamica e meno rigida rispetto alla versione originaria, può delinearsi come uno strumento euristico ancora utile a guidare lo studio di fenomeni poliedrici e mutevoli come i discorsi d'odio contemporanei. Sembrerà a prima vista una mossa eccessivamente semplice per dimostrarsi di una qualche utilità, ma se non si considerano i diversi elementi implicati nel processo comunicativo e le loro reciproche interazioni, difficilmente si potranno spiegare le caratteristiche di un fenomeno ormai normalizzato o comprendere che in alcune circostanze i discorsi d'odio richiedono più attenzione perché possono contribuire a legittimare modelli comportamentali fondati sulla cultura della discriminazione e della violenza.

Nelle pagine seguenti ci si muoverà in questa direzione, prendendo innanzitutto in esame alcune delle principali definizioni e concetti relativi alla letteratura consolidata sull'HS per evidenziarne la loro parziale adeguatezza a spiegare le caratteristiche dei discorsi d'odio contemporanei; quindi si affronterà il discorso sulle *affordance* e le specificità degli ambienti digitali in relazione all'HS e si riadatterà il modello multidimensionale di Lasswell al fine di restituire una visione d'insieme del fenomeno che tenga conto del contributo delle diverse dimensioni coinvolte e delle dinamiche di interazione e feedback tra di esse.

Studiare l'hate speech: ieri e oggi

La letteratura consolidata sull'HS, nata ormai più di cinquanta anni fa, può essere ancora un utile punto di partenza per studiare come si sviluppano oggi i discorsi d'odio tra spazi tecnologicamente mediati e non? La risposta a questo interrogativo non è affatto semplice, soprattutto in considerazione delle profonde trasformazioni che hanno segnato il sistema mediale e della centralità acquisita oggi dalla dimensione comunicativa.

Ciò premesso, per entrare nel merito della questione occorre partire dai significati originariamente attribuiti all'hate speech dagli studiosi di diritto che hanno introdotto questa espressione nel dibattito Usa a cavallo tra gli anni '70 e '80, in riferimento ad alcuni tipi di discorsi razzisti più corrosivi. Già il contesto nel quale è nata l'espressione e la focalizzazione sul razzismo ci fanno immediatamente capire come la questione sia strettamente legata a quella del difficile equilibrio tra la libertà di espressione, tutelata negli Stati Uniti dal primo emendamento, e il rispetto dei diritti umani.

Tra i primi a occuparsi del tema figura Delgado (1982), che individua tre elementi in base ai quali un discorso può essere qualificabile come HS e, dunque, sanzionabile dalla legge: (i) l'*intenzione* esplicita dell'oratore di sminuire la vittima attraverso il riferimento alla razza; (ii) l'*impatto* del discorso sulla vittima, ossia che quest'ultima interpreti il discorso per come è stato inteso (denigratorio), e (iii) la *percezione oggettiva* che si tratti di un

insulto razziale, vale a dire che una "persona ragionevole" lo possa riconoscere come tale. Se è immediatamente chiaro da questa definizione che non è presente un'attenzione alla dimensione comunicativa dell'HS, toccata soltanto indirettamente dall'intenzione dell'oratore di degradare la vittima, leggendo questi requisiti viene da chiedersi se siano ancora utili per affrontare lo studio dei discorsi d'odio oggi. È davvero così dirimente sapere che da parte del comunicatore vi sia un'esplicita volontà di danneggiare il target o che il messaggio sia interpretato secondo questa lettura dalla vittima?

Lasciando al momento da parte queste domande, tra i padri fondatori degli hate studies va certamente menzionata Matsuda (1989), che adotta una definizione di HS focalizzata molto sul contenuto del discorso che, per poter rientrare in questa categoria, deve contenere una discriminazione di tipo razziale (inferiorità razziale), essere persecutorio, odioso e degradante, prendere di mira gruppi o membri di gruppi tradizionalmente oppressi, e muovere da una evidente volontà del comunicatore di danneggiare il target.

Ora, se è abbastanza intuitivo che il contenuto è una dimensione rilevante anche in un'ottica comunicativa e molti lavori contemporanei continuano infatti a fondarsi sull'analisi del contenuto dei messaggi, due osservazioni meritano la nostra attenzione. Innanzitutto un approccio meramente ancorato sul contenuto poteva funzionare in un contesto come quello di fine '900, in cui il discorso pubblico sull'HS riguardava attori e contesti comunicativi circoscritti e le sue ricadute si legavano soltanto ai soggetti che potevano essere toccati da quel messaggio. Il processo di normalizzazione dell'HS nel discorso pubblico contemporaneo evidenzia, invece, che non è tanto o solo il contenuto dei messaggi d'odio postati o diffusi online a fare la differenza, quanto invece il modo in cui essi transitano tra i diversi media e piattaforme, come vengono trasformati e ri-distribuiti da innumerevoli soggetti tra loro diversi (cittadini, attori politici, media, etc.), gli effetti a catena che possono innescare, impossibili da delimitare nel tempo, nello spazio come pure nel numero dei soggetti su cui potrà produrre delle ricadute. In secondo luogo, nonostante il contenuto del messaggio sia la dimensione d'indagine che si è mantenuta più stabile nel corso degli anni, le sue caratteristiche sono però mutate. Matsuda fa riferimento ai contenuti riguardanti la discriminazione razziale nei confronti di gruppi o classi di individui "storicamente oppressi" (1989), ma, anche ampliando lo sguardo ad altri autori, i bersagli dell'odio hanno sempre connotati molto specifici; Moran parla, ad esempio, di target tradizionalmente svantaggiati (2016), Marwick e Miller di gruppi emarginati (2014), mentre altri autori preferiscono elencare le caratteristiche in base alle quali identificare i gruppi bersaglio che includono il genere, la razza, la religione, l'etnia, il colore, l'origine nazionale, la disabilità o l'orientamento sessuale (Cohen-Almagor 2011). Attualmente, però, le forme di rancore e di odio sono più mutevoli per motivazioni, forme e significati, e accanto ai target tradizionali, ve ne sono di nuovi e del tutto contingenti, quasi impossibili da prevedere. Così, durante il periodo di lockdown dovuto all'emergenza del Covid-19 si è visto rientrare tra i bersagli d'odio di molti italiani categorie decisamente insolite come runners, proprietari di cani domestici, cittadini lombardi o proprietari di seconde case nella veste di "untori". È facilmente prevedibile che ad ogni nuova emergenza potranno affacciarsi ancora altri bersagli, al punto da rendere molto difficile ancorare l'HS a una lista predefinita di target. Oltre ai target che si randomizzano, i discorsi d'odio diffusi nel web si

associano sempre più spesso a contenuti falsi e complottistici (Udupa et al. 2020) caratterizzati dal fatto di richiamare comportamenti di condivisione e partecipazione degli utenti, ma rispetto ai quali, comprendere se le reali intenzioni di chi li condivide siano il danneggiamento del target o il puro divertimento, non solo è molto complesso ma anche impossibile da cogliere se ci si sofferma soltanto all'esame del contenuto.

Questi brevi rimandi all'attualità hanno l'obiettivo di evidenziare come "i tratti comuni" (Sellars 2016) in base ai quali gli approcci di studio consolidati hanno definito l'HS, appaiono poco efficaci a spiegare le caratteristiche di un fenomeno ormai normalizzato e *ordinario*. Sebbene nel corso degli anni le definizioni si siano ampliate e arricchite, quest'impostazione originaria ha prevalso per lungo tempo e la tendenza dominante nella vastissima letteratura dedicata all'HS – di cui non è possibile dare conto in questa sede – è stata di intendere i discorsi d'odio come una deroga alla norma (Fumagalli, 2019) e di ancorarne la definizione ai seguenti requisiti: l'*intenzionalità*, ovvero della volontà esplicita del comunicatore di incoraggiare l'odio, il *contenuto* chiaramente interpretabile come un incitamento idoneo a provocare atti di odio o di violenza, i *target* specifici di questi atti e il *danno* provocato (Benesch 2012; Delgado 1982; Marwick e Miller 2014; Massey 1992; Moran 1994; Parekh 2012; Strossen 2001; Ward 1997). Vi è da dire che nel tempo vi sono stati importanti tentativi di rinnovamento dell'approccio, orientati talvolta a svincolare l'HS da requisiti troppo stringenti ripensando anche i gruppi target (Parekh 2012[2]), oppure a enfatizzare l'importanza di numerose variabili che concorrono a identificare un discorso d'odio, tra cui, la situazione o circostanza del discorso (Fumagalli, 2019), gli atteggiamenti dell'ascoltatore/pubblico (Langton 2012; Lawrence III 1990), l'autorevolezza di chi comunica (Benesch 2012; Gelber 2019) e il contesto in cui si colloca il discorso (Benesch 2012).

Ciononostante, l'impostazione oggi dominante continua a nostro avviso a scontare ancora alcuni limiti, in particolare, quello di affrontare l'HS principalmente come un'anomalia, un "discorso patologico" (Udupa e Pohjonen 2019) da diagnosticare, curare e bandire, e non come un fenomeno sociale e comunicativo di cui comprenderne gli aspetti più specifici, le dinamiche di circolazione e diffusione, le forme che assume, gli effetti di contagio che provoca sulle audiences che, per motivazioni le più diverse (ludiche, identitarie, etc.), sono portate spesso a riprodurre comportamenti aggressivi e discriminatori a loro volta. Se questa esigenza di ripensare l'approccio di studio dell'HS ha preso recentemente slancio in alcuni ambiti di ricerca, tra cui quello etnografico, segnalando l'importanza di uscire dal perimetro di un'indagine focalizzata sui contenuti che possono costituire un reato (Udupa e Pohjonen 2019), riteniamo che un altro passo importante in questa direzione possa derivare dall'adozione di una prospettiva teorica legata alla sociologia della comunicazione. Nel prossimo paragrafo si affronteranno le caratteristiche dei discorsi d'odio nel passaggio all'online, per poi esaminare i nuovi piani di lettura necessari allo studio delle pratiche di HS ordinarie che affollano il web.

L’odio online, cosa cambia?

L’odio online non muta certo nella sostanza la nozione tradizionale dei discorsi d’odio. In entrambi i casi, si è in presenza di comportamenti di soggetti e/o gruppi, più o meno organizzati, che puntano a umiliare, denigrare e trattare in modo discriminatorio altre persone accomunate da determinate caratteristiche (Ziccardi 2016). Tuttavia, in una società piattaformizzata (van Dijck, Poell e de Waal 2018), si modificano anche i flussi comunicativi, ridefiniti attraverso la logica degli algoritmi di visualizzazione sulle timelines degli utenti. Di conseguenza, nel ragionare sull’HS online, appare subito limitante il tentativo di definirlo focalizzando l’attenzione esclusivamente alla dimensione del contenuto, poiché acquisiscono un peso rilevante le più ampie dinamiche e relazioni comunicative sottostanti. Pur non mutando come si è detto nella sostanza, l’odio negli ambienti digitali richiede l’adozione di una prospettiva che esamini l’intersezione tra persone, tecnologie e pratiche, tenendo conto di come le *affordance* delle piattaforme (boyd 2014) contribuiscano a modellare queste pratiche, ridefinendo significati e caratteristiche dei discorsi d’odio.

Il primo elemento da considerare in questo senso è la capacità di propagazione dei messaggi all’interno di social media come Twitter, Facebook, YouTube o Instagram, che incoraggiano e semplificano la *diffusione* di contenuti da parte degli utenti. Direttamente legato a questo aspetto, è il dato della *visibilità* che distingue i contenuti maggiormente condivisi nelle reti di discussione online. Entrambe queste proprietà dei social media, diffusione e visibilità, nel caso dei discorsi d’odio che sono per definizione emotivamente connotati, possono contribuire ad accrescerne la capacità e velocità di propagazione con il rischio di trasformarli in messaggi virali. Si è riscontrato, infatti, che i contenuti (specie video) a forte tasso emozionale e, in particolare, quelli caratterizzati da emozioni negative, sollecitano maggiori comportamenti partecipativi da parte delle communities (sharing, liking, retweeting, ecc.) diffondendosi perciò più velocemente ed estensivamente (Ledwich, Zaitsev 2019; Nithyanand, Schaffner e Gill 2017). Senza dimenticare, poi, che i messaggi d’odio, una volta entrati in circolazione sono destinati a *permanere* online (al pari di qualsiasi altro messaggio), poiché sono le stesse piattaforme a supportarne la durata nel tempo (boyd 2014). Come ha notato il CEO dell'"Online Hate Prevention Institute", è chiaro che "più a lungo il contenuto rimane disponibile, più danni può infliggere alle vittime e responsabilizzare gli autori". [3] Tuttavia, essendo la permanenza una caratteristica insita della rete, non esiste alcuna garanzia circa l’eliminazione definitiva dei contenuti e di conseguenza, anche i discorsi estremi, violenti e pericolosi, una volta rimossi o bannati, possono riemergere sotto altre forme nella stessa piattaforma o ripresentarsi all’interno di altri circuiti web (Gagliardone et al. 2015).

Infine, appartiene alla logica di funzionamento degli ambienti generati dai social media e dal web anche la possibilità di *ricercare* i contenuti online, attraverso i motori di ricerca sia interni che esterni alle piattaforme, che permettono di ritrovare eventuali messaggi o materiali video slegati, però, dal contesto in cui sono stati prodotti e condivisi, con tutte le

conseguenze che ne possono derivare sul piano dell'uso e dell'interpretazione (Bentivegna e Boccia Artieri 2019).

Sebbene ciascuna piattaforma abiliti o limiti in modo più specifico la produzione e diffusione dei diversi tipi di messaggio, queste *affordance* creano nell'insieme un ambiente nel quale lo studio e l'interpretazione dei discorsi d'odio richiede approcci più specifici, dal momento che gli speaker, gli obiettivi, le motivazioni sottostanti, le tattiche, i contenuti e gli strumenti utilizzati non hanno medesime caratteristiche e significati riscontrabili nei discorsi d'odio tradizionali. A complicare ulteriormente il quadro, concorre, poi, il discorso sull'anonimato. Che si tratti di un anonimato "percepito" e "non reale" sembra essere di secondaria importanza allorché gli utenti, dietro lo "scudo" offerto dalla tecnologia (Ziccardi 2016), sono portati spesso a inasprire toni e linguaggi del confronto. Altrettanto rilevante nel favorire la circolazione di contenuti violenti online, oltre alla percezione di anonimato, contribuisce anche un diffuso senso di "deindividuazione" provato dagli utenti del web, ovvero dal fatto di percepire la propria identità individuale come meno rilevante rispetto all'essere parte di un gruppo, con l'effetto di disinibirne i comportamenti. Più forte all'interno di alcune piattaforme come YouTube (Halpern e Gibbs 2013) e Twitter (Oz et al. 2017) e meno in altre, questa sensazione di disinibizione concorre a sua volta a fare degli ambienti digitali degli spazi di potenziale crescita dell'HS. A renderli, in altre parole, un terreno favorevole alla diffusione dei contenuti d'odio sia da parte di gruppi organizzati ("eserciti") mossi da obiettivi condivisi e pianificati, sia da parte di utenti comuni, che, soprattutto a ridosso di specifici eventi o incidenti, tendono ad aggregarsi nella forma di "sciami" accomunati dal mettere in pratica comportamenti discriminatori più diffusi e spontanei (Gagliardone 2019). Il riferimento a questo tratto distintivo della produzione (organizzata o spontanea) dei discorsi di odio fa intuire a sua volta la complessità dell'operazione di individuazione e descrizione delle varie forme che essi possono assumere.

Le caratteristiche degli spazi digitali descritte fin qui, in un contesto in cui le forme di discriminazione e di intolleranza verso la diversità si sono progressivamente insinuate all'interno del tessuto sociale, rappresentano una sorta di acceleratore per la produzione, circolazione e propagazione dell'HS online. Allo stesso tempo, però, diventa anche più chiaro che la mutevolezza e fluidità delle forme che può assumere l'odio nel web è tale da rendere evidenti i limiti e le difficoltà connesse all'obiettivo di voler ricondurre l'HS a specifiche normative, basate su una rigida classificazione dei vari comportamenti sanzionabili. Si pensi, ad esempio, alle policies adottate dalle piattaforme digitali nel tentativo di arginare l'estendersi di questo fenomeno, da cui emerge un interesse focalizzato prevalentemente alla costruzione di casistiche, più o meno articolate, ma inevitabilmente soggette a revisioni e integrazioni derivanti dai mutamenti culturali, sociali e politici. Leggendo, ad esempio, le Linee guida della comunità di YouTube riguardo all'HS[4], si coglie come, a fronte di una lista estremamente articolata di caratteristiche che possono essere associate ai discorsi di odio e di esempi specifici, non vi sia un'analoga articolazione delle forme che tali discorsi possono assumere né vi è alcun riferimento ai contesti nei quali possono essere inseriti. Contesti che, talvolta, possono trasformare forme discriminatorie e di odio in qualcosa di liquidabile come "just funny stuff" (Haynes

2019: 3123). D'altro canto, non si può dimenticare che le piattaforme si muovono lungo un crinale complesso che, da un lato, vede la difesa della libertà di espressione come diritto fondamentale, dall'altro, deve tutelare gli utenti dagli abusi derivanti dall'appartenenza ad alcune categorie, pena l'instaurazione di un clima di violenza e odio tale da far abbandonare la piattaforma. Se in questo delicato esercizio di equilibrismo è indiscutibile lo sforzo da esse compiuto per ridurre l'estensione dell'HS, è anche chiaro che non rientra nelle loro finalità l'obiettivo, a noi caro, di comprendere le ragioni alla base della crescita dell'aggressività negli ambienti digitali e le implicazioni che ne derivano per i processi partecipativi degli utenti alla discussione online.

Le quattro "W" di un discorso d'odio

Per affrontare l'analisi dei discorsi d'odio da una prospettiva comunicativa partiremo dallo schema proposto da Harold Lasswell (1948) per descrivere un atto di comunicazione. Riadattando quel modello allo studio dell'HS online, si evidenzierà come le diverse dimensioni si condizionino reciprocamente e in maniera costante, al punto che esaminarle in maniera disgiunta e autonoma non permette di rispondere agli interrogativi di partenza di questo contributo e di spiegare i meccanismi e i tratti distintivi delle pratiche d'odio ordinarie diffuse online. Dato che il contesto in cui ci muoviamo è quello delle piattaforme digitali con le loro specificità, vediamo ora le altre dimensioni e relative sottodimensioni da considerare nello studio dei discorso d'odio online.

Il comunicatore (Who)

Prestare attenzione a chi comunica il messaggio è sempre stato un elemento centrale negli studi di comunicazione, dove l'autorevolezza della fonte viene considerata tradizionalmente come una variabile in grado di influenzare l'efficacia del processo comunicativo. In questa sede, tuttavia, tale questione assume un'importanza ancora più rilevante dal momento che il grado di autorevolezza e notorietà pubblica di chi comunica un discorso d'odio si traduce simmetricamente nell'aumento del potenziale corrosivo del messaggio. Ovvero, come vedremo più avanti, nelle ricadute che avrà il discorso d'odio, non tanto o solo per le conseguenze dirette sui target dello speech (immigrati, minoranze etniche o religiose, LGBT, ecc.), ma in termini di capacità di legittimazione di modelli comportamentali fondati sulla cultura della sopraffazione e della violenza. D'altronde già Goffman (1956) evidenziava come la posizione (di potere) di chi comunica incida sulla capacità di definizione della situazione di un messaggio, per cui separare un messaggio, nello specifico un discorso d'odio, dalla sua fonte, e dunque studiarne il contenuto senza tenere conto di chi lo ha prodotto, significherebbe non coglierne appieno il suo significato. Legato a questo aspetto, inoltre, va ricordato che l'autorevolezza della fonte assume in questa sede sfumature diverse, che hanno a che fare tanto con la posizione oggettiva dell'oratore – in termini di ruolo, status, potere – quanto con la notorietà conquistata sul

campo della rete, in ragione della sua capacità di influenzare un network di persone a tal punto rilevante da assicurare ai propri messaggi ampia visibilità e propagazione istantanea. Se pensiamo alla logica di funzionamento dei social media questi due tipi di potere spesso coincidono, perché gli hub centrali che guidano le dinamiche di opinione in rete corrispondono solitamente alle stesse élite tradizionali (giornalisti, politici, istituzioni), alle celebrity o ai blogger. Basti pensare, prendendo come esempio le élite politiche, al numero di followers di cui godono Barack Obama e Donald Trump (Twitter) o, nel caso italiano, Matteo Salvini (Facebook). Si tratta di un aspetto non secondario che evidenzia come, al di là delle premesse che hanno enfatizzato la dimensione orizzontale e paritaria della sfera pubblica online, in realtà la circolazione dei messaggi, inclusi quelli di HS, dipende da meccanismi di re-intermediazione da parte di questi snodi centrali, nella condizione oggettiva di poter influenzare il dibattito dentro e fuori la rete. Quando infatti la fonte del messaggio d'odio è un soggetto politico o un personaggio noto al grande pubblico, non solo aumenta la sua capacità di propagazione online, ma gli effetti si riverberano anche all'interno dei media mainstream, grazie a un effetto di rimbalzo dei contenuti tra media online e offline (Rega, 2019). Nel caso della campagna presidenziale del 2016, ad esempio, Donald Trump è riuscito attraverso i suoi tweet al vetriolo a conquistare un potere "algoritmico" sui social media (più followers, reazioni, commenti, retweet, like) che si è tradotto simmetricamente nella capacità di dominare la copertura mediatica e influenzare l'agenda politica (Faris et al. 2017). In altre parole, grazie a un uso attento dei social media, i suoi temi, l'immigrazione e la questione musulmani/Islam e, ancora più importante, le cornici interpretative che li hanno accompagnati improntate alla cultura della discriminazione e dell'esclusione, hanno prevalso anche all'interno degli altri media, diventando il centro del dibattito politico nel corso dell'intera campagna. Allo stesso tempo è altrettanto chiaro che, in un'ottica di azione e retroazione, anche i membri dell'audience che entrano in contatto, più o meno casualmente, con un discorso di odio di Trump o di un altro soggetto possono, a loro volta, assumere i panni del "comunicatore", rilanciandone il contenuto. Se questo aspetto sarà ripreso e sviluppato più avanti (4.3 Il destinatario), va da subito colto come l'imprevedibilità e la mobilità del ruolo del comunicatore – oltre che la sua coincidenza, talvolta, con l'audience – rendono la sua individuazione un compito non sempre possibile da realizzare.

In relazione al comunicatore del messaggio, infine, non si possono ignorare alcuni lavori di ricerca sull'HS riguardanti i commenti e le discussioni degli *ordinary user*. È interessante notare che quando gli "haters" sono cittadini comuni, come emerso da alcuni studi realizzati su piattaforme di social media (Twitter, Reddit), sono più spesso di sesso maschile e tendono ad essere membri di comunità densamente collegate, dove circolano contenuti di HS e si è soliti ritwittare i messaggi d'odio vicendevolmente (Costello e Hawdon 2018). Un meccanismo, questo, collegato alle motivazioni latenti che ispirano i comportamenti violenti, ma che segnala al tempo stesso la natura *relazionale* dei processi comunicativi; la condivisione dei messaggi d'odio, infatti, risponde spesso all'esigenza da parte delle persone di confermare la propria appartenenza a un gruppo o community, consolidando indirettamente l'identità del soggetto. Non è un caso, infatti, che l'aumento di comportamenti online ispirati all'odio e alla violenza è fortemente associato alla presenza

sempre più numerosa di gruppi di hate speech (Costello e Hawdon 2018) all'interno dei quali la circolazione dei contenuti d'odio è pressocché normalizzata e le pratiche di produzione e/o diffusione di tali contenuti da parte dei partecipanti diventando modalità spesso intrattenitive che rafforzano il senso di identità e vicinanza tra i soggetti.

Il messaggio (What)

Se il secondo piano di lettura riguarda il messaggio, è utile distinguerne due sottodimensioni: la *sostanza/contenuto* del discorso e i suoi elementi di *forma*, ovvero il tono e le parole scelte dal comunicatore. In riferimento alla componente sostanziale del messaggio, vanno valutate innanzitutto quelle modalità di intolleranza e discriminazione delle persone prese in considerazione dalla letteratura consolidata sull'HS e, dunque, i contenuti che stigmatizzano il target sulla base di caratteristiche ascritte (sesso, nazionalità, razza, religione, ecc.). Forme di questo tipo, sono la xenofobia, il razzismo, l'esterofilia, l'omofobia, la misoginia, che possono anche materializzarsi attraverso l'uso di stereotipi offensivi per etichettare i membri di gruppi target (Papacharissi 2004) o di espressioni d'odio e disprezzo nei loro confronti.
A ben vedere, prendendo ad esempio il caso dei rappresentanti politici, emerge nell'immediato che non si tratta di modalità difficilmente riscontrabili. Ricordiamo ancora tutti le parole con cui Donald Trump annunciò la sua candidatura alle primarie repubblicane del 2016, quando, rivolgendo l'attenzione agli immigrati messicani, disse: "Stanno portando la droga. Stanno portando il crimine. Sono stupratori. E alcuni, presumo, sono brave persone".[5] Ma lo stesso vale anche prendendo a riferimento il caso italiano e, in particolare, i messaggi di Matteo Salvini contro gli immigrati, tra cui un video postato su Twitter durante la campagna per le Politiche 2018 accompagnato dalle seguenti parole: "ENNESIMA protesta in provincia di Padova: non sono soddisfatti dell'accoglienza in hotel! Roba da matti! TUTTI A CASA!!! #4marzovotoLega" (titolo del video: Secondo voi questi scappano da guerre??. https://twitter.com/matteosalvinimi/status/958064490284769280).

In entrambi i casi, tornando alla definizione di Parekh (2012), si tratta di messaggi diretti contro un individuo/gruppo in base a una caratteristica normativamente irrilevante, che stigmatizzano il gruppo target (messicani e profughi negli esempi) attribuendogli implicitamente o esplicitamente qualità considerate sgradevoli e che implicano che esso sia visto come una presenza indesiderabile e dunque oggetto di ostilità. Il problema è che se in questi casi è facile riconoscere la presenza di hate speech, lo stesso non accade esaminando discorsi offensivi di vario genere che circolano quotidianamente online. Ciò che è apparso chiaro nelle ricerche di questi ultimi anni riguardanti i discorsi d'odio negli spazi web, è che non solo siamo spesso in presenza di target più estemporanei, random e difficilmente associabili all'HS, ma contemporaneamente le pratiche e i processi che accompagnano il discorso d'odio sono cambiati in funzione dei linguaggi e delle estetiche tipiche di Internet, dove fenomeni come il trolling, il doxing o il luzl (culture del piacere) rappresentano una parte costitutiva della cultura della rete. La discriminazione online (razziale, sociale, culturale ecc.) e i discorsi carichi di avversità, pur rafforzando visioni

improntate all'esclusione o denigrazione, sono resi piacevoli e divertenti (Udupa 2019) e dunque più difficili da riconoscere.

Ciò non significa che le tradizionali forme di discriminazione scompaiono, ma più realisticamente che si combinano a nuovi tipi di HS. Esaminando per esempio i discorsi d'odio più comuni all'interno dei social media (Twitter e Whisper), Mondal e colleghi (2018) hanno riscontrato che, oltre ai messaggi basati sulla razza, prevalgono nettamente quelli che offendono il target sulla base di aspetti comportamentali (insicuro, goffo, lento) o fisico-estetici (obeso, brutto) rispetto ai messaggi di discriminazione fondati sulla religione, la disabilità, il genere o l'etnia. Pur non costituendo un reato penale, queste forme di ostilità che popolano il web possono comunque danneggiare le persone e lo stesso discorso vale per quei tipi di HS che sono espressione di un antagonismo tout court, inteso alla creazione di un "nemico" che non necessariamente rimanda ai gruppi storicamente oppressi (Matsuda 1989) o emarginati (Marwick e Miller 2014).

Oltre alla sostanza del messaggio bisogna poi porre attenzione al *tono* e alle *parole* scelte dal comunicatore, tali da rendere quello speech potenzialmente più infiammante. Prendendo il caso degli appelli alla rabbia o alla paura, intesi per definizione a provocare delle risposte viscerali da parte del pubblico, essi si basano su toni allarmistici e registri apodittici che, se nello speech dal vivo dipendono soprattutto dal tono di voce del comunicatore, nella comunicazione online si traducono nell'uso deliberato del maiuscolo, nel ricorso a testi ingranditi, punti esclamativi e simboli iconici di vario genere (triangoli e cerchi rossi, stelle, emoticon etc.). In altre parole, sia online che offline, la tonalità manifesta dei messaggi d'odio appare relativamente semplice da rintracciare. Più problematica è, invece, la questione del lessico adottato dagli odiatori in rete, i quali, al fine di aggirare le policies delle piattaforme ed evitare di incorrere in segnalazioni e rimozioni del contenuto, negli ultimi anni hanno ammorbidito il linguaggio. Se un tempo era più semplice imbattersi in messaggi caratterizzati da un lessico esplicito e dunque da stereotipi etnici o sociali e parole d'odio dispregiative per natura (es: negro, frocio, ritardato; De Mauro 2016), oggi l'HS che circola nei social media è più indiretto e velato, basato sull'uso di termini comuni ma piegati all'odio (es: immigrato; Ferrini e Paris 2019, p. 39), sul ricorso ad allusioni ed espressioni ironiche o indirette. Senza dimenticare, infine, che un altro espediente segnalato da diverse ricerche consiste nell'uso di codici e sottocodici di esclusiva o prevalente comprensione all'interno di determinate communities ma più criptici o addirittura privi di senso per un osservatore esterno (molto diffuso, ad esempio, tra i membri dell'Alt-Right).

Il destinatario (To Whom)

La terza prospettiva di analisi dei discorsi d'odio fa riferimento al destinatario del messaggio, rispetto al quale diversi sono gli elementi che assumono rilevanza e che ci portano anche a fare delle distinzioni rispetto ad altri approcci di studio, in particolare, rispetto al "dangerous speech" (Benesch 2012). Riguardo all'audience, infatti, la studiosa evidenzia l'importanza, nel determinare il livello di pericolosità di un discorso, della

dimensione fisica del pubblico (grandezza e numerosità) e della sua dimensione emotiva, notando come un pubblico rabbioso o sensibile alla paura, per via ad esempio di minacce, traumi o violenze subite in passato (non necessariamente di persona), possa essere influenzato più facilmente da parte di un istigatore e reagire violentemente a sua volta. Tali ricerche, però, si concentrano su paesi dilaniati da guerre intestine e conflitti inter-etnici di lunga data (Benesch et al. 2020).

Uscendo dallo stretto perimetro dei discorsi volti a ispirare l'uso della violenza fisica contro altre persone e rivolgendo l'attenzione alle innumerevoli forme di ostilità, più o meno esplicite, che circolano diffusamente sul web, va subito chiarito che la natura del pubblico della rete muta radicalmente e che la sua suscettibilità è spesso il frutto di un altro tipo di dinamiche. Ciò premesso, partiamo da una prima considerazione, ovvero che le audiences online sono attive e connesse per definizione, secondo logiche partecipative che vanno oltre le sole pratiche di ricezione. Gli utenti della rete sperimentano, infatti, una doppia condizione quella di "avere un pubblico" e, contemporaneamente, di "essere parte di un pubblico" (Boccia Artieri 2012) e perciò possono fruire i contenuti di HS ma al tempo stesso interagire con essi e produrne a loro volta, generando potenzialmente delle proprie audiences. In tal senso è chiaro che, partecipando alla diffusione dei messaggi d'odio, commentandoli, trasformandoli e re-distribuendoli online, i pubblici della rete vanno anche considerati nella veste di un "secondo o un terzo comunicatore" che interviene nella messa in circolazione dell'ostilità, con più o meno incisività a seconda del network di persone in grado di influenzare.

In secondo luogo è altrettanto vero che i pubblici possono essere parte di comunità più o meno connesse tra loro, e che la produzione e diffusione dei contenuti d'odio nel web può essere tanto il frutto delle attività di audiences strutturate in gruppi definiti da interessi stabili e di vario genere – politico, sportivo, sociale, ludico, etc. – quanto di pubblici non strutturati, caratterizzati da dinamiche di aggregazione più fluide e spontanee, definite da interessi estemporanei e contingenti (*ad hoc publics*; Bruns e Burgess, 2011). In questo secondo caso, è soprattutto l'istantaneità delle dinamiche comunicative e aggregative tipiche del web ad essere alla base di fenomeni come lo *shitstorm* in cui, utenti guidati principalmente dal trasporto emotivo (affect), partecipano nella forma di "sciami" a campagne d'odio di diverso tipo che possono consumarsi anche nel giro di un solo giorno. Il classico schema binario "Noi" vs "Loro", secondo il quale la creazione di identità si basa sulla distinzione da identità avversarie, assume perciò nuove sfaccettature e valenze, che talvolta rimandano a fratture sociali di lunga data, ma altrettanto spesso rimandano a meccanismi più transitori ma importanti da esaminare, perché rivelatori di un malessere sociale più diffuso e persistente. Nel caso di gruppi definiti da interessi più stabili, inoltre, la contrapposizione tra i membri del proprio gruppo (l'*in-group*) e quelli esterni (l'*out-group*) evidenzia anche la doppia funzione del discorso d'odio che, da un lato, rafforza l'identità e solidità interna al gruppo, agevolando anche la socializzazione (e talvolta il reclutamento) di nuovi membri; dall'altro, accresce i sentimenti di rabbia e avversità contro il gruppo esterno portando, contestualmente, a una crescita di meccanismi di antagonismo generalizzato.

Se all'interno dei social media questo legame tra la polarizzazione delle audiences e i discorsi ostili è stato attentamente considerato (Humprecht et al. 2020), occorre altresì valutare in che modo il rapporto tra il comunicatore e il pubblico può a sua volta influenzare queste dinamiche. Il successo di un messaggio d'odio, infatti, dipende anche dall'esistenza di un *terreno comune* di credenze e atteggiamenti tra il portatore d'odio e gli ascoltatori (Langton 2012; Lawrence III 1990), per cui la dirompenza di questi messaggi si lega in gran parte alla capacità del comunicatore di rispondere alle attese del proprio uditorio. Si conferma, dunque, quanto già anticipato sulla natura profondamente relazionale dei processi comunicativi che deve essere tenuta presente anche nello studio del discorso d'odio, perché è evidente che se il livello di empatia e identificazione tra il comunicatore e l'audience è più forte, anche le reazioni a un messaggio di HS saranno più violente, materializzandosi nella produzione di ulteriore ostilità da parte del pubblico secondo un meccanismo di riproduzione circolare dell'odio e del rancore. L'esistenza di questo meccanismo, di cui esistono già dei riscontri empirici, apre la strada all'ultimo piano di analisi da considerare nel nostro modello, ovvero quello degli effetti.

Gli effetti (With What Effect)

Premesso che gli effetti dei messaggi d'odio non possono essere valutati in astratto ma dipendono da tutti gli elementi sin qui considerati, porci il problema di esaminarli ha un senso non scontato e che è bene chiarire. Come si è detto, la complessità dell'oggetto di studio è tale che non esiste una definizione certa e immutabile di cosa sono i discorsi d'odio e, contemporaneamente, gli studi sul tema contribuiscono talvolta ad alimentare l'incertezza e le ambiguità terminologiche intorno al fenomeno, utilizzando spesso come sinonimi concetti tra loro simili ma non sovrapponibili (hate speech, incivility, rudeness). Ora, a prescindere dalle definizioni e categorizzazioni da utilizzare, riteniamo che sia più utile, al fine di cogliere il senso delle pratiche ordinarie di HS online, osservare le loro conseguenze in un'ottica comunicativa. Se in letteratura hanno prevalso due principali interpretazioni (Fumagalli 2019), una focalizzata su come il comunicatore può danneggiare il gruppo target (Matsuda 1989; Waldron 2012) e l'altra su come può motivare ad agire contro il gruppo target (Langton 2012; Lawrence III 1990), in questo approccio vogliamo concentrare l'attenzione sugli effetti che i discorsi d'odio online provocano sugli utenti delle piattaforme digitali esposti a questi messaggi e in grado di reagire a loro volta e prendere la parola. Affrontare l'HS attraverso un framework comunicativo significa, infatti, interrogarsi sul clima di aggressività che permea oggi la discussione online, esaminando perciò il tenore dei commenti e delle risposte che scaturiscono dai messaggi d'odio. Sfruttando la caratteristica dei social media di connotarsi come spazi di interazione e scambio, si tratta di estendere l'analisi alle risposte degli utenti, osservando, più in generale, la qualità della discussione e del confronto che si animano online in presenza di contenuti d'odio. L'esempio classico che viene studiato in questa direzione, sono i forum delle testate informative, nati originariamente come un luogo di empowerment per la sfera pubblica, dove cittadini eterogenei per status, interessi, background possono avere accesso libero e paritario al dibattito su temi di interesse

reciproco. Al di là del fatto che in molti casi questi spazi sono stati costretti a chiudere per la presenza di linguaggi violenti, diversi studi hanno evidenziato gli effetti detrimenti dell'hate speech sulla partecipazione e sulla qualità e il valore democratico della discussione. Non solo la sua presenza influenza negativamente l'atmosfera della discussione, inibendo in molti casi la presa di parola da parte di altri utenti al confronto (Ziegele et al. 2018), ma accresce anche la polarizzazione e gli atteggiamenti stereotipati sui gruppi sociali da parte delle persone, compromettendo il potenziale di deliberazione del dibattito online (Chen 2017). Esattamente l'opposto di quello che dovrebbe essere il circolo virtuoso della partecipazione alla discussione attraverso il confronto reciproco.

Altrettanto significativo e da tener presente è poi l'effetto di "imitazione", ovvero la tendenza delle persone esposte a messaggi o comportamenti denigratori, incivili e violenti a reiterarli a loro volta (Gervais 2015; Gervais 2016; Rega e Marchetti 2019), secondo un meccanismo di *contagio* tra comunicatore e audience. In presenza di comunicatori autorevoli e con elevato potere di influencer è chiaramente più facile che si possano innescare comportamenti imitativi più ampi ed esperienze di *negative affect* condivise, come l'ostilità e la rabbia collettive. Dinamiche di questo tipo sono state rintracciate, ad esempio, in conseguenza di molti tweet d'odio veicolati da Donald Trump prima di diventare Presidente degli Stati Uniti (Pain e Masullo Chen 2019). Non solo, ma nel momento in cui il comunicatore è un soggetto politico o istituzionale, a prescindere dal fatto che il messaggio postato online sia un contenuto che incita inequivocabilmente alla violenza o se si tratti di un contenuto più ambiguo o velato, gli effetti che provoca possono comunque portare a una degenerazione della discussione tra gli utenti. Ciò si è verificato ad esempio in Italia durante le campagne elettorali delle Politiche 2018 e delle Europee 2019 e, in particolare, in relazione ai contenuti d'odio più o meno espliciti postati su Facebook dai leader e candidati di Lega e Fratelli d'Italia: i commenti che si sono scatenati in risposta a questi messaggi hanno mostrato con chiarezza il disprezzo nei confronti della controparte, soprattutto nei confronti di immigrati, rom, donne, supporter di altri partiti e candidati avversari; contemporaneamente, però, si è anche vista la tendenza a un peggioramento della discussione tra gli utenti, degenerata in uno scambio di ingiurie e aggressività reciproca (Amnesty International Report 2019[6]).

Nel complesso, emerge dunque il potere corrosivo dell'HS e il suo potenziale di accrescere la polarizzazione e il livore dei pubblici, alimentando persino la reiterazione di comportamenti aggressivi e discriminatori da parte dei cittadini. Per questo, specialmente nel caso di una retorica pubblica sempre più aggressiva, il suo impatto, soprattutto online, va ben oltre gli effetti diretti provocati sulle vittime o sui gruppi bersaglio, perché mina il valore democratico della discussione e dei processi partecipativi, legittimando, contemporaneamente, modelli comportamentali fondati sulla cultura della discriminazione, e della violenza.

Conclusioni

Il punto di partenza di questa riflessione sulle caratteristiche dei discorsi d'odio contemporanei è che ci troviamo dinanzi a un concetto in continua trasformazione quanto a forme e modalità assunte, un "moving concept" rispetto al quale l'adozione di categorie descrittive basate solo sul contenuto del messaggio non permette di coglierne i tratti e i meccanismi distintivi. Nel passaggio all'online, infatti, non solo il numero degli attori coinvolti nei discorsi d'odio si è ampliato e i ruoli da essi assunti sono diventati più indistinti, ma anche gli effetti che tali messaggi sono in grado di innescare non sono più circoscrivibili nel tempo, nello spazio e nel numero delle vittime potenzialmente coinvolte. Senza dimenticare, poi, che il passaggio al digitale ha anche coinciso con una nuova visibilità e diffusione del fenomeno, tale da farci ritenere non più rimandabile la necessità di aprire una riflessione su quali siano le specificità dell'HS negli spazi web e quali siano le ragioni alla base della sua proliferazione. Abbandonando una certa visione dominante che valuta l'HS come un discorso patologico o come un'anomalia da bandire o censurare, e puntando invece a sintonizzarci con le pratiche culturali e digitali messe in campo dagli utenti nella produzione dei discorsi di odio, abbiamo provato a ragionare su questo fenomeno da una prospettiva sociologico-comunicativa che considera l'HS nei termini di un processo comunicativo. Al pari degli altri, dunque, anche in questo caso va evidenziata la sua natura *relazionale* e, di conseguenza, il fatto che il suo significato dipende dall'interazione tra i diversi elementi coinvolti, la fonte, il messaggio, il pubblico, tenendo sempre presente il contesto della comunicazione e gli effetti che scaturiscono dai rapporti tra le diverse variabili. Se si scollegano questi elementi o si analizzano in maniera isolata, difficilmente si può restituire il senso complessivo del processo comunicativo e capire perché uno stesso messaggio può avere un potere di influenza e un impatto di più ampia portata in determinate situazioni e circostanze. Per esempio, se il comunicatore è un soggetto politico-istituzionale, se utilizza il proprio account di social media per diffondere informazioni distorte su temi sensibili, se il pubblico è in uno stato di particolare vulnerabilità e così via. Pensiamo solo un momento alla recente attualità e, in particolare, alle scelte che hanno indotto piattaforme come Twitter e Facebook a segnalare o rimuovere alcuni messaggi di Donald Trump. Ad esempio, a ridosso dell'uccisione di George Floyd ("When the looting starts, the shooting starts", tweet del 29-05-2020) o nel caso di affermazioni infondate riguardanti il Covid-19, tra cui quelle relative alla quasi immunità dei bambini al virus (post su Facebook contenente la video-intervista rilasciata da Trump a Fox News il 5-08-2020). Le stesse affermazioni sostenute da un cittadino comune con un limitato numero di followers, una scarsa capacità di influenzare opinioni e comportamenti collettivi o di suscitare empatia da parte del pubblico, probabilmente non sarebbero state considerate come comunicazioni decisive e pericolose. Il punto essenziale, detto più esplicitamente, è che il potere di influenza di una comunicazione dipende dall'insieme delle dimensioni implicate nel processo comunicativo; non soltanto il messaggio in sé, ma anche le caratteristiche del comunicatore, il contesto in cui rilascia il discorso, il pubblico a cui si rivolge, le sue aspettative e la sua suscettibilità, valutando contemporaneamente le dinamiche di azione e retroazione tra tutti questi elementi contemporaneamente. Dinanzi all'ampia mole di ricerche sull'HS che analizzano (selettivamente) porzioni specifiche del fenomeno e focalizzano l'attenzione soprattutto sul

messaggio senza considerare gli altri elementi, questo contributo ha cercato di riadattare il modello multidimensionale di Lasswell allo studio dei discorsi d'odio, enfatizzando però, differentemente dal politologo statunitense, la necessità di esaltare i rapporti di reciprocità e feedback tra le diverse dimensioni convolte nel discorso d'odio.

È chiaro che disporre di un framework teorico-analitico attraverso cui studiare l'HS è soltanto un primo passo e in prospettiva sarà necessario applicare questo modello a uno specifico caso di studio per mostrarne la sua utilità. In questa sede, tuttavia, ci interessava mettere a punto una proposta teorico-operativa che potesse guidare la ricerca su questo fenomeno e sulle peculiarità assunte oggi rispetto al passato. Perché se è vero che l'odio online non differisce dall'odio *tout court* è altrettanto vero che possono verificarsi slittamenti comunicativi progressivi tali da dare vita alla trasformazione di forme di discriminazione in qualcosa di apparentemente innocente come "just funny stuff".

Note biografiche

Sara Bentivegna, insegna Comunicazione Politica alla "Sapienza" Università di Roma. Tra le più recenti pubblicazioni, *Le teorie delle comunicazioni di massa e la sfida digitale* (con Giovanni Boccia Artieri, Bari 2019) e *A colpi di tweet* (Bologna 2015).

Rossella Rega, insegna Giornalismo e Nuovi Media e Media Industry and strategic communication presso il Dipartimento di Scienze Sociali, Politiche e Cognitive dell'Università di Siena. Tra le sue ultime pubblicazioni, *Les Fans d'Apple: enquête sur les réseaux sociaux* (Parigi 2016).

Bibliografia

Benesch, S. (2012). Dangerous speech: A proposal to prevent group violence. *Voices That Poison: Dangerous Speech Project*. Preso da: https://dangerousspeech.org/wp-content/uploads/2018/01/Dangerous-Speech-Guidelines-2013.pdf.

Benesch, S., Buerger, C., Glavinic, T., e Manion, S. (2020). Dangerous Speech: A Practical Guide. *Dangerous Speech Project*, 1–14. Preso da: https://dangerousspeech.org/guide/

Bentivegna, S., e Boccia Artieri, G. (2019). *Le teorie delle comunicazioni di massa e la sfida digitale*. Roma: Laterza.

Boccia Artieri, G. (2012). *Stati di connessione. Pubblici, cittadini e consumatori nella (Social) Network Society*. Milano: Franco Angeli.

boyd, danah. (2014). *it's complicated. The social lives of networked teens*. New Haven: Yale University Press. Preso da https://www.danah.org/books/ItsComplicated.pdf

Bruns, A., e Burgess, J. (2011). The Use of Twitter Hashtags in the Formation of Ad Hoc Publics, (August), 25–27. Preso da http://eprints.qut.edu.au/46515

Chen, G. M. (2017). *Online Incivility and Public Debate: Nasty Talk*. New York: Palgrave Macmillan.

Cohen-Almagor, R. (2011). Fighting Hate and Bigotry on the Internet. *Policy & Internet*, *3*(3), 89–114. https://doi.org/10.2202/1944-2866.1059

Costello, M., e Hawdon, J. (2018). Who Are the Online Extremists Among Us? Sociodemographic Characteristics, Social Networking, and Online Experiences of Those Who Produce Online Hate Materials. *Violence and Gender*, 5(1), 55–60. https://doi.org/10.1089/vio.2017.0048.

De Mauro, T. (2016). Le parole per ferire - Tullio De Mauro - Internazionale. Preso da: https://www.internazionale.it/opinione/tullio-de-mauro/2016/09/27/razzismo-parole-ferire.

Delgado, R. (1982). Words that wound: A tort action for racial insults, epithets, and name-calling. *Harv. CR-CLL Rev.133*, *17*.

Faris, R., Roberts, H., Etling, B., Bourassa, N., Zuckerman, E., e Benkler, Y. (2017). Partisanship, propaganda, and disinformation: Online media and the 2016 US presidential election. *Berkman Klein Center for Internet & Society Research Paper*. Preso da: http://nrs.harvard.edu/urn-3:HUL.InstRepos:33759251.

Ferrini, C., e Paris, O. (2019). *I discorsi dell'odio : razzismo e retoriche xenofobe sui social network*. Roma: Carocci.

Fumagalli, C. (2019). Discorsi d'odio come pratiche ordinarie. *Biblioteca Della Libertà*, *54*(224), 55–75. https://doi.org/10.23827/BDL_2019_1_3

Gagliardone, I., Gal, D., Alves, T., e Martinez, G. (2015). *Countering Online Hate Speech*. (Intergovernmental Panel on Climate Change, Ed.), *Climate Change 2013 - The Physical Science Basis*. United Nations Educational, Scientific and Cultural Organization.

Gagliardone, I. (2019). Defining Online Hate and Its " Public Lives ": What is the Place for " Extreme Speech "?. *International Journal of Communication*, *13*, 3068–3087. https://doi.org/1932–8036/20190005

Gelber, K. (2019). Differentiating hate speech: a systemic discrimination approach. *Critical Review of International Social and Political Philosophy*, 00(00), 1–22. https://doi.org/10.1080/13698230.2019.1576006

Gervais, B. T. (2015). Incivility Online: Affective and Behavioral Reactions to Uncivil Political Posts in a Web-based Experiment. *Journal of Information Technology & Politics*, *12*(2), 167–185. https://doi.org/10.1080/19331681.2014.997416

Gervais, B. T. (2016). More than Mimicry? The Role of Anger in Uncivil Reactions to Elite Political Incivility. *International Journal of Public Opinion Research*, *29*(3), edw010. https://doi.org/10.1093/ijpor/edw010

Goffman, E. (1956). *The presentation of self in everyday life*. New York, NY: Doubleday.

Halpern, D., e Gibbs, J. (2013). Social media as a catalyst for online deliberation? Exploring the affordances of Facebook and YouTube for political expression. *Computers in Human Behavior*, *29*(3), 1159–1168. https://doi.org/10.1016/j.chb.2012.10.008

Haynes, N. (2019). Writing on the Walls : Discourses on Bolivian Immigrants in Chilean Meme Humor. *International Journal of Communication*, *13*, 3122–3142. Preso da: https://ijoc.org/index.php/ijoc/article/view/9109/2714.

Humprecht, E., Hellmueller, L., e Lischka, J. A. (2020). Hostile Emotions in News Comments: A Cross-National Analysis of Facebook Discussions. *Social Media + Society*, *6*(1), 1-12. https://doi.org/10.1177/2056305120912481

Langton, R. (2012). Beyond belief: Pragmatics in hate speech and pornography. In I. Maitra & M.K. McGowan (eds) (pp. 72–93). *Speech and Harm: Controversies over Free Speech*. Oxford: Oxford University Press.

Lasswell, H. D. (1948). The structure and function of communication in society. In L. Bryson (Ed.). *The Communication of Ideas*. New York, New York, USA: Harper and Brothers.

Lawrence III, C. R. (1990). If he hollers let him go: Regulating racist speech on campus. *Duke Law Journal*, *1990*(3), 431–483. Preso da: https://scholarship.law.duke.edu/dlj/vol39/iss3/2

Ledwich, M., e Zaitsev, A. (2019). Algorithmic Extremism: Examining YouTube's Rabbit Hole of Radicalization. *ArXiv Preprint ArXiv:1912.11211*. Preso da: https://arxiv.org/abs/1912.11211

Marwick, A., e Miller, R. (2014). Online Harassment, Defamation, and Hateful Speech: A Primer of the Legal Landscape. Preso da http://ir.lawnet.fordham.edu/clip

Massey, C. R. (1992). Hate Speech, Cultural Diversity, and the Foundational Paradigms of Free Expression. Preso da: http://repository.uchastings.edu/faculty_scholarshiphttp://repository.uchastings.edu/faculty_scholarship/1376

Matsuda, M. J. (1989). Public Response to Racist Speech: Considering the Victim's Story. *Michigan Law Review*, *87*(8), 2320. https://doi.org/10.2307/1289306

Mondal, M., Silva, L. A., Correa, D., e Benevenuto, F. (2018). Characterizing usage of explicit hate expressions in social media. *New Review of Hypermedia and Multimedia*, *24*(2), 110–130. https://doi.org/10.1080/13614568.2018.1489001

Moran, M. (1994). Talking about Hate Speech: A Rhetorical Analysis of American and Canadian Approaches to the Regulation of Hate Speech. *Wisconsin Law Review*, *1994*, 2320–2321. https://doi.org/10.2307/1289306

Nithyanand, R., Schaffner, B., e Gill, P. (2017). Measuring Offensive Speech in Online Political Discourse. Preso da: http://arxiv.org/abs/1706.01875

Ott, B. L. (2017). The age of Twitter: Donald J. Trump and the politics of debasement. *Critical Studies in Media Communication*, *34*(1), 59–68. https://doi.org/10.1080/15295036.2016.1266686

Oz, M., Zheng, P., & Chen, G. M. (2018). Twitter versus Facebook: Comparing incivility, impoliteness, and deliberative attributes. *New media & society*, 20(9), 3400-3419. https://doi.org/10.1177/1461444817749516

Pain, P., e Masullo Chen, G. (2019). The President Is in: Public Opinion and the Presidential Use of Twitter. *Social Media + Society*, *5*(2), 1-12. https://doi.org/10.1177/2056305119855143

Papacharissi, Z. (2004). Democracy online: civility, politeness, and the democratic potential of online political discussion groups. *New Media & Society*, *6*(2), 259–283. https://doi.org/10.1177/1461444804041444

Parekh, B. (2012). Is there a case for banning hate speech? In *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (pp. 37–56). Cambridge University Press. https://doi.org/10.1017/CBO9781139042871.006

Rega, R. (2019). Discorsi d'odio e parole ostili come specchio della realtà politica contemporanea. *Rivista di Studi Politici*, n. 4/2019, 153-174.

Rega, R. e Marchetti, R. (2019). L'Incivility nelle Politiche 2018. Fine del dibattito pubblico?". *Comunicazione politica*, 1, 15-38.

Sellars, A. (2016). Defining Hate Speech. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2882244

Strossen, N. (2001). Incitement to hatred: Should there be a limit. S. Ill. ULJ, 25, 243–279. Preso da: https://digitalcommons.nyls.edu/cgi/viewcontent.cgi?article=1185&context=fac_articles_chapters

Udupa, S., e Pohjonen, M. (2019). Extreme Speech| Extreme Speech and Global Digital Cultures—Introduction. *International Journal of Communication*, 13, 19. Preso da: https://ijoc.org/index.php/ijoc/article/view/9103

Udupa, S. (2019). Nationalism in the Digital Age : Fun as a Metapractice of Extreme Speech. *International Journal of Communication*, *13*, 3143–3163. https://doi.org/1932–8036/20190005

Udupa, S., Gagliardone, I., Deem, A., e Csuka, l. (2020). Hate speech, information disorder, and conflict. Preso da http://ssrc-cdn1.s3.amazonaws.com/crmuploads/new_publication_3/the-field-of-disinformation-democratic-processes-and-conflict-prevention-a-scan-of-the-literature.pdf

Van Dijck, J., Poell, T., e De Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford University Press.

Waldron, J. (2012). *The harm in hate speech*. Harvard University Press.

Ward, K. D. (1997). Free Speech and the Development of Liberal Virtues: An Examination of the Controversies Involving Flag-Burning and Hate Speech. *U. Miami L. Rev.733*, *52*. Preso da https://repository.law.miami.edu/umlr/vol52/iss3/4

Ziccardi, G. (2016). *L'odio online: violenza verbale e ossessioni in rete*. Milano: Cortina Raffaello.

Ziegele, M., Jost, P., Bormann, M., e Heinbach, D. (2018). Journalistic counter-voices in comment sections: Patterns, determinants, and potential consequences of interactive moderation of uncivil user comments. *Studies in Communication | Media*, *7*(4), 525–554. https://doi.org/10.5771/2192-4007-2018-4-525

Note

[1] Nell'articolo faremo riferimento all'hate speech (più spesso nella forma contratta HS) e ai discorsi d'odio in maniera intercambiabile.

[2] Parekh (2012) riconduce l'HS a tre caratteristiche essenziali: (i) è diretto contro un individuo specifico o un gruppo di individui in base a una caratteristica arbitraria o normativamente irrilevante, (ii) stigmatizza il gruppo target attribuendogli implicitamente o esplicitamente qualità considerate indesiderabili, (iii) implica che il gruppo target sia visto come una presenza indesiderabile perciò legittimamente oggetto di ostilità.

[3] Andre Oboler, 31 Ottobre 2014 (traduzione a nostra cura; citato in Gagliardone et al. 2015).
[4] Per le norme sull'incitamento all'odio di YouTube si vedano i seguenti link: https://www.youtube.com/intl/it/about/policies/#community-guidelines; https://support.google.com/youtube/answer/2801939?hl=it (consultazione del 14/8/2020).
[5] "When Mexico sends its people, they're not sending their best (…). They're not sending you. They're not sending you. They're sending people that have lots of problems, and they're bringing those problems with us. They're bringing drugs. They're bringing crime. They're rapists. And some, I assume, are good people"; citazione tratta da: https://time.com/4473972/donald-trump-&/.
[6] L'analisi della campagna delle Europee 2019 è disponibile online: https://www.amnesty.it/cosa-facciamo/elezioni-europee/.