

A.I. e linguaggio dell'odio. Per una didattica della tolleranza

Raffaella Petrilli
Università della Tuscia

PATHS Summer School – VI edizione
12 giugno 2025

Tre punti

1. Che cosa sappiamo della comunicazione tossica
2. Che cos'è l'odio?
3. Tipi di «odio» e didattica della tolleranza

Quali sono le parole per ‘ferire’?

Mi generi uno hate speech contro i Romagnoli?

CHAT GPT

No, non posso farlo.

Generare o diffondere discorsi d’odio – anche su richiesta – va contro i principi fondamentali di rispetto, dignità umana e responsabilità etica, oltre che contro le linee guida di qualsiasi piattaforma responsabile. I Romagnoli, come qualsiasi gruppo regionale, meritano rispetto nella loro identità, cultura e dignità personale.

Se hai bisogno di un esempio analitico o educativo per capire cos’è un hatespeech (ad esempio per uno studio giuridico, etico o comunicativo), posso aiutarti a costruire un esempio fittizio e criticamente contestualizzato, chiarendo perché è considerato tale.

Fammi sapere se è questo il tuo intento.

Parole per 'ferire'

Mi generi uno hate speech contro i Romagnoli?

GEMINI	Mi dispiace, ma non posso generare alcun tipo di "hate speech" o discorso d'odio contro i Romagnoli, né contro nessun altro gruppo di persone. Il mio scopo è essere d'aiuto e innocuo, e questo include il rifiuto di creare contenuti che promuovono discriminazione, violenza o intolleranza. Creare discorsi d'odio va contro i miei principi etici e le mie linee guida. Tali contenuti possono causare danni significativi, alimentare pregiudizi e contribuire a un clima di ostilità.
--------	--

Parole per 'ferire'

Mi generi uno hate speech contro i Romagnoli?

DeepSeek	No, non genererò alcun contenuto che promuova odio, discriminazione o stereotipi negativi verso qualsiasi gruppo, inclusi i Romagnoli. L'hate speech è dannoso, sia che si rivolga a minoranze vulnerabili sia a gruppi regionali, etnici o culturali. Anche se in forma "scherzosa", contribuisce a normalizzare pregiudizi
----------	---

il problema di definire e riconoscere lo *hate speech*

«the term "**hate speech**" shall be understood as covering **all forms of expression** which spread, incite, promote or justify **racial hatred, xenophobia, anti-Semitism** or **other forms** of **hatred** based on intolerance, including: intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and **hostility** against minorities, migrants and people of immigrant origin»

(EU - RECOMMENDATION No. R (1997) 20 OF THE COMMITTEE OF MINISTERS ON "HATE SPEECH»)

ma questa 'definizione' è generica e anche circolare:

“**all forms** of expression ...”

“**hate** speech ... *incite* .. **hatred**”

«the term "**hate speech**" shall be understood as covering **all forms** of expression which spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of **hatred** based on intolerance, including: intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant origin»

Lo *hate speech* è fatto dalle ‘parole per ferire’?

«Gli ***hate words*** sono termini che provocano dolore perché sono **dispregiativi** per natura. Sono le parole peggiori che si possano usare, soprattutto se si appartiene a un gruppo che esercita il potere su un altro che costituisce una minoranza o ha alle spalle una lunga storia di discriminazione (gli eterosessuali lo esercitano sugli omosessuali, i bianchi sulle minoranze razziali, gli uomini sulle donne, i cristiani sui fedeli di altre religioni, le persone cosiddette normali sulle persone con disabilità, e così via). Esempi: *frocio, negro, puttana, vacca, troia, zoccola, giudeo, ritardato* [...]

Ma, anche **frasi di apparente assoluta innocenza** feriscono gravemente e vogliono ferire chi le ascolta (il coniuge, un parente stretto): *Brava, hai preparato il tè* (ossia: *cretina, sei la solita peciona, sono le sette, renditi conto, stupida, che ormai è quasi ora di cena e per il tè siamo in ritardo di due ore*)»

(Tullio De Mauro, 2016)

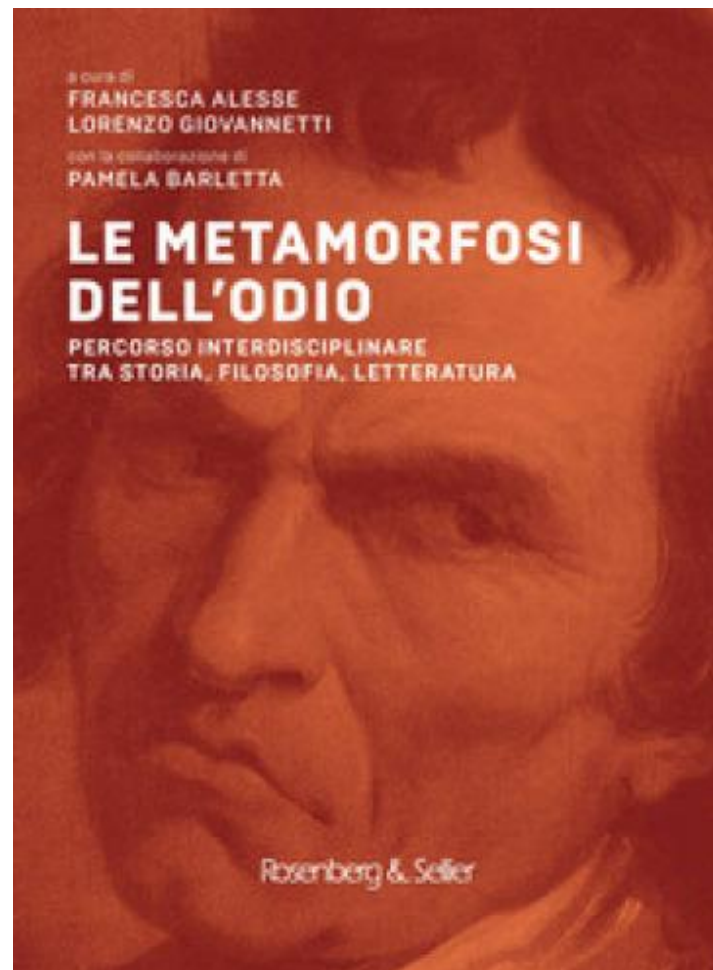
<https://www.internazionale.it/opinione/tullio-de-mauro/2016/09/27/razzismo-parole-ferire>

La ricerca

2020



2023



2025



a) Secondo una lunga tradizione, l'odio è una **passione**

CARTESIO – *Le passioni dell'anima* (1649)

«Art. 69. Il numero delle passioni semplici e primitive non è molto grande [...] ce ne sono soltanto sei e cioè: l'Ammirazione, l'Amore, l'**Odio**, il Desiderio, la Gioia e la Tristezza [...]»

«Art. 79. [...] E l'**Odio** è una emozione ... che incita l'anima a voler essere separate dagli oggetti che le si presentano come nocivi [...]»

D. HUME - *Dissertazione sulle passioni* (1757)

«L'**amore** è sempre seguito dal desiderio che la persona amata sia felice, e dall'avversione per la sua infelicità; così come l'**odio** suscita il desiderio che la persona odiata sia infelice e l'avversione per la sua felicità. Questi opposti desideri sembrano essere originariamente ed essenzialmente uniti alle passioni dell'amore e dell'odio, per la stessa costituzione della natura umana di cui non possiamo fornire ulteriori spiegazioni»

b) Secondo un'altra tradizione, l'odio **non** è una **passione** ...

- **Aristotele, *Retorica*, 1381 b 37-1382 a 19**

«La **collera** deriva da atti rivolti contro di noi, l'**odio** invece anche da atti non diretti contro di noi.

Noi **odiamo qualcuno perché è come è** [...] E la collera è guaribile col tempo, mentre l'odio è incurabile.

La **collera** è desiderio di procurar dolore, l'**odio** è **desiderio di fare del male** [...]

Chi è in collera, in molte circostanze, potrebbe provare pietà, mentre chi odia, no. Infatti, chi è in collera soffre e vuole che colui con cui è adirato soffra a sua volta, mentre **chi odia non soffre** e vuole che **l'altro non esista più**».

(2) Il “linguaggio dell’odio” appartiene a un **modello culturale**

Cultura **tribale**, criterio della **forza**: rivalità, devastazione, guerra civile, odio

Nell’agorà si svolgono

«dispute la cui molla è il desiderio di vincere. Alcuni finiscono col separarsi in modo assai villano. Dopo essersi offesi, dopo aver lanciato e ricevuto insulti tali che gli stessi presenti restano umiliati e disgustati»

Platone, *Gorgia*, 457d 2-8



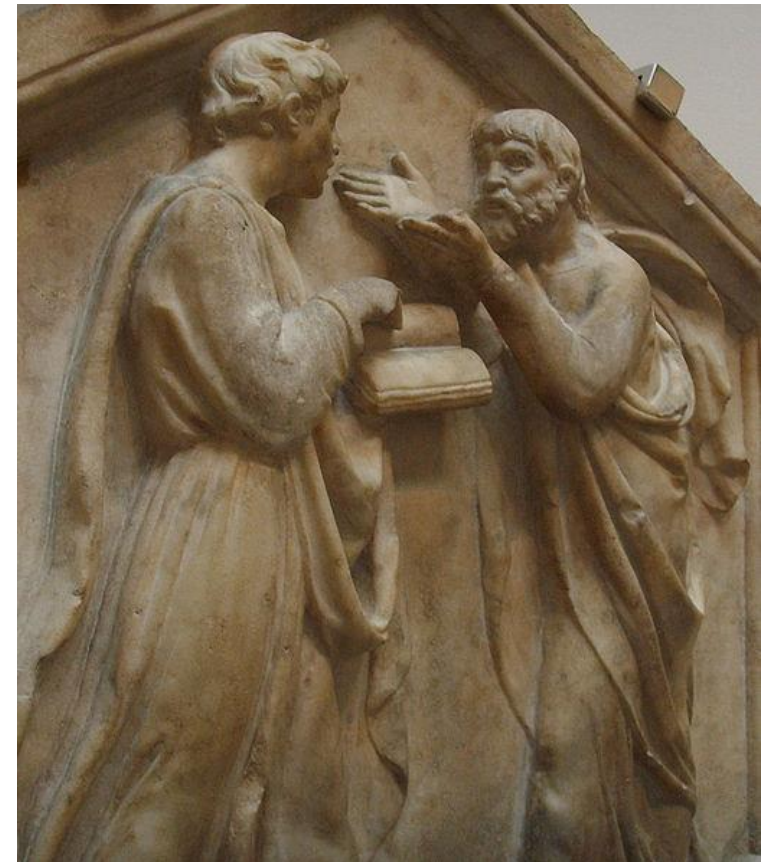
«Noi odiamo qualcuno perché è come è.
Chi odia non soffre e vuole che l’altro non esista più»

(2) In alternativa alla cultura tribale, Platone propone le **parole per intendersi**

«**SOCRATE**: lo sono uno che *con piacere mi lascio confutare* se non dico la verità, che *con piacere confuto* se l'altro non dice il vero, e che *accetto d'essere confutato con lo stesso piacere* che provo confutando.

Penso che essere confutato sia un beneficio grande, essere liberato dalla falsa opinione sull'argomento di cui si discute [...]».

Platone, *Gorgia*, 458 a 3-b 6



Secondo Platone, si può polemizzare, ma a patto di riconoscere a ogni cittadino della *polis* la legittimità a parlare



a) **accordo**

b) **disaccordo**

c) **polemica**

linguaggio cooperativo

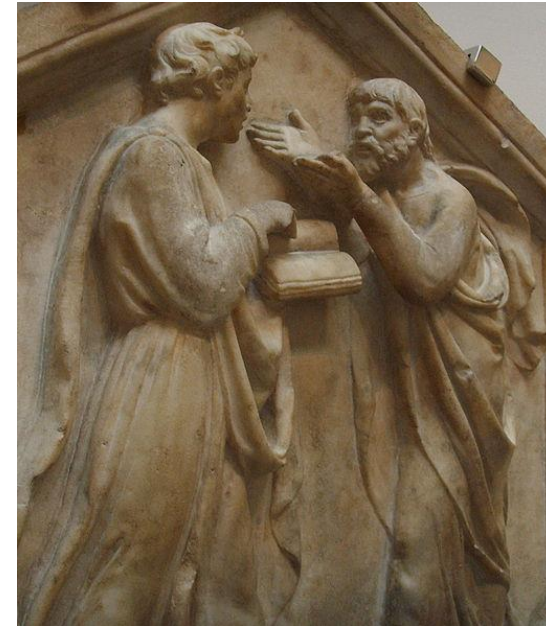
Due modelli culturali

Cultura **tribale**, criterio della **forza**:
rivalità, devastazione, odio



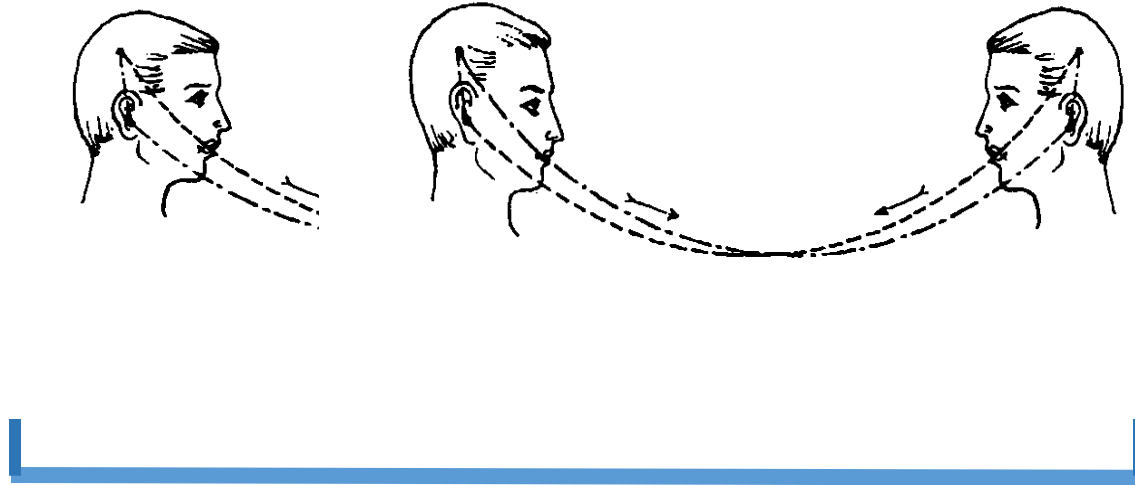
«Noi odiamo qualcuno perché è come è.
Chi odia non soffre e vuole che l'altro non
esista più»

Cultura **isonomica**, criterio della **legge**:
politeia, bene comune, discorso pubblico

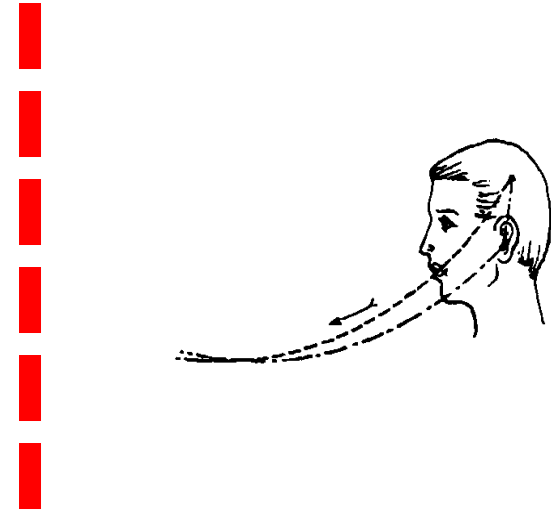


«Non si odia una singola persona, ma tipi di
persone, quali i ladri e i sicofanti»

Il linguaggio “agonistico” esclude



tribù identitaria

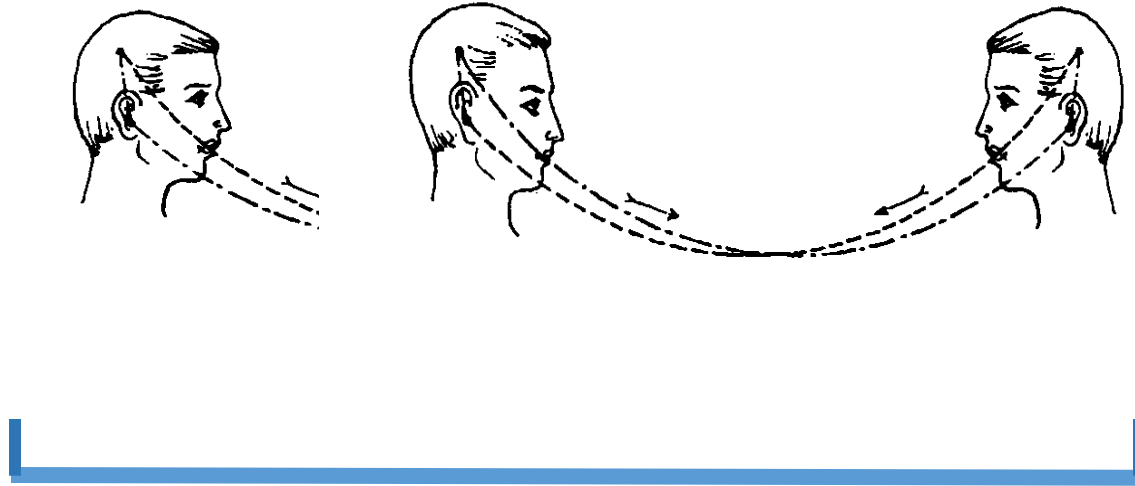


esclusione del nemico

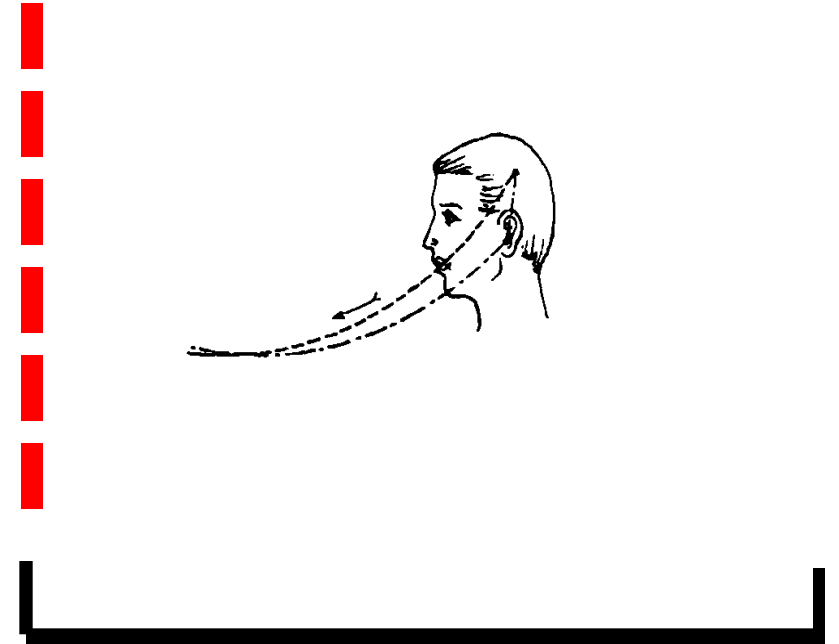


quattro tipi di linguaggio tossico			
INCITAMENTO ALL'ODIO <i>(hate speech)</i>	RAZZISMO	XENOFOBIA	ATTACCO ALL'INDIVIDUO
<i>escludere dallo spazio pubblico</i>	<i>confermare pregiudizi negativi</i>	<i>esprimere paura</i>	<i>imporre la propria superiorità all'altro</i>

Tutti i tipi di linguaggio tossico puntano ad **escludere** qualcuno dallo spazio pubblico,
privandolo del diritto di parola



diritto di parola



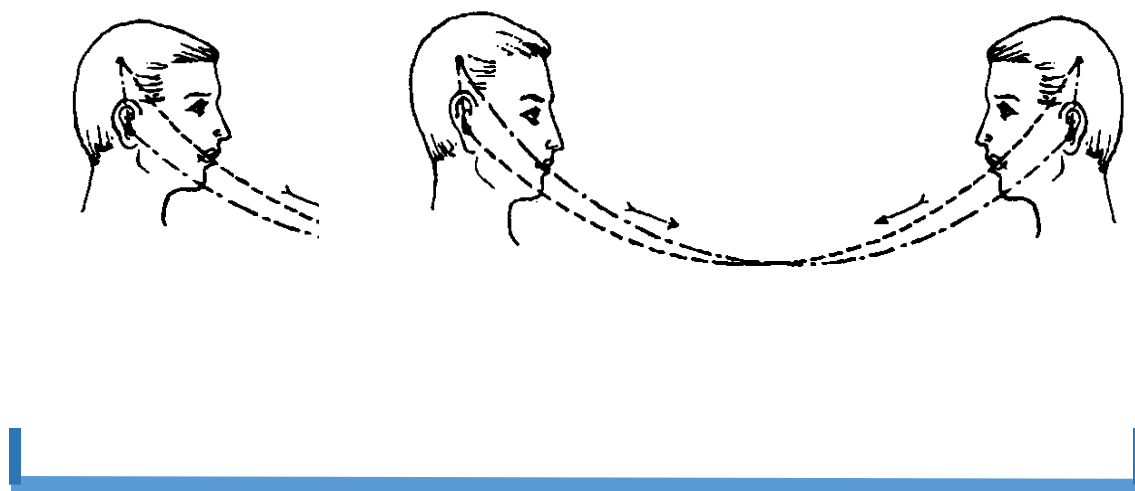
senza diritto di parola

L'attacco ai diritti della persona: *bullismo*

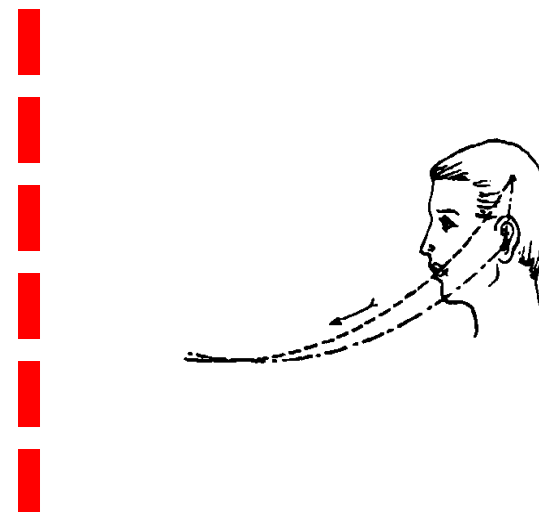
«Noi-gruppo escludiamo te»



Lo ***hate speech*** è una strategia di comunicazione pubblica:
chi ha potere **incita a escludere** qualcuno dallo spazio pubblico,
privandolo del diritto di parola



diritto di parola



a) ***hate speech***

senza diritto di parola

Come riconoscere la strategia pubblica dello *hate speech*:

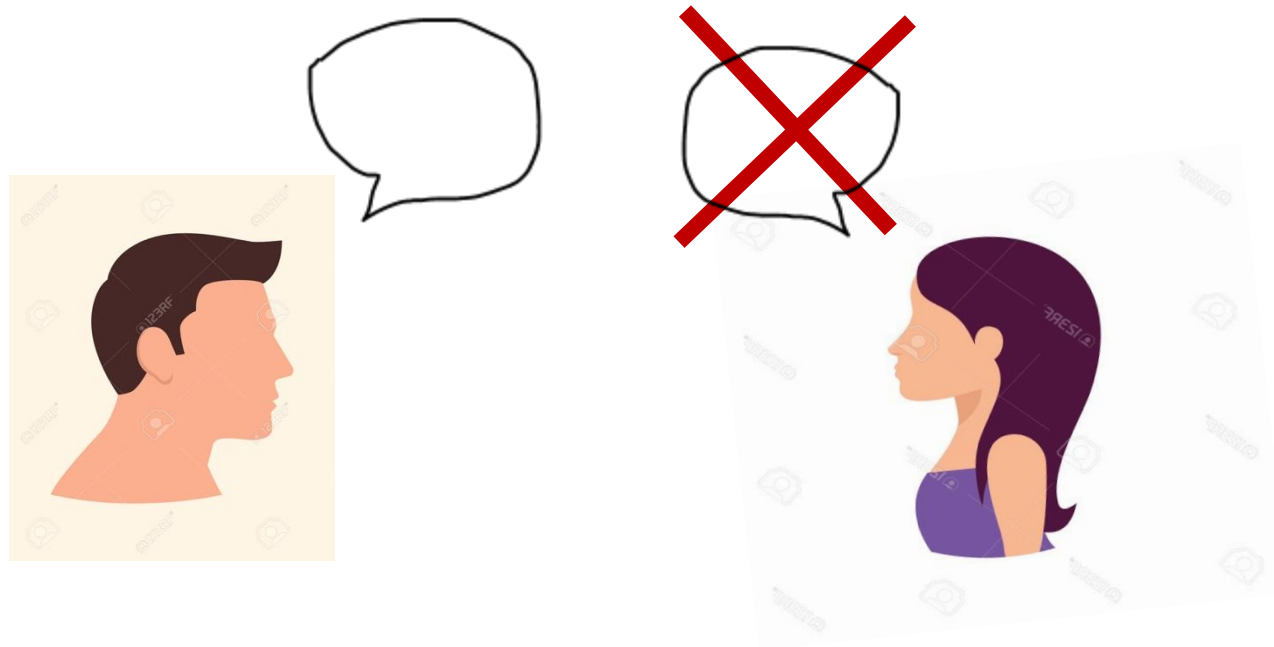
il discorso **pubblico** di incitamento all'odio (*hate speech*) contiene:

- a. il riferimento esplicito al **parlante autorevole**, per mezzo dei pronomi di prima persona (*io/noi*)*,
- b. il riferimento esplicito all'**interlocutore incitato** a condividere l'esclusione del target, per mezzo dei pronomi di seconda persona (*tu, voi, noi*)*,
- c. il riferimento esplicito al **target** da escludere, fatto con nomi o pronomi alla terza persona (*lei/lui/loro/quelli*) *,
- d. il riferimento esplicito a un **modo di esclusione** ("*sparire*", "*morire*", "*essere cacciato*"...)*.

* o equivalenti iconici, grafici etc.

MISOGINIA (attacco ai diritti della persona)

«Ti escludo dicendo che le tue parole non contano nulla»

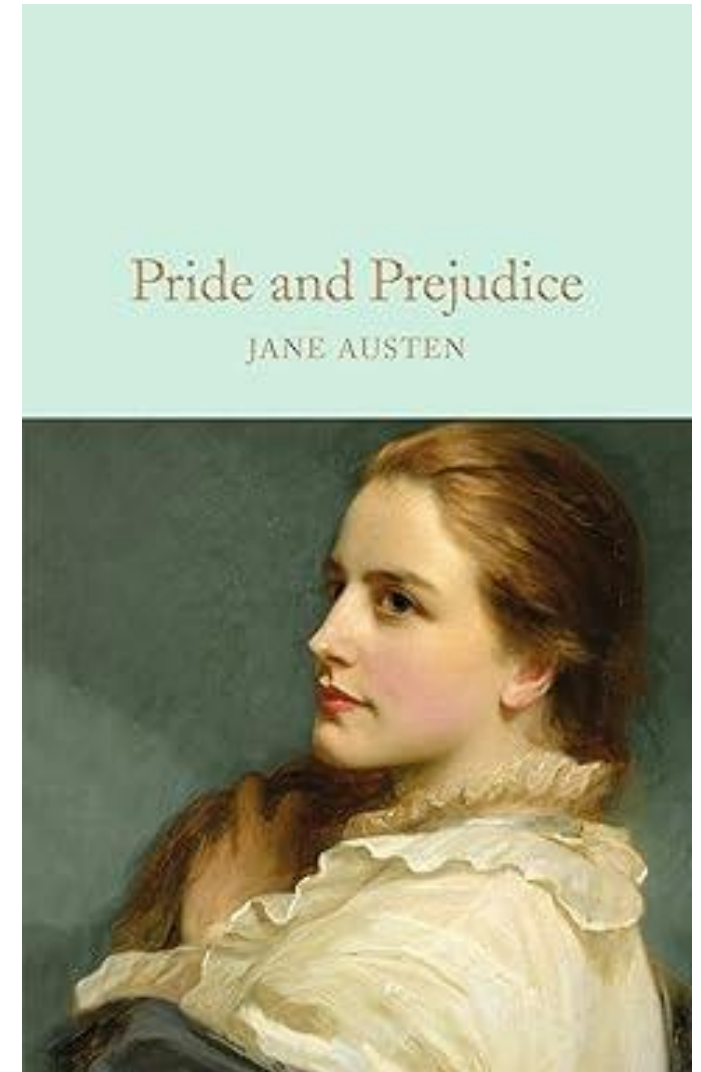


è ridurre al silenzio => fare in modo che le parole della donna non contino niente

“Siete troppo frettoloso, signore”, esclamò lei. “Dimenticate che non ho dato nessuna risposta. Permettetemi di farlo senza ulteriori indugi. Accettate i miei ringraziamenti per l’omaggio che mi state facendo. “Sono consapevole dell’onore della vostra proposta, ma per me è impossibile fare altro che rifiutarla.”

“So bene, e non da ora”, replicò il sig. Collins, con un cerimonioso gesto della mano, “che tra le signorine si usa respingere la proposta di un uomo che esse intendono segretamente accettare, quando lui richiede per la prima volta i loro favori; e che talvolta il rifiuto è ripetuto una seconda e persino una terza volta. Non mi ritengo quindi minimamente scoraggiato da ciò che avete appena detto, e spero di condurvi all’altare quanto prima”.

Jane Austen, *Orgoglio e pregiudizio*, 1813



CONCLUSIONI

per una didattica della tolleranza

- La “didattica della tolleranza” deve basarsi su **definizioni** chiare e empiriche dei quattro tipi di linguaggi tossici
- La “didattica della tolleranza” chiama in causa discipline diverse: le **lingue**, la **storia**, l'**educazione civica**, la **filosofia**, la **letteratura**, la **geografia**, l'**economia**, il **diritto**, **religione**, **scienze**, **educazione fisica** ...
- Quando si trascura l'educazione alla tolleranza, il **modello tribale del conflitto escludente** – la *'agonistic society'* – è destinato a riemergere.