For a Bottom-Up Approach to the Linguistic and Legal Definition of Hate Speech
A Case Study of Offences Against Women

Francesca Ferrucci
Università degli Studi della Tuscia, Viterbo, Italia

Abstract  In addressing the legal definition of Hate Speech (HS), some important linguistic issues arise: the necessity of semantic precision of normative texts conflicts with the difficulty of tracing distinctive features of the phenomenon; the balancing with freedom of expression poses the question on the boundaries between opinions and verbal violence. Comparative studies on national laws show significant gaps in definitions, highlighting their negative effects. This paper discusses the problem analysing a corpus of 2,720 Facebook negative comments against women with decisional positions in Europe. Recurring textual strategies are identified testing a bottom-up approach, argued as the appropriate method to meet the current research challenges.

Keywords  Hate Speech. Woman. Recipient. EU Legislation. Freedom of speech. CMC. Text Analysis.

Summary  1 Legal and Linguistic Definitions of Hate Speech. – 2 Limits of Current Definitions and Challenges for Their Implementation. – 3 A Case Study of Offences against Women. – 3.1 Roles and Actions in the Structures of Narratives. – 3.2 Woman as Unworthy Interlocutor. – 3.3 Delegitimizing Woman’s Opinion and Power. – 3.4 Violent Orders. – 3.5 Downgrading Woman’s Identity to Her Physical Appearance. – 3.6 Downgrading Woman’s Identity to the Family Role. – 4 A Bottom-Up Approach for a Comprehensive Definition.
1 Legal and Linguistic Definitions of Hate Speech

In addressing the legal definition of Hate Speech (henceforth HS), some important linguistic issues arise: there is the need of semantic precision, as requested by normative texts, but it is difficult to find the distinctive features of the phenomenon; its sanction has to be balanced with the principle of freedom of speech, but it is not clear what are the boundaries between opinions and verbal violence, not always corresponding to those between declarative utterances and speech acts (cf. Austin 1962; Searle 1969; Schlieben-Lange 1975). The present article has the purpose to provide a systematic review of the adopted solutions, mainly developed in a jurisdictional perspective, and to discuss them with an interdisciplinary focus that, by integrating inputs from linguistics, psychology and sociolinguistics, can produce an advancement with respect to the state of the art. An overview of the current definitions is first provided, highlighting the different used parameters and some of their limits; the challenges for their implementation are then identified, analysing documented HS events that hardly fit into the sanctioning perimeters. A case study, based on the HS against women with public decisional roles in Italy and in Europe, is finally presented, with the aim to experiment a bottom-up approach in the definition of the phenomenon, argued as the appropriate method to meet the various manifestations of discrimination and to achieve more comprehensive defining criteria.

Comparative studies on national laws show significant gaps in HS definitions and highlight their negative effects: if, on the one hand, the diversity of cultures in which norms are produced must be taken into account, on the other hand the lack of neatly defined criteria can reduce the effectiveness of the laws themselves, it can cause confusion in the judgment of single cases and produce an abuse of the term in the public debate, hindering deterrence.

While many hate-based acts are spelled out and subsequently criminalized within national Criminal Codes, strict definitions of hate crime and hate speech have generally not been delineated. (PRISM Project 2015, 48)

However, it appears very important to have a common and consistent definition at the international level, in order to counter the HS in all media, particularly those of Computer-Mediated Communication (henceforth CMC), where offences may become a viral phenomenon beyond single local communities (ELSA 2014, 291).

---

1 See also Danish Institute for Human Rights 2017.
2 On the semiotics of the new media see Cosenza 2014; Manetti 2008.
If we consider the main circulating definitions of HS, the most frequent defining parameters are pragmatic: first of all, the sender’s intention (hurting, offending) or motivation. For example, definitions in Croatia and Netherlands are respectively:

hate crime is a crime committed because of race, colour, religion, national or ethnic origin, disability, gender, sexual orientation or gender identity of another person.

Offences with a discriminatory background. (PRISM Project 2015, 48)

An overview of the trends in many European countries is available, in the same study by the PRISM European Project, referring to Bulgaria, Belgium, Croatia, Cyprus, Czech Republic, Finland, Germany, Greece, Ireland, Italy, Latvia, Lithuania, Malta, Poland, Portugal, Romania, Spain and Sweden, where the key concept is “bias”:

All of the 18 responding countries specified certain bias categories in their legislation, which help to identify segments of society that may be particularly targeted in acts of discrimination, hate crime and hate speech. Every respondent mentioned variations of the terms race, religion, ethnic origin, nationality, and sexual orientation [...]. Other commonly mentioned categories include age, disability, and language. (PRISM Project 2015, 49)

UK’s definition considers the different channels where HS may occur and highlights the perlocutionary effects on the recipient (incitement to hatred, intimidation):

hate speech is any speech, gesture or conduct, writing, or display which is forbidden because it may incite violence or prejudicial action against or by a protected individual or group, or because it disparages or intimidates a protected individual or group. The law may identify a protected individual or a protected group by certain characteristics. (ELSA 2014, 269)

In Italy, the bias categories mentioned in laws are sex, race, language, religion, ethnicity, nationality, political opinion, personal and social conditions; in this framework, it is worth noting the exclusion of LGBT and disabled people from the Mancino Law against Hate crimes (Italian Constitution 1948, artt. 3, 8, 19; Law no. 645/1952; Law no. 654/1975; Legislative Decree 122/1993; Law no. 205 of 1993; Legislative Decree 286/1998; Legislative Decrees 70/2003, 215/2003 and 216/2003; Law no. 85/2006; Legislative Decree 198/2006; Law no. 101/2008; Legislative Decree 5/2010; Civil Code artt. 1418, Penal Code artt. 302, 402-5, 594-5, 724) (cf. PRISM Project 2015, 142-76).
What is meant by “protected individual or group” is better explained in the definition by OSCE (Organization for Security and Co-operation in Europe), that is used as a working basis by some European States. It describes hate crime as:

- a criminal act motivated by bias towards a certain group. For a criminal act to qualify as a hate crime, it must meet two criteria:
  - The act must be a crime under the criminal code of the legal jurisdiction in which it is committed;
  - The crime must have been committed with a bias motivation.
- “Bias motivation means that the perpetrator chose the target of the crime on the basis of protected characteristics. A “protected characteristic” is a fundamental or core characteristic that is shared by a group, such as race, religion, ethnicity, language or sexual orientation. (PRISM Project 2015, 48)”

Another important definition is the one by the Council of Europe:

- 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, migrant and people of immigrant origin. (cit. in PRISM Project 2015, 10)

All these definitions refer to what, in the pragmatic theory of language, is considered a fundamental aspect of communication, including motivations, intentions, attitudes and feelings of sender and recipient; but they leave unsolved the problem of how to determine a priori the motivations of haters and emotional reactions of hate, in the perspective of sanction and criminalization.

Another parameter for crime definition is logic-semantic: the truth-value of an utterance. If the offences are false, then they must be punished.

Whoever, with the exception of the cases indicated in the previous article, by communicating with more people, offends the reputation of others, is punished with imprisonment for up to one year or a fine of up to one thousand thirty-two euros. If the offense con-

---

4 The text is a model for many HS definitions in study groups and commissions: see for example the Report by Italy’s Chamber of Representatives, where the key concept is the core characteristic of a person as the basis of the hatred (cf. Camera dei Deputati 2017).

5 Recommendation (97)20 of the Council of Europe, 30 October 1997.
sists in attributing a specific fact, the penalty is imprisonment of up to two years, or a fine of up to two thousand sixty-five euros.6

This strategy for determining what is sanctionable is significantly different from the previous ones, as a violent offence may be ‘technically’ true, for example when denigrating persons who actually belong to a religious group presented as ‘undesirable’. It seems to overlap only partially with the HS.

Finally, a strategy frequently implemented in HS definitions is to list the victims, as also shown by the Council of Europe’s text mentioned above and in the norm holding in Austria:

whoever incites or publicly urges to violence in a manner that is likely to endanger the public order, or in a way perceivable by the general public, against a church or a religious society, or against a group of people defined according to the criteria of race, colour, language, religion or belief, nationality, descent or national or ethnic origin, sex, disability, age or sexual orientation or against the member of such a group explicitly because of its affiliation, shall be punished with an imprisonment up to two years. (ELSA 2014, 10)

The reference to a “group of people” or a “member of such a group explicitly because of its affiliation” is another choice that, relying on the concept of “protected characteristics”, presents the risk of overgeneralizing the HS cases.

2 Limits of Current Definitions and Challenges for Their Implementation

While catching the main aspects of the phenomenon, the list of victims can improperly leave out many episodes targeting single persons, where their belonging to a social group is not recognizable as a triggering factor of the HS. Mounting evidence shows HS against VIPs, teachers, Healthcare Professionals, public officers and in general against those considered ‘counterparts’ in specific situations.7

6 Original text: “Chiunque, fuori dei casi indicati nell’articolo precedente, comunicando con più persone, offende l’altrui reputazione, è punito con la reclusione fino a un anno o con la multa fino a milletrentadue euro. Se l’offesa consiste nell’attribuzione di un fatto determinato, la pena è della reclusione fino a due anni, ovvero della multa fino duemilasessantacinque euro” (Italy’s Penal Code art. 595, “Diffamazione”).

7 Women are paradigmatic of this phenomenon, being often targeted by HS with no explicit reference to the categories of sex or gender: it is then difficult to sanction it as an act against a member of group explicitly because of its affiliation, or on a basis
with targets and contexts that are difficult to classify for regulatory purposes. Thus, one limit of current HS definitions is an insufficient ‘descriptive adequacy’: they do not represent all HS events, producing false negatives (that is, events that are not included in the range of a targeted phenomenon although they should be). Should the laws mention all types of victims, the lists could be very long, and it could be very difficult to specify all circumstances where HS happens. On the other hand, where definitions rely on the pragmatic strength of HS, its potential to hurt, offend, intimidate or incite, we face the opposite risk of false positives (that is, events that are included in the range of a targeted phenomenon although they should not). For example, it is possible to hurt someone with strong criticism, without resulting in violence. This is obviously linked to the issue of freedom of speech as a fundamental right that must be ensured. In addition, it is not clear how the pragmatic strength can be objectively measured and based on distinctive features, thus meeting the need of a ‘strict definition’, as recommended by all relevant stakeholders in this field.

of protected characteristics or bias considered in laws. The corpus analysed in this paper includes many of these cases, focusing on women who play apical decisional roles; but there are also women in other roles who have to face HS in their day-to-day life, with less possibility to count on normative (and cultural) representation of these abuses. In Italy, for example, we can mention the recent experience by Mrs. Carola Rackete, the captain of the rescue ship Sea-Watch 3 of a NGO, who during the summer 2019 defied the Italian government ban and docked at Lampedusa for humanitarian reasons, after having saved many migrants: she has been strongly attacked by some internet users, but she does not match the types of victims used to define hate crimes. Evidences of HS against some workers’ categories, such employees in public services, are well documented in scientific papers: cf. for healthcare professionals Kapoor 2017. Among them, women are the majority of the targets; see also De Mauro 2016; Ferrucci 2019.

In Figure 1, false negatives are represented by the light blue areas in the horizontal axis, where the events of “general circulating HS” and/or “face-to-face HS” are not covered by the defining parameters; false positives are instead represented by the light blue areas in the vertical axis, where the pragmatic, semantic or logic features can be related to non-violent speech acts. Conversely, the overlapping areas in dark blue represent the descriptive adequacy, that is reached only in some cases. HS events are commonly analysed in literature distinguishing between general circulating HS (in media, public spaces, including cases where victims do not participate to the interaction) and face-to-face communication (cf. Gumperz, Hymes 1972).
In the following section, a case study will be presented, regarding offences against women in CMC. We will consider many examples of violent utterances that do not fit into the current HS definitions and suggest more comprehensive criteria. The framework of the research is the assumption that HS produces severe harms on the victims: general consequences reported in literature are feelings of fear, silencing, provocation to anger and frustration, restriction of ability to identify with the group’s characteristics and of identity-building processes. HS may lead to an overall disempowerment of the targeted persons; repetitive HS, if performed either in face-to-face communication or in general circulating mass media, may be traumatic: a ‘before’ and an ‘after’ of the event clearly separate the victim’s interior experience. So we have a twofold challenge.

First, we consider all types of HS events, even less frequent or less visible ones, starting from a collection of actual HS events that can represent the variety of the phenomenon. Only in this way we can reach a complete picture of what is to be defined in regulatory texts. Second, we have to consider the point of view of the recipient. That’s exactly the right perspective that should be adopted, mentioning the Australian law as a positive example:

> [t]o the extent that hate speech laws are sufficiently broad and flexible to address the types of speech events and spectrum of harms attested to by the interviewees, and to assess harm from the perspective of the targeted communities, members of Indigenous and racial/

---

9 The literature distinguishes between constitutive and consequential harms (Maitra, McGowan 2012, 6); namely, between harms that are occasioned in the saying of a HS event, and harms that occur as a result of it. See also Gelber, McNamara 2016: the authors show that both types of HS events represented in Figure 1 can incur harm constitutively, consequentially or simultaneously in both ways.
Research is needed to address the issue of descriptive adequacy, as it arises from the state of the art, also taking into account the ethical need to provide an answer, even in cultural terms, to all manifestations of this type of violence. For the academic community in communication sciences, study questions may be spelled out in the following ways: is it possible to clarify the distinctive features of HS, distinguishing it from impoliteness, lack of cooperation and legitimate criticism? Under which conditions it becomes impossible to maintain a common field of semantic negotiation? Which textual strategies, appropriately generalized, can constitute the basis for a definition that should be at the same time more rigorous and broader, therefore more effective also for sanctions?

3 A Case Study of Offences against Women

A corpus of 2720 negative comments in Facebook has been examined, for a total of 106,816 tokens, browsing the profiles of 15 women who, at the moment of the events (2017-19), had public decisional roles in Italy and in Europe. The target women were chosen on the basis of the criteria of visibility (members of national Parliaments or Governments or international Boards, mayors of capitals), country and institutional representativity (4 European countries, UK, 1 European Institution and 1 worldwide institution), primary language of the Facebook Profile (Italian, English, French, Spanish, German) and political representativity (affiliation to different parties or movements). The comments were manually collected and then labelled by name of the target women, role at the moment of the HS event, year of occurrence and textual characters: no text-mining software was used, also with the intention to catch HS events where violence arises at a level wider than single lexical items or utterances, such as in the narrative structures. In a final validation process, all the comments were...
ments were checked, in order to standardize the labels regarding the HS textual characters and remove those that, even with strong criticism, did not include any of them. The most frequent identified characters are discussed in the following and shown in the examples.\(^\text{13}\)

### 3.1 Roles and Actions in the Structures of Narratives

In 49.1% of the comments, the hater represents a narrative where the attacked woman is the victim of a (sexual) violence:

1. \textit{Anche se ho quasi 25 anni chiedo un regalo a Babbo Natale... Per Natale voglio stare chiuso in una stanza con te, soli, tu ed io... Solo noi e la mia accetta. Partirei con il taglio delle mani prima.}  
2. \textit{Du hast sie alle rein geholt und jetzt werden unsere Kinder von diesen Leuten vergewaltigt und ermordet.... das sollten sie mal mit deine Kinder machen... Ach nee du hast ja keine oder sollen sie es mal mit dir machen du würdest dich vielleicht noch freuen darüber.... das dich überhaupt einer nimmt. Was hast du aus unseren Land gemacht.}

Less frequently (4.4%), but interestingly, the hater assigns to the attacked woman also the role of killer or criminal, suggesting, explicitly or implicitly, the need to punish her:

3. \textit{No le he visto yo la persona tan inútil y tan enferma mental como tú cierras trato con musulmanes que te tienen por una furcia haces tratos con independistas siendo uno de ellos recuerda que todo esto pagarás la propia nación te dará lo que te tiene que dar.}

\[^\text{13}\] The comments are quoted in the original language, which is mostly related to the primary language of the Facebook Profile and to the geographic location of both the hater and the hated; only in the case of target women belonging to international boards, the audience is wider and multilingual. Here, the comment seems to be influenced by the sender’s native language or, in a minority of events, his/her choice in favour of a vehicular international language such as English. In the subheadings, for explanatory reasons the textual characters are discussed separately but frequently they occur together in the same comment: for instance, the woman is represented as crazy (unworthy interlocutor) and at the same time as a dangerous criminal (role in the structure of narrative): see example no. 3. In this case the comment has been counted as occurrence of both textual characters. Thus the given percentages are to be considered as overlapping sets.
4. Madame, ne citez plus jamais Simone Veil, vous êtes loin de lui arriver à la cheville. Cette grande dame à fait beaucoup pour les femmes, vous, vous voulez tuer tout le monde !

3.2 Woman as Unworthy Interlocutor

In 42.1% of the comments the victim is attacked with injurious qualifications and presented as unworthy interlocutor (as consequence of being crazy, mentally ill or similar):

5. Du irre Oma, du bist verwirrt und hast keine Ahnung vom normales Leben....

3.3 Delegitimizing Woman’s Opinion and Power

In 30.6% of the comments the woman’s opinion or public role is delegitimized:

8. Prima fa le “valige” meglio è per tutti.....Prenda atto della situazione e cambi mestiere...
9. when are you gonna pay the money back with your very poor italy..you are nothing without this EU, go home we dont want you and we dont need you...and peace in europe dream on, you have bring the extreme violence into the eu with your open borders..
10. ma ancora parla questa????????ma chi sei, chi te vole ??????

3.4 Violent Orders

In 26.8% of the comments the hater gives violent orders to the attacked woman:

11. sei una puttana andicappata vattene a casa hai la cosa giusta x una volta vaiiiiiiiii viaaaaaa
12. Tu es très vieux, demande Euthanasia !...et votre partenaire, même que.
13. Niemand kann sie mehr ertragen, nicht mal ihre eigenen Leute. Also hauen sie endlich ab!
3.5 **Downgrading Woman’s Identity to Her Physical Appearance**

In 8.2% of the comments the hater presents the woman’s identity referring to limits of her body and/or her (bad) physical appearance:

14. *patética como siempre. Ni sí ni no, ni blanco ni negro...esa solo sabe hacer pipí en mitad de la calle...*
15. *SEI PIÙ BRUTTA E VECCHIA DEI LANZICHENECCHI*
16. *Les mercenaires sont a la tête de l’état ou plutôt la mafia bourgeoise, vous les “prout prout” de la société. Avoir autant de mépris d’arrogance et de dédain pour les autres, c’est pitoyable et indigne de votre fonction, rappelez vous que votre merde de président n’est pas élu a la majorité des français. Votre classe détruit allègrement la planete en pronant le capitalisme et vous donnez des lecons. Heureusement que l’on voit pas vos cheville ca doit etre affrexu a voir tellement elle doivent etre gonflée. En bref je vous conchie ! (PS: changer de coiffure serieux on est au 21eme siecle)*

3.6 **Downgrading Woman’s Identity to the Family Role**

In 2.1% of the comments the hater presents the woman’s identity referring to her traditional role at home and within the family:

17. *Incroyable de se sentir encore indispensable quand on a échoué lamentablement, c’est bien français. On en veux plus de la vieille politique, on a vu les ravages de votre politique en Poitou Charentes et au ministère de l’écologie, on paye encore les portiques... Stop, il est temps pour vous de vous occuper de vos petits enfants.*

4 **A Bottom-Up Approach for a Comprehensive Definition**

The Facebook comments do not meet the parameters used in current HS definitions and at the same time can hardly be explained with the categories of impoliteness, lack of cooperation or criticism. They are violent, to the extent that they contain the sender’s and recipient’s simulacra\(^\text{14}\) that are incompatible with the maintenance of a common

---

\(^{14}\) The ‘traces’ of the subjectivities of Sender and Recipient in written texts.
field of semantic negotiation. The attacked woman cannot even an-
swer to the comment, as the answer itself would legitimate the nul-
lification of her image that already took place in the interaction. In
this sense, the comments are qualitatively different from manifes-
tations of opinions, even when strongly conflictive. In the corpus we
found, the textual features that show a common strategy are: destroy
the ground of the communication, demolish its components and on-
ly leave the victim the possibilities of either taking note of the dem-
olition, or to rejecting the comment as a whole, for example by eras-
ing or ‘censoring’ it.

19. what democracy you idiot? Europe has no freedom of speech.
anyone who speaks up against your agenda gets in real trou-
ble. People like me can say things because we have this thing
called 2nd Amendment in our country and we have real rights
and freedoms unlike all the serfs you are lording it over.

The principle of Freedom of Speech, maximally protected in the US
Constitution, explicitly conflicts with the opposite need to engage
with the recipients’ point of view. That is the present challenge of
the legal systems in the European area.

In facing this complexity, we adopted a bottom-up approach, start-
ing from a sample of real data and measuring ‘upon the field’ the de-
scriptive adequacy of possible generalizations. Based on women’s
perspective, HS appears as a speech that:
• narratively represents an act of illegal violence or coercion
against the woman;
• represents the woman in a humiliating way, also by qualifying
her as unworthy of interlocution;
• delegitimizes the woman’s right to express her ideas and/or to
exercise decision-making power;
• gives violent orders to the woman;
• a speech that downgrades the woman’s identity to her physical
appearance or family traditional role.

This method, which uses the categories of linguistics and text anal-
ysis to describe specific coding modes, could be replicated on oth-
er corpora, with the advantage to separate HS definition from ab-
stract categories of victims and hatred and to address multiple and
intersectional discrimination, an emerging feature of hate crimes
(cf. eMORE Project 2018).
Bibliography


