PERFORMATIVE HATE SPEECH ACTS.

PERLOCUTIONARY AND ILLOCUTIONARY UNDERSTANDINGS IN INTERNATIONAL HUMAN RIGHTS LAW

ALESSANDRO DI ROSA[1]

Abstract: The first part of this work analyses the concept of hate speech and its legal-philosophical foundations linked to freedom of speech, through the use of tools provided by current trends in the theory of performativity. The second part, in turn, aims to suggest two possible perspectives on the translation of these philosophical demands into positive legislation within human rights law: the first one based on a liberal conception of freedom as non-interference and a perlocutionary understanding of performative speech acts; the second one adopting a neo-republican interpretation of freedom as non-domination and an illocutionary understanding of speech acts. Finally, the work aims to critically sift through the application of the theory of performativity to the legal definitions that hate speech has acquired within this context.

Keywords: hate speech; freedom of speech; performativity; non-domination.

Summary: I. *HATE SPEECH*: DEFINING ISSUES; II. FROM SPEECH ACTS THEORY TO CONTEMPORARY THEORY OF PERFORMATIVITY; III. HATE SPEECH: *PERLOCUTIONARY* OR *ILLOCUTIONARY*?; III.1. A liberal model of freedom, human rights and performativity: performatives as perlocutionary speech-plus acts; III.2. A neo-republican model of freedom, human rights and performativity: performatives as illocutionary speech acts; III.3. Perlocutionary and illocutionary accounts: drawing a line between criminal and civil sanctions.

I. *HATE SPEECH*: DEFINING ISSUES

This work aims to analyse feasible applications of the theory of performativity to the various definitions that hate speech has acquired under international human rights law. Within this framework, a structure coherent with the dualist human rights analysis method (PECES-BARBA MARTÍNEZ 1995; ANSUÁTEGUI ROIG 2002; BARRANCO AVILÉS 2009; DE ASÍS ROIG 2010) will be followed. This implies the need to analyse the philosophical issues underlying hate speech and freedom of expression, on the one hand, and their reflection in positive law, on the other. Therefore, this paragraph will be dedicated to the analysis of the concept of hate speech.

The term '*hate speech*' indicates a legal category that originates from a reflection on the nature of speech and action in the US case-law, at the beginning of the 20th century, and which has recently spread to Europe (ZICCARDI 2016, 11-59). However, the main issue is to be found in the difficulty in drawing a definition, insofar as hate speech is

[1] Alessandro DI ROSA, PhD Candidate in *Law (Philosophy of Law)* and *Advanced Studies in Human Rights*, Double title co-tutorship PhD Programme, University of Parma - Universidad Carlos III de Madrid, Spain (alessandro.dirosa@unipr.it).
I would like to express all my gratitude to my friend Julian D. Grass for the mothertongue English copy-editing of this paper.

usually intended as a set of non-homogeneous behaviours (PAREKH 2012, 40). Indeed, there is no unique and universal definition[2], partly because it is the result of complicated legislative activity and the elaboration of international normative standards, which have materialised in the adoption of several declarations contributing to – and competing in – its understanding[3]. A useful research methodology to obtain a first draft of the definition of hate speech could be that of a historical approach observing how this concept has progressively emerged within the international normative context (see FARROR 1996; GASCÓN CUENCA 2016).

Without it being possible to dwell on this analytically, it is essential to recall the Universal Declaration of Human Rights (1948), in particular: Art. 7 (protection from discrimination and incitement to discrimination); Art. 19 (freedom of expression); Art. 29 (statutory reservation and respect for others' rights); and Art. 30 (prohibition of the abuse of right). These articles bear witness to the commitment, at an international level, to prevent the abuse of freedom of expression.

For the purpose of the definition of hate speech, the International Covenant on Civil and Political Rights (1966) is of paramount importance, with its Art. 20.2 setting forth a prohibition of «any advocacy of national, racial and religious hatred which constitutes incitement to discrimination, hostility or violence». Furthermore, it is worth underlining that the Covenant makes it mandatory for Member States to prohibit hate speech not only in public but also in private.

Moreover, the Convention on the Elimination of all forms of Racial Discrimination makes a distinction between racial discrimination (Art. 1) and racial hate speech (Art. 4), the latter defined as a «dissemination of ideas of racial superiority or justification or advocacy of racial hatred or discrimination». The Convention requires the establishment of criminal offences of racial hate speech by all Member States, not only for incitement but also for sheer dissemination of ideas of racial superiority[4]. Further, the Committee established by the Convention (CERD) defines hate speech in terms of a genuine abuse of freedom of expression[5].

At a regional-European level, it is worth considering the contribution of the Council of Europe in defining hate speech. The European Convention on Human Rights does

[2] UN Human Rights Council, 28th session, Report of the Special Rapporteur on minority issues, 5 January 2015, A/HRC/28/64.

[3] For an overview of the evolutionary trajectory of the treatment of hate speech, see MCHANGAMA 2015.

[4] However, it is worth underlying that some States have refused to interpret the locution «an offence punishable by law» in terms of the need for *criminal* sanctions (not only civil or administrative); anyway, to such an issue has answered the Committee in the 1985 report on Art. 4, entitled *Positive Measures Designed to Eradicate all Incitement to, or Acts of Racial Discrimination, Implementation of the International Convention on the Elimination of All Forms of Racial Discrimination*, U.N. Doc. CERD/2, 1985, stating that «the imposition of civil liability falls short of the requirement of Article 4 to declare certain acts or activities as an offence punishable by law», p. 185 (FARROR 1996, 52).

[5] Report of the Committee on the Elimination of Racial Discrimination, U.N. GAOR, 42nd session, sup. No. 18, p. 69, U.N. Doc. A/42/18, 1987.

not refer expressely to hate speech; however, the European Court of Human Rights, stemming from Art. 10 (freedom of expression) and Art. 17 (prohibition of the abuse of rights), has easily imposed limits on freedom of expression when it materialises in acts of hate speech (FARROR 1996, 62 e 76-77; see CARUSO 2011 e 2017; SPIGNO 2018, 473), regardless of it being in public or in private.

Council of Europe Recommendation No. 20 of October 30th 1997 defines hate speech as «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»[6]. The provision ends with an *inclusive formula*, adding «other forms of hatred» as possible situations disseminating forms of discrimination. This allows the claim that hate speech has an *open definition* subject to actual situation analysis and possible further extensions on grounds of new discriminations arising against specific minorities[7], such as those based on sexual orientation (see WEBER 2009, 4). Besides, Principle No. 2 of the Recommendation requires that governments should adopt *civil*, *criminal* and *administrative* legal measures, which could facilitate the judiciary to the necessary balancing between freedom of expression with human dignity and the protection of others' reputation and rights.

When it comes to the European Union, the EU Charter of fundamental rights (Nizza Charter, 2000-2007) also contains a rather broad notion of discrimination, referring to categories such as «sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation» (Art. 21). Thus, a potential notion of hate speech emerging from the European Union provisions would meet with a greater number of definitions based on categories which integrate the definition of discrimination itself, but within the limits of the lack of the "open definition", the latter emerging from Council of Europe Recommendation No. 20. Nevertheless, it is a provision that would appear to confirm the *plurality* of definitions that hate speech has.

In 2008, the Council of the European Union issued a Framework Decision against *racism* and *xenophobia*, requiring that all EU Member States *criminalise* conducts intended for public incitement of hatred against a group (or a member of that group) identified on the basis of ethnical belonging, skin colour, national or ethnical origin[8]. Furthermore, in its 2015 Resolution on the situation of fundamental rights within the European Union, the European Parliament explicitly refers to hate speech based on prejudice on grounds of disability, sexual orientation or gender identity[9].

[6] Council of Europe, Committee of Ministers, Recommendation No. R (97) 20, *of the Committee of Ministers to Member States on "hate speech"*, adopted by the Committee of Ministers on 30 October 1997 at the 607th meeting of the Ministers' Deputies, Scope, Section 2.

[7] The UN Committee on the Elimination of Racial Discrimination recognised that within the Council of Europe the definition of hate speech is much broader: see UN Human Rights Council, 28th session, Report of the Special Rapporteur on minority issues, p. 12.

[8] Council Framework Decision 2008/913/JAI of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law.

[9] European Parliament, *Situation of fundamental rights in the EU (2013-2014)*, Resolution C 316/3, n. 107.

A general definition of hate speech, based on the outcome of international and regional provisions, could therefore be as proposed by Alexander Brown (BROWN 2015, 4-5; italics added):

> «At any rate, the overall impression created by these characterizations is of *speech or other expressive conduct* that is in some sense *intimately connected with hatred of members of groups or classes of persons identified by certain ascriptive characteristics* (e.g.s race, ethnicity, nationality, citizenship, origin of birth, war record, religion, sexual orientation, gender or transgender identity, disability, age, physical appearance), where this connection is exemplified by familiar tropes *relating to hatred in the motive, content, or effect of the relevant speech or other expressive conduct*».

However, given the lack of unambiguous definitions at an international level, hate speech can be fully understood only through the analysis of the legal-philosophical debate (see HEINZE 2016). Furthermore, as will be shown, two significantly different approaches to the definition of hate speech stemming from ICCPR and ICERD can be outlined on grounds of the theory of performativity (see §§ 2, 3). The aim of the following section, therefore, is precisely that of identifying possible tools in the theory to analyse the treatment of hate speech adopted by international human rights law. Those tools will then be put into practice both with a liberal and a neo-republican uderstanding of freedom, in order to show the different outcomes that they could lead to.

II. FROM SPEECH ACTS THEORY TO CONTEMPORARY THEORY OF PERFORMATIVITY

A crucial issue related to the definition of hate speech is that of the means by which the hateful message is spread. Indeed, hate speech is traditionally shared by means of expressive conducts which involve, but are not limited to, the use of voice. Nevertheless, within the hate speech debate, a distinction between *speech* and *action* appears to lie behind the various positions authors have taken – at times explicitly, at times implicitly. This work will attempt to show how the theory of performativity could be useful in deconstructing the *speech* v. *action* distinction (BROWN 2015, 14, footnote; KOLTAY 2013, 27; *contra*: BARENDT 2007, 78-81) within a neo-republican understanding of freedom as non-domination.

The philosophical understandings of the distinctions between *speech* and *action* are quite variegated and underline the different approaches employed to endorse opposite solutions to hate speech issues. As Alexander Brown highlights, it is possible to identify two clusters of arguments aiming to discern "pure" speech from "other kinds" of speech (BROWN 2015, 14-15 footnote):

1) One cluster examines the distinction between «*pure speech* or *mere speech*» and «*speech-conduct* or *speech-plus*»[10]. When hate speech bans prohibit the latter, one

[10] A good example of this is the *fighting words doctrine*, among the doctrines on the First Amendment to the US Constitution which is used by the Supreme Court starting from the case *yinsky v. New Hampshire*,

does not question their compatibility with the constitutional guarantee of freedom of expression, which can be limited provided that it constitutes "something more" than mere speech (such a distinction, to my mind, lies behind the observations of those liberal thinkers who allow said limitation in some cases[11]);

2) the second cluster deals with the impossibility to differentiate between "pure" speech and "other kinds" of speech, which implies that every kind of speech is also an expressive conduct; nevertheless there are lawful and unlawful expressive conducts, the latter of which do not fall within the protection of freedom of expression (see SUNSTEIN 1993, 125; FISH 1994, 124-126; MARCUS 2008, 1025-1059).

Regardless of the strategy adopted, both could lead either to any action falling within freedom of expression, or to the opposite solution: that is, that most discursive expressions cannot be told apart from actions and cannot fall within free speech classical logic (GREENAWALT 1989a, 79-163).

For this reason, it is useful to conduct a study on the *theory of performativity* as a contemporary derivation from speech acts theory, which, according to the stance made in this paper, appears to be an essential tool to understand this issue clearly.

As is widely known, speech acts theory dates back to the '60s, namely to the publication of John Langshaw Austin's *How to Do Things with Words* (AUSTIN 1962), with performativity becoming one of the «key-concepts» of pragmatics (CONTE 1994, 382). Firstly, Austin distinguished between *constatives* and *performatives*[12]. Secondly, he made a further distinction between:

a) *locutionary acts*, the action «*of* saying something»;
b) *illocutionary* acts, focusing on the conventions that make it possible to perform acts «*in* saying something»;
c) *perlocutionary* acts, which focus on the consequences of the action performed «*by* saying something» (AUSTIN 1962, 94-107).

As to the differences between *illocutionary* and *perlocutionary* acts, Austin maintains that the former have *conventional consequences*, while the latter produce *material consequences*. The production of either type does not bear on the *intentional* or *unintentional* character of the act: there could be intentional illocutionaries that do not

315 U.S. 568 (1942), whereby «fighting words» are understood in terms of those which, in just being uttered, cause a harm or tend to incite to a simultaneous «breach of the peace».

[11] Indeed, this is the argument recalled e.g. by EMERSON 1963, 880-881; equally, analysing Emerson's theses, see SÁNCHEZ GONZÁLEZ 1992, 31-33.

[12] *Constatives* are describing statements, whereas *performatives* are statements by which one performs an action (AUSTIN 1962, 60); the latter depend on the existence of conventions and non-linguistic rules (HORNSBY 2008, 904), even though distinguishing between the two is actually impossible, because of every constative also being a performative (AUSTIN 1962, 52). This has made philosophers claim that Austin maintains the impossibility of distinguishing between *speech* and *action* (AVRAMIDES 2001, 66; HORNSBY 1994, 187), in a refusal of Cartesian philosophy (SBISÀ 2013, 76; see BRACKEN 1994; HEINZE 2016, 137; GELBER 2002, 50).

have positive outcomes, in the same way as there could be unintentional perlocutionaries that produce material consequences.

The example that follows can clarify the difference between the two. In giving the order «Shoot him!», the speaker is making an *illocutionary* speech act, which is giving an order based on conventional consequences brought about by language. Yet the speaker could also produce material consequences, a *perlocutionary* use of speech, which amounts to the material effects of the speech act: that is, for the listener to be effectively persuaded to shoot[13].

Both elements, in Austin's reflections, could come together with each speech act. However, one might analyse them as different features of language on which different stances on hate speech can be taken – this will be shown in the following interpretations of the speech acts theory, whereby illocutionary and perlocutionary acts come to be separated. Indeed, the *illocutionary* level could be separated from the *perlocutionary* one in that a single speech act could fail to produce perlocutionary consequences and still work at the illocutionary level (ZANGHELLINI 2003, 467-468).

After Austin, John Searle claimed that speech acts theory must only focus on *illocutionary acts* (SEARLE 2001, 221; see SEARLE 1965, 221-239; RABOSSI 1972; SBISÀ 2013, 39-40), which makes it possible to say that in his view the performative is reduced to illocutionary acts[14]. He then studies intentionality as an objective feature of language, underlining the role of social institutions, rules, conventions and contexts in the making of speech acts: that is to say, according to him, the meaning is the result of social practices (SEARLE 2002, 142-155)[15] and is not related only to what the speaker actually meant.

Thus, Searle helps to draw on Austin's distinction in conceiving a theory of action through language; all speech acts, he says, are actions whose effects depend on social conventions. Moreover, he clearly states that *intentionality* is a requisite for the performative, adding that while perlocutionary acts require physical action as well as intentionality, illocutionary acts are performed in their mere utterance (that is, intentionality is a self-sufficient criterion: illocutionaries do not depend on action).

Searle's debate with Jacques Derrida (see NEALON 2017, 1-5; ALFINO 1991, 143-152; PAYNE 1995, 5-6; NORRIS 1982), following the publication of the latter's *Signature,*

[13] However, this does not mean that the listener must *actually shoot* for the speech act to bring about perlocutionary effects: the material effects in this case are that the listener be *effectively persuaded to shoot*, not to *actually shoot*. One could be effectively persuaded to shoot (chemical substances in the brain actually change when that happens) and yet not shoot (i.e.: trying to do so, but the weapon does not work; trying to do so, but then getting distracted; actually doing so, but the weapon is fake; firstly getting convinced, but then changing one's mind). Philosophically and legally speaking, this makes sound difference.

[14] Moreover, Searle links speech acts to a general theory of action, maintaining that speech acts are not «the point of intersection of a theory of language and a theory of action»: as opposed to this, «a theory of language is part of a theory of action» (SEARLE 1969, 16), what truly makes it explicit that he too rejects Cartesian philosophy, subscribing to a theory of knowledge in material form (SMITH 2012, 2).

[15] Indeed, Searle argues for the existence of an «extra-linguistic institution» (SEARLE 1989, 548) which gives speech acts their actual meaning.

Event, Context (DERRIDA 1988), sheds light on the possibility of performatives always being the product of *repetition* beyond particular contexts (ALFINO 1991, 147-148), which, as Derrida maintains, makes them acquire deep meanings depending on their «citationality» (DERRIDA 1988, 6-18). What is useful about Derrida's «citationality» is that the particular context in which performatives are uttered, carry all the tracks of their previous uses in other contexts: that is to say, the meaning of a performative is a constant quote («citation») of other infinite speech acts.

Drawing on Derrida's work, contemporary scholars have introduced the notion of «echoing responsibility» (MEDINA 2006, 140). This could refer to the responsibility of hate speakers in putting across a message that is composed of all the meanings that it has acquired within the practices of a particular society, thus becoming performative as in *constructing social reality* (see BUTLER 1997). An example of this could be the meaning that the words «black» and «white» have acquired when associated with race issues and related verbal offences[16]. Moreover, the contemporary use of the theory of performativity makes it possible to take into consideration *social iterability*, which would prevent the adoption of an individualistic approach in coming up with a legal answer to the issue of hate speech. *Social iterability* outlines the effects of hate speech as not exclusively material and individual; rather, they are capable of transcending the particular context they originate in and insinuate themselves in social relationships, consequently contributing to their structuring based on the subordination of vulnerable subjects (see BUTLER 1997, 150-152; LOXLEY 2007, 134; LOIZIDOU 2007, 19).

In brief, such an evolution of speech acts theory is what lies behind its «contemporary uses», the analysis of which can be helpful in describing the useful tools that the contemporary theory of performativity provides. Indeed, after what has been described in philosophy of language as a «*linguistic turn*» (RORTY 1967; COLEBROOK 2010), which has given birth to speech acts theory, there emerges another defined as «*performative turn*» (MEDINA 2010, 286-287), which has allowed a resurgence of speech acts theory within a new philosophical debate on the issue of action in normative, political and social terms. In this respect, the contemporary theory of performativity underlines a potentially harmful and offensive role of speech acts (MARTÍNEZ GARCÍA 1999, 338): a highly relevant issue in the analysis of hate speech.

Furthermore, these issues could arguably go along with two different conceptions of liberty: a *liberal* and a *neo-republican* one, which, as will be shown, could account for different understandings of freedom of expression and hate speech.

III. HATE SPEECH: *PERLOCUTIONARY* OR *ILLOCUTIONARY*?

The theoretical elements analysed so far constitute what can be defined as the philosophical foundations of the contemporary theory of performativity. Such theory recalls speech act theory and relaunches it towards a "political" use. Two different uses

[16] As an example of the fact that race is a product of social construction, see the major works of the *Critical Race theorists*: CRENSHAW-GOTANDA-PELLER-THOMAS 1996; THOMAS-ZANETTI 2005.

of the theory of performativity could be made within different political-philosophical backgrounds, so as to come to two different answers to hate speech-related issues.

On one side, adopting a *perlocutionary* approach to hate speech, together with a distinction between *speech* and *action* and a liberal model of freedom, human rights and harm, could lead to an «oppositionist» answer to the issue of limiting freedom of expression when it comes to hate speech, accepting such limitations only in cases in which «*speech* shades into *action*» (see POST 1995, 639)[17].

On the other, an *illocutionary* approach to hate speech in line with a neo-republican understanding of freedom and human rights, appears to make way for a «prohibitionist» answer based on the rejection of the distinction between *speech* and *action*[18].

These two accounts of the use of the theory of performativity will be shown to possibly lie behind the normative definitions of hate speech given by two different instruments of international human rights law: the International Covenant on Civil and Political Rights (ICCPR) and the International Convention on the Elimination of All Forms of Racial Discrimination (ICERD). This distinction of approaches could arguably lead to draw a line between *criminal* and *civil* (and other forms of) State intervention against hate speech.

III.1. A liberal model of freedom, human rights and performativity: performatives as perlocutionary speech-plus acts

The use of performatives to give account of hate speech only as *perlocutionary* acts could be understood in connection with a liberal theory of freedom and freedom of expression, a liberal model of rights and harm and an acceptance of the distinction between *speech* and *action*. All these, as will be shown, could be read as the philosophical elements lying behind the definition of hate speech given by the ICCPR.

It is within classical liberalism that the foundation of freedom of expression is laid (ANSUÁTEGUI ROIG 1994, 2018; PÉREZ DE LA FUENTE 2010a). This perspective adopts an understanding of public opinion in its *subjective* sense, focusing on the subject of the action of freedom of expression and less so on the content (RODRÍGUEZ URIBES 1999, 81-83). In other words, it is a matter of analysing the *who* and *how* of freedom of expression.

Within the contemporary debate, this model is characterised by its general inclination to absolutism (as in the «*free speech absolutists*» model described by SUNSTEIN 1993, 5). This does not mean that *there are no limits* to freedom of expression (BAKER

[17] In fact, Post recognises that in some cases «speech shades into *conduct*» (italics added), but that would be a misleading definition, inasmuch as, legally speaking, speech is always conduct: what Post was making reference to was evidently action.

[18] For the distinction between *oppositionists* and *prohibitionists* in the hate speech debate, see HEINZE 2016, 24. Other distinctions could be adopted, such as KENNEDY's (2003, 36, 124) between «eradicationists» and «regulationists».

2012, 57-58 footnote), but those limits only emerge in case of serious harm or conflicts of rights. In particular, harm and conflicts of rights are based on an implicit assumption related to the distinction between *speech* and *action*: only when speech shades into action – or is likely to – the State must intervene in defence of other individuals' rights or to avoid greater harm (EMERSON 1963).

The main features of the liberal model within the hate speech debate can be listed as follows: 1) a negative conception of rights (GELBER 2002, 29-35; see HABERMAS 1994, 223); 2) a negative conception of freedom (BERLIN 2002) or, in some instances, liberty of moderns (CONSTANT 2001); 3) the faith in the *marketplace of ideas*[19] principle (*see* PÉREZ DE LA FUENTE 2010a).

These characteristics lead to identifying the absolutist model with a libertarian understanding (FISS 1996, 3; see PÉREZ DE LA FUENTE 2010a, 83-87). In fact, not all liberal arguments can be classed as libertarian; yet, the latter is the understanding at the heart of the points made for the prevalence of freedom of expression in the case of hate speech (HEYMAN 1998, 1307). Indeed, a negative conception of freedom as non-interference (BERLIN 2002) is arguably the general account of liberty – and freedom of expression – set forth not just by libertarians, but by liberals too[20] (PETTIT 2013, 175; *see* WALDRON 1987, 130). Further, such a conception of freedom as non-interference is what lies behind the account of freedom of expression of several liberal thinkers (EMERSON 1963, 895; SCHAUER 1982, 129; GREENAWALT 1989b, 121-122; see HEINZE 2016, 97).

The distinction between speech and action seems to stem from Mill's case of *corn-dealers*[21]: an excellent example of the liberal model's ability to account for the possibility of *words* shading into *action* (CUEVA FERNÁNDEZ 2012, 453), depending on the context. This suggests that there are certain limits to freedom of expression, which in Mill arise in connection with the harm principle.

It is common knowledge that the harm principle is the cornerstone of Feinberg's work (see FEINBERG 1984), which draws a line for criminal intervention in liberal States only to acts that cause *harm* to others[22]. The other principle, related to *offense* instead of harm (see FEINBERG 1985), only covers minor cases of psychological and mental evils, such as emotional distress, anxiety and irritation, and cannot always justify criminal intervention. Feinberg distinguishes between two categories of «unhappy but not necessarily harmful experiences [...]: those that *hurt* and those that *offend*» (FEINBERG 1984, 46). The first are physically hurting, whereas the latter are only «nonpainful mental

[19] See *Abrams v. United States*, 250 U.S. 616 (1919). SUNSTEIN 1993, XVIII recalls the relation between the liberal model and such a metaphor.

[20] For a reading of freedom as non-interference in Rawls, see SPITZ 1994. For Dworkin's understanding of freedom, in turn, see DWORKIN 1996, 21-26.

[21] The example of corn-dealers allows Mill to distinguish non-harmful opinions from those which incite to violence depending on the context; let us consider, for instance, the case of the opinion that corn-dealers cause poor starvation, which, although harmless in itself, can become harmful if expressed in a starving crowd standing before a corn-dealer's house (see MILL 1978).

[22] Feinberg's notion of *harmful acts* includes all those acts which cause serious harm to individuals by wrongfully thwarting their interests (FEINBERG 1984, 34).

states». Theoretically speaking, they could both harm, providing they become so serious and painful that the State cannot avoid intervening; but in Feinberg's description, the standard for seriousness is made up on grounds of physical pain (MACLURE 2017, 137). In relegating hate speech acts to the second category, it becomes clear that his harm principle is based on a distinction between *speech* and *action* (as well as an account of freedom as non-interference, which allows the State to interfere with individuals' freedom only in case of harm: FEINBERG 1984, 7-8).

As to the link between freedom as non-interference and the liberal rights-based models, let us consider Ronald Dworkin's understanding of *rights as trumps* (DWORKIN 1977). These rights make the case for a private space defined by the "wall" of individual rights, which trump all other considerations based on public policies (SKINNER 1986, 248). In Dworkin's liberal account, rights could be limited only by other rights: that means that in order to make the case for the limitation of freedom of expression in cases of hate speech, another competing right must be juxtaposed. Thus, freedom intended as non-interference in one's rights could be limited only for the sake of guaranteeing another individual's right.

In hate speech, finding another competing right to freedom of expression is a daunting task (see FIGUEROA ZIMMERMANN 2016, 4-5): the right to be free from discrimination, for instance, could be understood as limiting freedom of expression *if and only if* hate speech amounts to *actual* discrimination. That is to say, there must be an *action* of discrimination, as a perlocutionary understanding of performatives would suggest. Indeed, freedom as non-interference is aimed to avoid *acts* of interference in one's sphere (NADEAU 2004, 114; LABORDE-MAYNOR 2008, 16). When offending another person on the basis of their race (racist hate speech), for example, one could be punished for hate speech only if one incites others to violent *action* or *acts* of discrimination. The liberal model, therefore, accepts a general distinction between *speech* and *action*, but it does not accept that some speech acts are bad in themselves, regardless of the material consequences that they could bring about (as an illocutionary account would admit).

Dworkin himself, the key-figure of «deontological oppositionism» (HEINZE 2016, 11), by virtue of a liberal rights-based approach, declares his dislike of bans on hate speech (DWORKIN 1992, 61-64), claiming the necessary and insurmountable bond between democracy and free speech. Freedom of expression, Dworkin argues, is untouchable, in that it constitutes a universal human right (DWORKIN 2009, V) and therefore cannot be separated from human dignity, regardless of the content. In other words, freedom of expression includes also the right to offend (DWORKIN 1996, 218). In Dworkin's arguments it is possible to underline an implicit acceptance of the distinction between speech, which cannot harm in itself, and action, which instead can. This is most blatant in a paragraph in which – distinguishing between «upstream» laws against discrimination and «downstream» laws against hate speech – hateful expressions are appointed to the sphere of opinion forming, while the material consequences of sexism, intolerance, and racism (which fall in the domain of policies and laws against discrimination) are held to be damaging for effectively belonging to the sphere of action (DWORKIN 2009, VIII). Those values which are deemed as being harmful to third parties cannot be prevented from being

expressed: one can only prevent individuals from *acting* on such values (BAKER 2009, 142).

Furthermore, a position in the hate speech debate known as «consequentialist prohibitionism» (HEINZE 2016, 33-35) stands out for a double reading of the causation of harm in hate speech: *material/direct* or *immaterial/indirect*[23] (HEINZE 2016, 125-129 e 137-142). Although generally adopting a critical reading of liberal categories, and accepting hate speech bans in some cases, those who adhere to the *direct* view of causation accept the liberal distinction between *speech* and *speech-plus*. Academically speaking, this is the approach that lies behind the works of some exponents of *Critical Race Theory* (see THOMAS-ZANETTI 2005; CASADEI 2008; CASADEI 2016; ZANETTI 2017), such as Jean Stefancic, Richard Delgado and Mari J. Matsuda[24], who believe that hate speech causes negative material psychological consequences such as stress, depression, alcoholism, and therefore adopt a *perlocutionary* approach (DELGADO-STEFANCIC 2004, 14; MATSUDA 1989, 2332-2336).

This is even more apparent in libertarians' accounts of hate speech, as that of Nadine Strossen, who argues in favour of the distinction between «speech advocating unlawful conduct» and «the unlawful conduct itself» (STROSSEN 1990, 531). On these bases, she criticises those who claim that there is no difference between racist speech and racist conduct.

For all these reasons, a *perlocutionary* approach to performatives and hate speech could be understood under both a *harm-based* and a *rights-based* approach to freedom of expression, limiting State intervention in the prohibition of hate speech only to cases in which *speech* shades into *action* – or is very likely to –, thus causing harm or violating another individual's right by means of its material consequences. In fact, speech could do so, as underlined by the theory of performativity, thanks to perlocutionary speech acts. The exclusively *perlocutionary* approach, therefore, is arguably influenced by a liberal model of freedom of speech: the *content* of the expression is not relevant as such, but only the material effects it can produce according to the distinction between *mere speech* and *speech-plus* (*see* § 2).

All the features described so far could be said to underlie the model adopted by the ICCPR: namely, a distinction between *speech* and *action*, a *perlocutionary* account of performatives, an understanding of freedom as non-interference and a liberal model of rights and harm.

In order to demonstrate such conclusions, let us recall the definition of hate speech given by Art. 20, Par. 2, ICCPR based on the «advocacy of national, racial or religious hatred that constitutes *incitement to discrimination, hostility or violence*». Three elements are necessary for a conduct to fall within this definition (ZICCARDI 2016, 21): a) *willingness* or *intentionality* of the conduct; b) *effectivity* of the incitement; c) *verification*

[23] The immaterial/indirect approach, in turn, is referred to in the following paragraph.
[24] For their understanding of hate speech, see, among others, PÉREZ DE LA FUENTE 2010b, 152-153.

or *verifiability* of the acts. All three make it possible to shed light on the kind of approach this definition is based on; as a matter of fact, through the lenses of performativity, this falls within a *perlocutionary* understanding and a distinction of *speech* and *action*.

In the first place, this means that in order for it to be prosecutable, hate speech must be expressed in a performative conduct: it is a matter of employment of speech or other forms of expression, consistent with speech acts theory, aimed at sending out a message able to "do things with words". Further, such performative conduct must be the cause of *material effects*, either on its victims (being subject to discrimination, hostility or violence – *verification* – or there being a present danger that such effects will produce – *verifiability*), or on people subject to the act of incitement (*effectivity* of the incitement). This is based on a general distinction between *speech* and *action*. Let us recall the example of the order «Shoot him», this time in a "hate speech-fashion", as in «Shoot that nigger down!»: in this case, the listener should be effectively persuaded to act, or it should be very likely. This account could be therefore defined *perlocutionary*, in that it requires of the performative conduct a material consequence (or a likelihood). In short, such an account does not draw on the possibility of the production of "just" illocutionary consequences, as it requires the material consequence of the incitement.

In the second place, the conduct must be intentional. In stating so, as generally required by contemporary criminal law in constitutional states and systems of rule of law, it refuses Austin's account of intentionality, which, as shown earlier, was not a feature distinguishing illocutionary from perlocutionary acts. As to Searle, this could be consistent with his account providing that the responsibility of intentionality meets the requirements of criminal law for intentionality itself.

Moreover, the adoption of a model of freedom as non-interference in others' private sphere and enjoyment of rights is evident in the fact that the incitement brought about by hate speech, as its perlocutionary consequences, can be understood to possibly undermine e.g. the right to equality and the right to be free from discrimination, the right to physical integrity and the right to reputation. When State intervention is required for the protection of interests other than individual rights, in turn, the liberal account is able to give grounds to such intervention only in cases of material actions as perlocutionary outcomes, in connection with the harm principle (SORIAL 2013, 63). Therefore, such a model is arguably based both on a liberal *rights-based* model and on a *harm-based* model[25].

[25] See Art. 19.3 ICCPR: «The exercise of the rights provided for in paragraph 2 of this article carries with it special duties and responsibilities. It may therefore be subject to certain restrictions, but these shall only be such as are provided by law and are necessary: *(a) For respect of the rights or reputations of others; (b) For the protection of national security or of public order (ordre public), or of public health or morals*» (italics added).

III.2. A neo-republican model of freedom, human rights and performativity: performatives as illocutionary speech acts

If adopting a neo-republican model of performativity and freedom, hate speech could be understood as producing relations of subordination in society. This constitutes the sheer conventional consequence of *illocutionary* speech acts bringing about social subordination (ZANGHELLINI 2003, 485-486), which can be read within a Foucaultian notion of genealogic power (see FOUCAULT 1977 and 1983) as «productive, or constitutive, of social realities, truths and subjectivities» (LEEZENBERG 2013, 290)[26].

These considerations can be understood, among others, thanks to Rae Langton's account of the *theory of subordination*, which she uses as a prohibitionist answer to pornography and hate speech. In a 1993 article (LANGTON 1993), Langton analyses the *illocutionary* dimension of speech and pornography[27] as acts of subordination (HORNSBY-LANGTON 1998, 21; LANGTON 1993, 296). However, in Langton's opinion, the strategy should *not* be to maintain that pornography and hate speech do not fall within the category of *speech* (and therefore they are *actions*), so as to give arguments for the lack of protection under the First Amendment. On the contrary, Langton does not see the solution in the distinction made between *speech* and *action* and does not see it as a tool useful for solving the issue[28].

Reinterpreting the *silencing effect* argument (see DELGADO-STEFANCIC 1996, 480-481; GELBER 2002, 209), Langton focuses on two distinct dimensions of pornography and hate speech: an *illocutionary* and a *perlocutionary* one. As to the *illocutionary* dimension, Langton recalls the necessity of the "authority condition" of the speaker, which, according to her, can be gained through discursive practices. This makes it possible to understand Langton's idea of the *performative* character of hate speech (LANGTON 2018, 1-2). Indeed, hate speech *creates* a sort of authority – which, however, needs some already consolidated social practices, albeit not formally – by enabling, normalising and affirming the acceptability of given behaviours (LANGTON 1993, 316). The authority of hate speech, therefore, is the informal authority of a social practice (LANGTON 2018, 6; see TIRRELL 2012) that allows hate speech itself to constitute an *act of linguistic subordination* regardless of its consequences.

However, Langton's view of hate speech also consists of a second dimension. In her analysis, the silencing effect of pornography prevents women from "doing things with words" themselves (HORNSBY-LANGTON 1998, 21; LANGTON 1993, 314). Indeed, she

[26] Indeed, Foucault analyses power in biopolitical forms, where the subject appears to be the product of construction, the victim of power's discipline. (*ex multis*, *see* O'LEARY 2010; DIGESER 1995, 40-43).
[27] Langton's account here rests on Catharine MacKinnon's work, who had maintained that pornography is a case of *hate speech against women* (MACKINNON 1993, 22),
[28] Indeed, as MacKinnon herself had stated: «It is not new to observe that while the doctrinal distinction between speech and action is on one level obvious, on another level, it makes little sense. In social inequality, it makes almost none. Discrimination does not divide into acts on one side and speech on the other. Speech acts. It makes no sense from the action side either. Acts speak. In the context of social inequality, so-called speech can be an exercise of power which constructs the social reality in which people live», MACKINNON 1993, 29-30.

theorises the so-called *illocutionary silence* (or «illocutionary disablement», LANGTON 1993, 315), that is the particular type of silence which occurs when people speak but fail to obtain the desired effects, as well as failing to put in place the actual discursive action intended originally. Thus, pornography (as well as hate speech) comes about as a practice able to silence women on an illocutionary level, making it impossible for them, in sexual practices, to illocutionary say «no» in view of the fact that, as an effect of pornography itself, it would be perceived as a «yes». In this account of the performative, the illocutionary and perlocutionary levels are not independent from one another. However, Langton's use of illocutionary disablement as a perlocutionary consequence, in Austin's terms, has to do with conventional consequences and not material ones. Therefore, she gives preminence to the *illocutionary* dimension of speech acts, implying perlocutionaries as a side-effect.

This is what allows Langton to deal with *incitement* to hatred as a form of speech act both *illocutionary* and *perlocutionary*, maintaining that the verbs *incite*, *promote* and *advocate* contain both the former – in that they constitute the action of incitement, promotion and justification of hatred itself just in being uttered, without there being need of material consequences – and the latter – for recalling prelocutionary consequences (LANGTON 2012, 75; LANGTON-HASLANGER-ANDERSON 2012, 758).

In such an analysis, Langton arguably succeeds in overcoming the traditional distinction between illocutionary and perlocutionary approaches, suggesting a reading able to keep together the two at the same time. Further, Langton appears to base such an account on an understanding of freedom of speech which goes beyond the liberal one, inasmuch as she calls for State intervention to prohibit hate speech so as to support freedom of expression and provide individuals with illocutionary freedom («freedom of illocutionary acts», JACOBSON 1995, 75-76).

In fact, Langton's account could arguably be explained through a neo-republican understanding of freedom as non-domination (PETTIT 1997, SKINNER 1998). In essence, Philip Pettit and Quentin Skinner made the case that the general understanding underlying liberalism of freedom as non-interference is not enough to understand current situations of lack of freedom. Non-domination is a more demanding account of freedom, in that it does not merely focus on *actual* interference as an *act*, but on *possible arbitrary interference* as a kind of *relationship* between subjects (freedom as «status»). That is to say, while non-interference aims to avoid *acts* of interference, non-domination focuses on *relations* between subjects (NADEAU 2004, 114; LABORDE-MAYNOR 2008, 16). This is why the general distinction between *speech* and *action*, within a neo-republican understanding of freedom, is incapable of providing a satisfactory reading of the relations of lack of freedom within hate speech. Indeed, being dominated is the condition of dependency on the structural social power excised by those who can arbitrarily influence the status of the dominated (LOVETT 2007, 119).

In this view, freedom can also be understood as the power to control one's own speech acts and not to be controlled by others': that is to say, not to be victim of «discursive unfriendly» interactions (PETTIT 2001, 69). Then, there is a powerful

dimension of domination, which could be described as *discursive domination*, lying behind subordination: it arises not when people are materially incited to act as a perlocutionary effect of speech acts, but through a subordinating use of illocutionary acts that are able to shape social reality (as in BUTLER 1997).

The *illocutionary* approach, as opposed to the exclusively perlocutionary one, gives also relevance to the *content* expressed, adhering to its need to be "virtuous" in an *objective* sense, as the republican model would maintain (RODRÍGUEZ URIBES 1999, 84). In other words, it is just the expression of content harmful to other people's dignity that outlines the irrelevance of the liberal distinction between *speech* and *action,* in order to resolve the issues underlying hate speech. For example, in saying «Back off, nigger!» to a black individual, it would be certainly relevant if my speech act incites a crowd of white men standing at my side to attack him (*perlocutionary* account); but the most relevant aspect would be that my reprehensible and offensive speech act contributes to a reality of social subordination in which white people are socially dominating over black people, regardless of the material consequences of my single conduct (*illocutionary* account). Further, recalling the element of «social iterability» (*see* § 2), the force of speech acts is often to be found not in the single speech act of an individual (individualistic approach), but in the much broader connection with its global social meaning contributing to shaping social reality.

Within the neo-republican model, the limits to freedom of expression are given far more value; this reflects into the republican understanding of rights (BARRANCO AVILÉS 2009). Fundamental rights embody virtue and are the point of convergence of individual interest and common good (RAZ 1995, 33; BARRANCO AVILÉS 2000, 89-90). What this model rejects is the individualistic principle, used to argue for the State non-intervention in freedom of expression; on the contrary, the government can be a «friend» of freedom of expression itself (FISS 1996, 2), while still posing limits on it as a way of guaranteeing it. Furthermore, some authors go so far as to suggest that non-domination should be the value underlying all human rights systems, giving an account of it as the general *right to rights* (BOHMAN 2009, 64-65)[29].

This means abandoning a liberal understanding of rights, which, as has been shown, requires that the State should intervene to limit an individual's right only when another competing right ought to be protected. Instead, within a neo-republican understanding of non-domination, the State can intervene irrespectively of the existence of a violation a competing right (BARRANCO AVILÉS 2001, 206-207). Nor is freedom as non-domination related to harm-based understandings of State intervention (HONOHAN 2002, 195-196).

[29] Consider Bohman's proposal: «[B]asic human rights can be understood as fundamentally concerned with basic freedoms, including rights against [...] domination. These basic freedoms are negatively and positively justified as necessary conditions for avoiding [...] domination and destitution on the one hand and for living a worthwile human life [...] on the other. [...] Human rights provide the criterion of justice for claims that can be made upon others on the basis of shared freedom, a freedom that is possible only when social coexistence is based on non-domination», (BOHMAN 2009, 64-65).

In the case of an illocutionary understanding of hate speech, for example, it is not relevant whether my racist hate speech is concretely able to interfere in someone's rights (such as not to be discriminated, to name one) or it causes serious harm: what is relevant is an objective situation of subordination which, inasmuch as it is incompatible with non-domination, should be removed by the State. That is to say, the State could also intervene limiting illocutionary uses of hate speech, to promote a more tolerant environment irrespectively of effective harm or violations of rights. This makes it possible to say that the illocutionary understanding of hate speech, combined with a neo-republican understanding of freedom as non-domination, makes the case for rejecting the distinction between *speech* and *action* as a basis for State intervention. As a matter of fact, there should not be any distinction between speech and action, but only between lawful and unlawful expressive conducts.

A similar understanding is reflected in some authors' positions in the debate on hate speech. Jeremy Waldron, e.g., rejects solving the problem of regulating hate speech solely by means of the distinction between *speech* («words») and *action*. Instead, he goes as far as to state that hate speech cannot be considered just as an exercise of speech, thus accepting its *performative* dimension (WALDRON 2012, 166). Moreover, such a stance can arguably go together with an *illocutionary* view of performatives. Therefore, he believes hate speech to be a *malum in se*, regardless of its consequences, in that it is an intrinsic feature of language itself. Hate, inasmuch as indistinguishable just through the speech/action lenses, is visible; that is why Waldron denies its being immaterial.

Furthermore, within the so-called «consequentialist prohibitionism» (see *supra*, § 3.1), those who adopt an *immaterial/indirect* reading of the causation of the harm in hate speech (HEINZE 2016, 125-129 and 137-142), use phenomenological, socio-linguistic and deconstructionist studies, advocating for a theory of materiality of language and adopting a non-Cartesian approach: an indistinction between *res cogitans* and *res extensa* – in other words, between *speech* and *action* (HEINZE 2016, 138; see WILLIAMS 2011). On these grounds, they affirm the existence of diffused effects of hate speech that cannot (and should not) be empirically demonstrated. This is, among others, Charles Lawrence III's approach. Recalling cognitive psychology, he reminds of the existence of cultural stereotypes which cause the standardised repetition of unconscious behaviours whereby people learn judgements of value. If ignored, these behaviours become implicitly "allowed" with the effect of perpetuating the *status quo*, going so far as to structure social reality (LAWRENCE III 1987, 325-344). The existence of such stereotypes makes it possible, from Lawrence's point of view, for hate speech to constitute a harm in its mere utterance (*illocutionary* approach). According to Lawrence, therefore, the distinction between *speech* and *action* blurs to the point of vanishing, at least in hate speech (HEINZE 2016, 138). As a matter of fact, he claims that «[r]acism is both 100% speech and 100% conduct» (LAWRENCE III 1990, 444).

When it comes to the prohibition of hate speech within an *illocutionary* approach, unlike the ICCPR, the ICERD sets international norms as a defence against racist hate speech, which appear to qualify as negative the sheer dissemination of an illocutionary message that is derogatory to its victims, irrespectively of the actual incitement, the latter

being set forth as another different conduct. Indeed, from the definition of hate speech that can be found in the ICERD, there appears to emerge an approach which is compatible also with an *illocutionary* understanding. The Convention defines racial hate speech as «all *dissemination of ideas* based on racial superiority or hatred, *incitement* to racial discrimination, as well as all acts of violence or incitement to such acts against any race or group of persons of another colour or ethnic origin, and also the provision of any assistance to racist activities, including the financing thereof» (italics added).

In such a definition of hate speech, the different situations of incitement to discrimination, violence and *dissemination of ideas* based on racial superiority are all set apart. That is to say, dissemination of such ideas is an autonomous conduct different from the other ones. Such conduct, in particular, is much broader than the one adopted by the ICCPR, which makes the element of incitement a necessary requirement[30] (SPIGNO 2018, 288).

The contemporary theory of performativity can be helpful, within the approach hereby adopted, to understand the difference between the conduct of *incitement* and the much broader *dissemination of ideas*. As a matter of fact, as opposed to the first, the latter can fall within an *illocutionary* performative account: the "harm" of hate speech is to be found in the discriminatory subordination automatically brought about by it, irrespectively of the effect of incitement that it could have on third parties and/or of the negative effect on the victim. Harm, therefore, should not be empirically demonstrated.

The regional context of the Council of Europe also shows openness to the recognition of the illocutionary dimension of hate speech. As observed earlier, the Council of Europe recommendation n. 20 of October 30th 1997 defines hate speech as «*all forms of expression* which *spread*, incite, promote or *justify* racial hatred, xenophobia, antisemitism 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»[31]. Thus, one ought not to take into consideration only those forms of incitement to hatred or promotion of hatred, but also conducts that *spread* or *justify* hatred. It is therefore worth underlining that even the Council of Europe appears to recognise a *malum in se* in the illocutionary dimension of hate speech[32].

[30] This is also noticed by MENDEL 2012, 417-429, who anyway maintains that such an interpretative effect, even though logic, should be considered prohibited under Art. 19 ICCPR, which allows for limitations to freedom of speech only providing that restrictions are necessary. However, what is still to be demonstrated is that a restriction based on the sheer *dissemination of ideas*, rather than on *incitement*, would not be adequate with a view to posing necessary restrictions.

[31] Council of Europe, Committee of Ministers, Recommendation No. R (97) 20, *of the Committee of Ministers to Member States on "hate speech"*, adopted by the Committee of Ministers on 30 October 1997 at the 607th meeting of the Ministers' Deputies, Scope, Section 2 (italic added).

[32] The point is straightforwardly made by the Parliamentary Assembly of the Council of Europe, Report Doc. 9263, 12 October 2001, *Racism and xenophobia in cyberspace*, Draft Recommendation (http://www.assembly.coe.int/nw/xml/XRef/X2H-Xref-ViewHTML.asp?FileID=9540&lang=EN), in stating that it «considers racism not as an opinion, but as a crime».

III.3. Perlocutionary and illocutionary accounts: drawing a line between criminal and civil sanctions

However, such an *illocutionary* dimension is agreed upon reluctantly when discriminatory acts that are included among the criminally punishable conducts are not expressed through *action* but "only" through expression of thought (see RODRÍGUEZ MONTAÑÉS 2012), as is the case for ICERD. Indeed, contemporary criminal law in Member States would not accept the prohibition of illocutionary conducts of subordination, because, as shown earlier, they are understood to be "mere speech" and not speech shading into action. Furthermore, they can be not fully intentional and cannot be faced with an exclusively individualistic approach (as criminal law generally requires).

As to the question of *intentionality*, it is clear that while it would be possible to demonstrate it in cases of incitement or promotion of hatred – *perlocutionary* approach – the same would be impossible in cases where the speaker is contributing in a linguistic manner to the social subordination of others – *illocutionary* approach – to the extent of undermining their freedom as non-domination through «echoing responsibility» and «social iterability» (*see* § 2). This alone could be enough to exclude criminal intervention – which generally requires intentionality of conduct – for illocutionary subordinating hate speech.

As to the *individualistic* approach issue, the contribution to social subordination shall not be accounted for by taking into consideration individual actions, as these would not be enough to fully demonstrate the graveness of the issue of linguistic domination. But in doing so, criminal intervention would again go blatantly beyond its own limits.

In order to resolve this issue, therefore, it could be useful to draw a line between *perlocutionary* and *illocutionary* hate speech, as follows.

According to G. A. Kaufman, two are the essential requisites in order for a conduct to legally qualify as hate speech: 1) it has to be aimed at a vulnerable group, namely historically subject to discrimination; 2) at the same time, it has to express a humiliating message against that group, that is to say, which harms its dignity. Then, in order to intervene with criminal sanctions, that conduct has to meet either of the following requisites defining its *intentionality*: 3) the «malice», namely the *incitement* aimed at *third parties* to humiliate or marginalise the aforementioned groups or the people that belong to it; 4) direct «intentionality», as in aiming the humiliating or marginalising behaviour directly at a *group* or to *individuals* belonging to it (KAUFMAN 2015, 151-179).

Indeed, adopting the already mentioned definitions, criminally punishable could be those conducts of *incitement to discrimination* of historically discriminated groups aimed at third parties (such as crowds), as well as conducts of humiliation or discriminatory exclusion directly aimed at a historically discriminated group or at one of its members, in the latter case providing that material consequences arise from *speech* shading into *action*. Thus, conducts of hate speech defined through a *perlocutionary* approach would be criminally prosecutable.

In case hate speech conducts do not fall within those mentioned above, they can still be analysed as *illocutionary* hate speech. In this regard, it could be acceptable that they not be criminally prosecuted; however, although different from the other ones, such conducts also form part of the broad definition of hate speech, so that in their case it will be necessary to think of a non-criminal intervention. For instance, the instrument of civil tort liability could be used as well as an active and positive intervention of the State in favour of allowing hate speech victims to express themselves, so as to counter the discriminations they are subject to, which would allow to react to social subordination. Further, major State spending on human rights education could be another means of combating illocutionary subordinating hate speech.

This does not mean that the distinction between *speech* and *action*, rejected through the theory of performativity, would come back on stage to allow drawing a line in favour of the action of criminal law. Preferably, on the grounds of the fact that all expressive conducts are actions and all actions are expressive conducts, the distinction between *illocutionary* and *perlocutionary* approaches can make it possible to ask which consequences of expressive conducts are suitable to be subject to criminal sanctions and which, instead, to other legal instruments. In addition, a single conduct may fall under the two at the same time, therefore requiring both types of State intervention.

This would make it possible to develop a comprehensive approach to the issue of hate speech, that is able on the one hand to keep together the *illocutionary* and *perlocutionary* dimensions while giving a thorough legal answer that is not limited only to criminal interventions and, on the other hand, that is able to draw a line between this and other forms of legal intervention.

FINAL REMARKS

In conclusion, it is worth reminding that different conducts fall under the broad definition of '*hate speech*'. In order to make the case for State intervention against it, a line has to be drawn in such an intervention. Indeed, when hate speech qualifies in terms of *incitement*, the State should be allowed to intervene with criminal sanctions; as opposed, when it shows up in terms of *dissemination of ideas*, criminal sanctions are arguably not the best tool for countering these conducts, for the aforesaid reasons.

Nevertheless, it is possible to argue in favour of the need of statal intervention in order to protect victims' human rights in accordance with other instruments that should be imagined on grounds of a non-exclusively liberal understanding of freedom as non-interference – that is, giving value to an understanding of freedom as non-domination, or otherwise, being free from social subordination perpetrated through speech acts.

Such an approach, as a matter of fact, makes it possible to focus also on the *content* and not only on the *means* of the expression itself. The content of the message, under the general human right to freedom of speech, can indeed be detrimental to freedom from non-domination, therefore it is up to the State to give a legal answer capable of guaranteeing that social subordination is not brought about through hate speech. Indeed,

not only can hate speech be analysed within a *perlocutionary* approach – which, as shown earlier, would be more consistent with a liberal understanding of freedom as non-interference, freedom of speech and the distinction between *speech* and *action* –, but also within an *illocutionary* approach that is able to take into consideration power relations between the speaker and the hate speech target. These relations can be understood through the lenses of the theory of subordination and freedom as non-domination, as well as by recalling the need to abandon the *speech* v. *action* distinction.

In other words, social subordination, as an effect of hate speech, is a phenomenon which cannot be analysed only through the lenses of a liberal understanding of freedom of expression and perlocutionary tools, which separate speech from action and language from its social consequences. On the contrary, there is a broader requirement to understand freedom of speech based on non-domination within a neo-republican approach, which would be able to put into place an *illocutionary* attention to the implicit linguistic (and therefore social) subordination of hate speech victims as a side-effect of the collapse of the traditional liberal distinction between *speech* and *action*.

REFERENCES

ALFINO, M. (1991). Another Look at the Derrida-Searle Debate. *Philosophy and Rethoric*, 24(2), pp. 143-146.

ANSUÁTEGUI ROIG, F.J. (1994). *Orígenes doctrinales de la libertad de expresión*. Madrid: Universidad Carlos III de Madrid, Boletín Oficial del Estado.

ANSUÁTEGUI ROIG, F.J. (2002). *Poder, ordenamiento jurídico, derechos*. Madrid: Dykinson.

ASÍS ROIG, R. DE. (2010). *Sobre el concepto y el fundamento de los derechos. Una aproximación dualista*, Madrid: Dykinson.

AUSTIN, J.L. (1962). *How to Do Things with Words. The William James Lectures delivered at Harvard University in 1955*. Oxford: Clarendon Press.

AVRAMIDES, A. (2001). Intention and Convention. In: B. Hale, and C. Wright, eds., *A Companion to the Philosophy of Language*. Oxford: Blackwell, pp. 60-86.

BAKER, C.E. (2009). Autonomy and Hate Speech. In: I. Hare, and J. Weinstein, eds., *Extreme Speech and Democracy*. Oxford: Oxford University Press, pp. 39-158. ttps://doi.org/10.1093/acprof:oso/9780199548781.003.0009

BAKER, C.E. (2012). Hate Speech. In: M. Herz, and P. Molnar, eds., *The Content and Context of Hate Speech. Rethinking Regulation and Responses*. Cambridge: Cambridge University Press, pp. 57-80. https://doi.org/10.1017/CBO9781139042871.007

BARENDT, E. (2007). *Freedom of Speech*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199225811.001.0001

BARRANCO AVILÉS, M.C. (2000). Notas sobre la libertad republicana y los derechos fundamentales como límite al poder. *Derechos y Libertades*, 9, pp. 68-72.

BARRANCO AVILÉS, M.C. (2001). El concepto Republicano de Libertad y el modelo constitucional de derechos fundamentales. In: *Anuario de Filosofía del Derecho*, 18, 2001, pp. 205-226.

BARRANCO AVILÉS, M.C. (2009). *Teoría del Derecho y derechos fundamentales*. Lima: Palestra.

BERLIN, I. (2002). Two Concepts of Liberty [1969]. In: Id., *Liberty*, ed. by H. Hardy. Oxford: Oxford University Press, pp. 166-217. https://doi.org/10.1093/019924989X.003.0004

BOHMAN, J. (2009). Cosmopolitan Republicanism and the Rule of Law. In: S. Besson, J.L. Martí, *Legal Republicanism. National and International Perspectives*, Oxford: Oxford University Press, pp. 60-77. https://doi.org/10.1093/acprof:oso/9780199559169.003.0003

BRACKEN, H. (1994). *Freedom of Speech: Words Are Not Deeds*. Westport: Praeger.

BROWN, A. (2015). *Hate Speech Law. A Philosophical Examination*. New York: Routledge. https://doi.org/10.4324/9781315714899

BUTLER, J. (1997). *Excitable Speech. A Politics of the Performative*. New York: Routledge.

CARUSO, C. (2011). Ai confini dell'abuso del diritto: l'*hate speech* nella giurisprudenza della Corte Europea dei Diritti dell'Uomo. In: L. Mezzetti, and A. Morrone, eds., *Lo strumento costituzionale dell'ordine pubblico europeo. Nei sessant'anni della Convenzione per la salvaguardia dei diritti dell'uomo e delle libertà fondamentali (1950-2010)*. Torino: Giappichelli, pp. 339-352.

CARUSO, C. (2017). L'*hate speech* a Strasburgo: il pluralismo militante del sistema convenzionale. *Quaderni Costituzionali*, 37(4), pp. 963-981.

CASADEI TH., a cura di. (2008). "Razza", discriminazioni, istituzioni. Fascicolo monografico della *Rivista trimestrale di scienza dell'amministrazione*, 52.

CASADEI, TH. (2016). *Il rovescio dei diritti umani: razza, discriminazione, schiavitù*. Roma: DeriveApprodi.

COLEBROOK, C. (2010). The Linguistic Turn in Continental Philosophy. In: A.D. Shrift, ed., *The History of Continental Philosophy*, vol. VI, *Poststructuralism and Critical Theory's Second Generation*. Chicago: The University of Chicago Press, pp. 279-309.

CONSTANT, B. (2001). *La libertà degli antichi, paragonata a quella dei moderni* [1819], traduzione e cura di Giovanni Paoletti. Torino: Einaudi.

CONTE, A.G. (1994). Forma Performativa. *Anuario de Filosofía del Derecho*, 11, pp. 381-392, p. 382.

CRENSHAW, K., and GOTANDA, N., and PELLER, G., and THOMAS, K., eds. (1996). *Critical Race Theory: The Key Writings That Formed the Movement*. New York: New Press.

CUEVA FERNÁNDEZ, R. (2012). El «discurso del odio» y su prohibición. *Doxa. Cuadernos de Filosofía del Derecho*, 35, pp. 437-455. https://doi.org/10.14198/DOXA2012.35.18

DELGADO, R., and STEFANCIC, J. (1996). Ten Arguments against Hate-Speech Regulation: How Valid? *Northern Kentucky Law Review*, 23, pp. 475-490.

DELGADO, R., and STEFANCIC, J. (2004). *Understanding Words That Wound*, Boulder: Westview Press, Boulder.

DERRIDA, J. (1988). Signature, Event, Context [1972]. In Id., *Limited Inc.*, ed. by G. Graff. Evanston (Ill.): Northwestern University Press.

DIGESER, P. (1995). *Our Politics, Our Selves? Liberalism, Identity, and Harm*, Princeton (NJ): Princeton University Press. https://doi.org/10.1515/9781400821716

DWORKIN, R. (1977). *Taking Rights Seriously*, Cambridge (Mass.): Harvard University Press.

DWORKIN, R. (1992). The Coming Battles over Free Speech. *The New York Review of Books*, 11 June 1992, pp. 55-64.

DWORKIN, R. (1996). *Freedom's Law. The Moral Reading of the American Constitution*. Oxford: Oxford University Press.

DWORKIN, R. (2009). Foreword. In I. Hare, and J. Weinstein (eds.), *Extreme Speech and Democracy*. Oxford: Oxford University Press, pp. V-IX.

EMERSON, T. (1963). Toward a General Theory of the First Amendment. *Yale Law Journal*, 72, pp. 877-956. https://doi.org/10.2307/794655

FARROR, S. (1996). Molding the Matrix: The Historical and Theoretical Foundations of International Law concerning Hate Speech. *Berkeley Journal of International Law*, 14(1), pp. 1-98.

FEINBERG, J. (1984). *The Moral Limits of the Criminal Law*, vol. I, *Harm to Others*. Oxford-New York: Oxford University Press.

FEINBERG, J. (1985). *The Moral Limits of the Criminal Law*, vol. II, *Offense to Others*. Oxford-New York: Oxford University Press.

FIGUEROA ZIMMERMANN, F. (2016). Bajo sospecha: El liberalismo y la regulación del Hate Speech. In: *Libertades Públicas. Colección breviarios*, 6, pp. 1-17.

FISH, S. (1994). *There's No Such Thing as Free Speech, and it's a Good Thing, Too*. Oxford: Oxford University Press.

FISS, O. (1996). *The Irony of Free Speech*. Cambridge (Mass.): Harvard University Press.

FOUCAULT, M. (1977). *Discipline and Punish: The Birth of the Prison* [1975], trad. byA. Sheridan. Hardmondsworth: Penguin.

FOUCAULT, M. (1983). The Subject and Power. In: H.L. Dreayfus, and P. Rainbow, *Foucault: Beyond Structuralism and Hermeneutics*. Berkeley: University of California Press.

GASCÓN CUENCA, A. (2016). Estándares internacionales de protección ante el discurso del odio en el ámbito internacional. In: Id., *El discurso del odio en el ordenamiento jurídico español: su adecuación a los estándares internacionales de protección*. Cizur Menor: Aranzadi, pp. 23-73.

GELBER, K. (2002). *Speaking Back. The free speech versus hate speech debate*. Philadelphia: John Benjamins Publishing Company.
https://doi.org/10.1075/dapsac.1

GREENAWALT, K. (1989a). *Speech, Crime, and the Uses of Language*, Oxford University Press, Oxford, 1989, pp. 79-163.

GREENAWALT, K. (1989b). Free Speech Justifications. In: *Columbia Law Review*, 89, 1, pp. 119-155.
https://doi.org/10.2307/1122730

HABERMAS, J. (1994). Derechos humanos y soberanía popular: las concepciones liberal y republicana, trad. by J. González Amuchastegui. *Derechos y Libertades*, 3, pp. 215-230.

HEINZE, E. (2016). *Hate Speech and Democratic Citizenship*. Oxford: Oxford University Press.
https://doi.org/10.1093/acprof:oso/9780198759027.001.0001

HEYMAN, S. J. (1998). Righting the Balance: An Inquiry into the Foundations and Limits of Freedom of Expression. In: *Boston University Law Review*, 78, pp. 1275-1392.
https://doi.org/10.2139/ssrn.157818

HONOHAN, I. (2002). *Civic Republicanism*. London-New York: Routledge.
https://doi.org/10.4324/9780203460894

HORNSBY, J. (1994). Illocution and its significance. In: S. Tsohatzidis, ed., *Foundations of Speech Act Theory: Philosophical and Linguistic Perspectives*. London: Routledge, pp. 187-207.

HORNSBY, J. (2008). Speech Acts and Performatives. In: E. Lepore, and B.C. Smith, eds., *The Oxford Handbook of Philosophy of Language*, Oxford: Oxford University Press, pp. 893-910.
https://doi.org/10.1093/oxfordhb/9780199552238.003.0035

HORNSBY, J., and LANGTON, R. (1998). Free Speech and Illocution. *Legal Theory*, 4, pp. 21-37.
https://doi.org/10.1017/S1352325200000902

JACOBSON, D. (1995). Freedom of Speech Acts? A Response to Langton. *Philosophy & Public Affairs*, 24(1), pp. 64-79.

https://doi.org/10.1111/j.1088-4963.1995.tb00022.x

KAUFMAN, G.A. (2015). *Odium dicta. Libertad de expresión y protección de grupos discriminados en internet*, Consejo Nacional para Prevenir la Discriminación, Mexico.

KENNEDY, R. (2003). *Nigger: The Strange Career of a Troublesome Word*. New York: Vintage.

KOLTAY, A. (2013). *Freedom of Speech. The Unreachable Mirage*. Budapest: Wolters Kluwer.
https://doi.org/10.2139/ssrn.2216655

LABORDE, C., AND MAYNOR, J. (2008). The Republican Contribution to Contemporary Political Theory. In: Iid. (eds.), *Republicanism and Political Theory*. Oxford: Blackwell Publishing, pp. 1-28.
https://doi.org/10.1093/acprof:oso/9780199550210.003.0001

LANGTON, R. (1993). Speech Acts and Unspeakable Acts. *Philosophy & Public Affairs*, 22(4), pp. 293-330.

LANGTON, R. (2012). Beyond Belief: Pragmatics in Hate Speech and Pornography. In: I. Maitra, and M.K. Gowan, eds., *Speech and Harm: Controversies over Free Speech*. Oxford: Oxford University Press, pp. 72-93.
https://doi.org/10.1093/acprof:oso/9780199236282.003.0004

LANGTON, R. (2018). The Authority of Hate Speech. To be published in: J. Gardner, L. Green, and B. Leiter, *Oxford Studies in Philosophy of law*, vol. III, Oxford: Oxford University Press.
https://doi.org/10.1093/oso/9780198828174.003.0004

LANGTON, R., HASLANGER, S., and ANDERSON, L. (2012). Language and Race. In: G. Russell, and D. Graff Fara, eds., *The Routledge Companion to Philosophy of Language*. London: Routledge, pp. 753-767.

LAWRENCE III, CH.R. (1987). The Id, the Ego, and Equal Protection: Reckoning with Unconscious Racism. *Stanford Law Review*, 39, pp. 317-388.
https://doi.org/10.2307/1228797

LAWRENCE III, CH.R. (1990). If He Hollers Let Him Go: Regulating Racist Speech on Campus", in *Duke Law Journal*, 1990, pp. 431-483.
https://doi.org/10.2307/1372554

LEEZENBERG, M. (2013). Power in speech actions. In: M. Sbisà, and K. Turner, *Pragmatics of Speech Actions*, Berlin-Boston: De Gruyter Mouton, pp. 287-312.
https://doi.org/10.1515/9783110214383.287

LOIZIDOU, E. (2007). *Judith Butler. Ethics, Law, Politics*. New York: Routledge-Cavendish.

LOVETT, F. (2007). *A General Theory of Domination & Justice*, Oxford: Oxford University Press.

LOXLEY, J. (2007). *Performativity*. London: Routledge.

https://doi.org/10.4324/9780203391280

MACKINNON, C.A. (1993). *Only Words*. Cambridge (Mass.): Harvard University Press.

MACLURE, J. (2017). The Regulation of Hateful and Hurtful Speech: Liberalism's Uncomfortable Predicament. In: *McGill Law Journal*, 63, 1, pp. 133-154. https://doi.org/10.7202/1054353ar

MARCUS, K.L. (2008). Higher Education, Harassment, and First Amendment Opportunism. *William and Mary Bill of Rights Journal*, 16, pp. 1025-1059.

MARTÍNEZ GARCÍA, J.I. (1999). Los derechos humanos y el poder del lenguaje. *Derechos y Libertades*, 4(7), pp. 335-357.

MATSUDA, M.J. (1989). Public Response to Racist Speech: Considering the Victim's Story. *Michigan Law Review*, 87, 1989, pp. 2320-2381. https://doi.org/10.2307/1289306

MCHANGAMA, J. (2015). The Problem with Hate Speech Laws. *The Review of Faith & International Affairs*, 13(1), pp. 75-82. https://doi.org/10.1080/15570274.2015.1005919

MEDINA, J. (2006). *Speaking from Elsewhere. A Nex Contextualist Perspective on Meaning, Identity, and Discursive Agency*. Albany: State University of New York Press.

MEDINA, J. (2010). The Performative Turn and the Emergence of Post-Analytic Philosophy. In: A.D. Shrift, ed., *The History of Continental Philosophy*, vol. 7, *After Poststructuralism: Transitions and Transformations*, ed. by R. Braidotti, Chicago: The University of Chicago Press, pp. 275-305.

MENDEL, T. (2012). Does International Law Provide for Consistent Rules on Hate Speech?. In: M. Herz, and P. Molnar, eds., *The Content and Context of Hate Speech. Rethinking Regulation and Responses*. Cambridge: Cambridge University Press, pp. 417-429. https://doi.org/10.1017/CBO9781139042871.029

MILL, J.S. (1978). *On Liberty* [1859]. Indianapolis: Hackett Publishing.

NADEAU, CH. (2004). Non-Domination as a Moral Ideal. In D. Weinstock, Ch. Nadeau (eds.), *Republicanism: History, Theory and Practice*. London: Frank Cass Publishers, pp. 104-116

NEALON, J.T. (2017). Jokes and the Performative in Austin And Derrida; or, The Truth is a Joke? *Cultural Critique*, 95, pp. 1-24. https://doi.org/10.5749/culturalcritique.95.2017.0001

NORRIS, C. (1982). *Deconstruction: Theory and Practice*. London: Methuen.

O'LEARY, T. (2010). Michel Foucault. In: A.D. Shrift, ed., *The History of Continental Philosophy*, vol. 6, *Poststructuralism and Critical Theory's Second Generation*. Chicago: The University of Chicago Press, pp. 67-89.

PAREKH, B. (2012). Is There a Case for Banning Hate Speech? In: M. Herz, and P. Molnar, eds., *The Content and Context of Hate Speech. Rethinking Regulation and Responses*, Cambridge: Cambridge University Press, pp. 37-56.

https://doi.org/10.1017/CBO9781139042871.006

PAYNE, M. (1995). Introduction. In: S. Cavell, ed., *Philosophical Passages: Wittgenstein, Emerson, Austin, Derrida*, Cambridge (Mass.): Blackwell, pp. 5-6.

PECES-BARBA MARTÍNEZ, G. (1995). *Curso de Derechos Fundamentales. Teoría general*, con la colaboración de R. De Asís Roig, Carlos R. Fernández Liesa, Ángel Llamas Cascón. Madrid: Universidad Carlos III de Madrid – Boletín Oficial del Estado.

PÉREZ DE LA FUENTE, O. (2010a). Libertad de expresión y el caso del *lenguaje del odio*. Una aproximación desde la perspectiva norteamericana y la perspectiva alemana. *Cuadernos Electrónicos de Filosofía del Derecho*, 21, pp. 67-104.

PÉREZ DE LA FUENTE, O. (2010b). El enfoque español sobre lenguaje del odio. In: Id., J. Oliva Martínez, and J. Daniel, eds., *Una discusión sobre identidad, minorías y solidaridad*. Madrid: Dykinson, pp. 133-155.

PETTIT, PH. (1997). *Republicanism. A Theory of Freedom and Government* 1997, Oxford: Oxford University Press.

PETTIT, PH. (2001). *A Theory of Freedom: From the Psychology to the Politics of Agency*, Cambridge: Cambridge University Press.

PETTIT, PH. (2013). Two Republican Traditions. In A. Niederberger, and P. Schink, *Republican Democracy. Liberty, Law and Politics*. Edinburgh: Edinburgh University Press, pp. 169-204.
https://doi.org/10.3366/edinburgh/9780748643066.003.0008

RABOSSI, E.A. (1972). Locuciones e ilocuciones: Searle y Austin. *Revista Hispanoamericana de Filosofía*, 18(6), pp. 3-41.

RAZ, J. (1995). Rights and Politics. In: *Indiana Law Journal*, 7, 1, 1995, pp. 27-44.

RODRÍGUEZ MONTAÑÉS, T. (2012). *Libertad de expresión, discurso extremo y delito. Una aproximación desde la Constitución a las fronteras del derecho penal*. Valencia: Tirant lo Blanch.

RODRÍGUEZ URIBES, J.M. (1999). *Opinión pública. Concepto y modelos históricos*, Madrid-Barcelona: Marcial Pons, Ediciones Jurídicas y sociales.

RORTY, R. (1967). *The Linguistic Turn. Recent Essays in Philosophical Method*. Chicago-London: The University of Chicago Press.

SÁNCHEZ GONZÁLEZ, S. (1992). *La libertad de expresión*. Madrid: Marcial Pons, pp. 31-33.

SBISÀ, M. (2013). Locution, Illocution, Perlocution. In: M. Sbisà, and K. Turner, eds., *Pragmatics of Speech Actions*, Berlin-Boston: De Gruyter Mouton, pp. 25-75.
https://doi.org/10.1515/9783110214383.25

SCHAUER, F. (1982). *Free Speech: A Philosophical Enquiry*. Cambridge: Cambridge University Press.

SEARLE, J.R. (1965). What is a Speech Act? In: M. Black, ed., *Philosophy in America*, London: Cornell University Press, pp. 221-239.

SEARLE, J.R. (1969). *Speech Acts*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139173438

SEARLE, J.R. (1989). How Performatives Work. *Linguistics and Philosophy*, 12, pp. 535-558. https://doi.org/10.1007/BF00627773

SEARLE, J.R. (2001). J.L. Austin. In: A.P. Martinich, and D. Sosa, eds., *A Companion to Analytic Philosophy*, London: Blackwell, pp. 218-230.

SEARLE, J.R. (2002). Individual Intentionality and Social Phenomena in the Theory of Speech Acts. In: Id., *Consciousness and Language*, Cambridge: Cambridge University Press, pp. 142-155. https://doi.org/10.1017/CBO9780511606366.010

SKINNER, Q. (1986). "The Paradoxes of Political Liberty. In: *The Tanner Lectures on Human Value*, VII. Salt Lake City: University of Utah Press.

SKINNER, Q. (1988). *Liberty before Liberalism*, Cambridge: Cambridge University Press.

SMITH, B. (2012). John Searle: From Speech Acts to Social Reality". In Id., ed., *John Searle*, Cambridge: Cambridge University Press, pp. 1-33. https://doi.org/10.1017/CBO9780511613999.001

SORIAL, S. (2013). Free Speech, Hate Speech, and the Problem of (Manufactured) Authority. In: *Canadian Journal of Law and Society*, 29, 1, pp. 59-75. https://doi.org/10.1017/cls.2013.43

SPIGNO, I. (2018). *Discorsi d'odio. Modelli costituzionali a confronto*. Milano: Giuffrè.

SPITZ, J.-F. (1994). The Concept of Liberty in 'A Theory of Justice' and Its Republican Version. In: *Ratio Juris*, 7, 3, 1994, pp. 331-347. https://doi.org/10.1111/j.1467-9337.1994.tb00187.x

STROSSEN, N. (1990). Regulating Racist Speech on Campus: A Modest Proposal. *Duke Law Journal*, pp. 484-573. https://doi.org/10.2307/1372555

SUNSTEIN, C.R. (1993). *Democracy and the Problem of Free Speech*. New York: The Free Press (Macmillan).

THOMAS, K., and ZANETTI, GF., eds. (2005). *Legge, razza, diritti. La* Critical Race Theory *negli Stati Uniti*. Reggio Emilia: Diabasis.

TIRRELL, L. (2012). Genocidal Language Games. In: I. Maitra, and M.K. Gowan, (eds.), *Speech and Harm: Controversies over Free Speech*. Oxford: Oxford University Press, pp. 174-221. https://doi.org/10.1093/acprof:oso/9780199236282.003.0008

WALDRON., J. (1987). Theoretical Foundations of Liberalism. In: *The Philosophical Quarterly*, 37, 147, pp. 127-150.

https://doi.org/10.2307/2220334

WALDRON., J. (2012). *The Harm in Hate Speech*. Boston (Mass.): Harvard University Press.
https://doi.org/10.4159/harvard.9780674065086

WEBER, A. (2009). *Manual on Hate Speech*. Council of Europe Publishing, Settembre 2009, (disponibile al link: https://rm.coe.int/1680665b3f).

WILLIAMS, S.H. (2011). Democracy, Freedom of Speech, and Feminist Theory: A Response to Post and Weinstein. *Virginia Law Review*, 97(3), pp. 603-615.

ZANETTI, GF. (2017). *Critical Race Theory*: temi e problemi degli studi critici sulla "razza". In: M.G. Bernardini, and O. Giolo, eds., *Le teorie critiche del diritto*. Roma: Pacini Giuridica, pp. 35-49.

ZANGHELLINI, A. (2003). Jurisprudential Foundations for Anti-Vilification Laws: The relevance of speech act and foucauldian theory. In: *Melbourne University Law Review*, 27, pp. 458-487.

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