Hate speech – its legal, linguistic and philosophical conceptualisation

Maciej Zweiffel
Jagiellonian University in Krakow
maciej.zweiffel@uj.edu.pl
ORCID: 0000-0002-4010-3907

ABSTRACT

Objective: The multiplication of information exchange in the internet era inevitably also extends into aggressive communication, especially because communicating via digital devices produces a false feeling of impunity. In this context, it is worth inquiring how to define hate speech (1) and in what manner such elements of the speech act are realised – identity, sender, intention (illocutionary content), receiver, perlocution – in the case of a hateful communication (2). Methods: To reach these research goals, it is necessary to analyse basic legal concepts describing hate speech, then compare them and determine the context of communication acts with similar intention and defined in the penal code. These include defamation, insult, and punishable threat in the penal code, which are akin to hate speech. Results and cognitive value: This consideration has been presented with the help of the philosophy of language, to build a model of such communication and also to show its limiting features, i.e., this kind of transmission of information seems to contradict the essence of communication.

KEY WORDS
defamation, insult, hate speech, speech act, prototype, extension, non-occasional feature

The considerations presented here raise the question of what kind of linguistic activity hate speech is. This linguistic view involves a conscious narrowing of the set of phenomena under study and the use of specific tools of thought. The first indicates that the object of analysis can be primordial linguistic messages, or those that can be translated into the form of a statement (e.g., “this picture shows that…”). This limitation of the field of interest is not only due to the volume of the article but is also connected with the claim (here, of course, it is taken as an assumption) that natural language is a universal means of human communication. As for the
concepts used in the analysis, they include speech acts (in J. Searle’s terms), speech functions (developed by R. Jakobson), and the following elements of cognitive linguistics: cognitive definition (J. Bartmiński), prototype and extension (J.R. Taylor).

The discussion will begin with an analysis of the legal provisions juxtaposing hate speech and defamation, insult and threat, which are close in meaning to it. This comparison is intended to point out the peculiarities of hate speech against the background of these other forms of communication. Then, after proposing the linguistic characteristics of hate messages (prototypical approach), their relationship to the phenomenon of communication itself will be discussed.

Defining hate speech as a special type of communicative act proceeds from concepts that are clearly distinguishable from it (defamation, criminal threat) to the closest insult to it. The analysis of the data found, in this case legal documents, combined with linguistic methods, is intended to identify the features or peculiarities of hate speech.

Within the framework of the considerations presented, the definition of hate speech that has been developed will also make it possible to show the dangers to communication itself resulting from the formulation of such messages.

Defamation, insulting and threats

To approximate the concept of hate speech from a legal perspective, let’s turn to the semantically close terms of the Penal Code. Article 212 para. 1 of the Code defines defamation as slandering “another person, a group of persons, an institution, a legal person or an organizational unit without legal personality of such conduct or qualities that may degrade it in public opinion or expose it to loss of confidence necessary for a given position, profession or type of activity.”¹ Thus, the purpose of a defamatory communication is to humiliate the addressee, i.e. “to insult someone’s dignity, humiliate, disgrace” (Szymczak, 1998). Outlining a preliminary cognitive definition, it can be said that the activity in question consists in presenting someone or something higher (someone trustworthy, something highly valued, good) in terms of someone or something lower (someone unworthy, something negative, bad), e.g., a person as an animal, a praiseworthy or positive action (work, love, etc.) in the form of something disgusting (e.g. thievery, prostitution). This usually falls into the formulas: X, who seems to be someone worthy, is actually Y, i.e., someone unworthy, e.g., this businessman is actually a crook; this headmistress is a sadist. Alternatively, person X, known in public life for his highly regarded qualities a, b or c, is attributed their opposite. Typical defamation cases include accusing someone of financial abuse, providing sexual services, defrauding someone of something, fraud, behaviour unworthy of the position held, drunkenness, greed, dishonesty, and meanness (Wernik, 2018). Also crucial in the case of defamation is the issue of reference to the categories of truth and falsehood, for, if X, who seems to be someone of integrity, is accused of fraud and this accusation turns out to correspond to the facts, then such defamation becomes a “true accusation” (article 213 of the Penal Code) and, of course, loses the characteristics of a criminal act, becoming a socially desirable action, undertaken in “defense of a socially legitimate interest” (article 213 para. 2 item 2 of the Penal Code).

Insult, probably because of the multiplicity of behaviours that can be examples of it, is defined in article 216 of the Penal Code only as an action of a public nature and with a clearly identified addressee (“Whoever insults another person in his presence or even in his absence, but

¹ Law of June 6, 1997 – Penal Code (Journal of Laws of 1997, No. 88, item 553), as amended. Further references to the Penal Code will be indicated in parentheses by specifying an article, paragraph or point.
in public or with the intention that the insult should reach that person, shall be subject to a fine or the penalty of restriction of liberty” – article 216 para. 1 of the Penal Code). As A. Wernik (2018) points out, unlike defamation: “an insulting statement cannot […] be verified in terms of truth or falsity. This is because, most often, an insulting statement is a verbal utterance, involving the use of words that are generally considered abusive, insulting, vulgar or ridiculing. It can also be a gesture, e.g., spitting at someone, showing the middle finger.” Thus, while defamation is a communicative act in which – referring to R. Jakobson’s typology – the descriptive function dominates, the essence of an insult is the transmission of extremely negative emotions towards a person. Therefore, it can trigger an additional communication code – body language (article 216 para. 3 of the Penal Code) – which, in defamation ex definitione cannot occur, as it is impossible to assess its truthfulness (or, in other words, its truthfulness results from its relation not to the facts, but to the person of the addressee; it is, for example, a true, authentic expression of someone’s feelings).

Legal concepts that may be related to hate speech include criminal threats (article 190 of the Penal Code), which includes “threatening another person to commit a crime” (article 190 para. 1 of the Penal Code), and related stalking (article 190a of the Penal Code), which is, as indicated in the provision, “persistent harassment” that arouses in a person “a sense of threat, humiliation or anguish or significantly invades his privacy” (article 190a para. 1 of the Penal Code). Both crimes – like defamation or insult – can be treated as communicative behaviour, as the key elements of the communicative act appear in them: the sender, the receiver and the code through which these unfriendly (or “negative”) contents are transmitted. In the case of criminal threats and harassment, which can be treated as a series of threats, the performative function of speech and the pattern of the illocutionary act of promising become apparent with particular clarity, with, of course, an important modification regarding the expectations of the addressee.

Referring to the classic considerations of John Searle, the act of threat can be presented as follows. The sender N utters (formulates – if there is no direct contact) the message (utterance) W, so that the addressee A receives this message and decodes it. N, by uttering (formulating) W, promises A sincerely and unabashedly that he will perform a future action G toward A, with A preferring that N not do G (Searle, 1987, pp. 77-78).

The term “sincere and unadulterated” applies to all the communication activities under analysis, for, in each case, the defamation, insult or threat requires, as Searle (1987) puts it, “normal conditions on entry and exit” (p. 77). This framework includes, for example, sincerity of intent, not playing any stage role or not telling an unpleasant joke, correct decoding of messages and the corresponding perlocutionary effect occurring in the addressee (being offended, indignation, fear, etc.). Moreover, in these three cases, the perlocutionary effect turns out to be crucial, since the case for defamation, insult or threat is brought by the addressee of such messages (this will further turn out to be important when juxtaposed with hate speech addressed, for example, to immigrants unfamiliar with the Polish language).

Hate speech – the legal approach, the scope of the concept
As can be seen from the above remarks, the concept of hate speech is not included or defined in the provisions of the Penal Code. With this in mind, Ombudsman A. Bodnar proposed “creating a government strategy to counter hate crimes” (Jędrzejczyk, 2021). The government, however, assumes that “Polish criminal law already has adequate instruments of criminal-legal response to criminal behaviour motivated by discriminatory grounds” (Jędrzejczyk, 2021).

Arguably, the provisions of the Penal Code already discussed are also involved, but articles 256 and 257 are crucial. The first covers the prohibition of propagating totalitarian regimes and inciting...
“hatred on the basis of national, ethnic, racial, religious differences or on the basis of irreligiousness” (article 256 para. 1 of the Penal Code). The second focuses on racism and defines as a punishable act the public insulting of “a group of people or an individual person because of his national, ethnic, racial, religious affiliation or because of his irreligiousness” (article 257 of the Penal Code).

Compared with the previously analysed defamation, insult or threat, the distinguishing feature of “incitement to hatred”, which, in this context, can be treated as a close denominator of “hate speech”, turns out to be the relatively clear-cut scope of the meaning of such communications. The hatred expressed must be related to national origin, ethnicity, race (a category now rarely used) and religious worldview. The issue of the prohibition of the propagation of totalitarianism (article 257 of the Penal Code) makes it still necessary to include anti-Semitism (Nazism) in this field of meaning, and this is where the enumeration should end, unless the prohibition of communist agitation is linked to counteracting the stigmatisation of any social group (creating a “class enemy” for immediate needs and hatefully attacking it; examples are provided by the teachers’ strike or the persecution of judges by those in power).

However, if one turns to the primary European document related to hate speech, the result is not much richer in content. The Council of Europe Recommendation R (97) 20 of October 30, 1997, which defines hate speech, speaks primarily of attacks based on (a) racial or ethnic, (b) xenophobic, and (c) religious grounds. The introduction also mentions, citing the earlier 1993 Vienna Declaration, countering social tensions, but not defined in terms of class, economics, culture or customs. The core of the very definition of hate speech, however, includes the terms “racism,” “xenophobia” and “intolerance” (Recommendation No. R (97) 20 of The Committee of Ministers to Member States on Hate Speech, 1997).

This core, when confronted with reality, turns out to be very narrow in meaning. This has been emphasised in the literature, as well as in reports by the Ombudsman’s office (see A. Calek et al., 2019; “30 przykładów…”, 2019; Kowalski, 2019). After all, in moments of social tension, connected, for example, with strikes organised by certain professional groups or resistance to changes proposed by those in power and involving a particular group, there is an increase – especially on social media – of hate attacks related precisely to friction between the community on strike or opposing certain changes and the rest of society. The divisions that are then drawn can of course be – and unfortunately very often are – exploited by political parties, which further aggravates the situation. As an illustration of the phenomena in question, it suffices to recall the 2019 teachers’ strike (Rusak, 2022) and the instability associated with the changes in the judiciary introduced by the Law and Justice governments.

In an examination of national regulations, it turns out that hate speech, with a basis that cannot be reduced to ethnic, national or religious issues, is an open set, updated, as it were, each time in relation to a specific political or social situation (strikes, legal changes involving some professional group, its wage demands, etc.).

The social theme analysed should be emphasized because hate speech is most often defined in close connection with the concepts of xenophobia or intolerance, with the two terms treated synonymously. But is hate speech directed against, for example, striking (or non-striking) teachers or judges protesting government actions related to xenophobia or intolerance? It seems not, if one considers the prototypical meaning of these expressions, centred upon dislike or fear of someone/something foreign, different.

During a conference discussion at the 6th PTKS Congress (September 22-24, 2022), Prof. Marek Kochan pointed out the one-sidedness of this definition, which focuses exclusively on right-wing extremism.
This raises an important problem in defining hate speech, since at first glance it appears to be distinguished from insult, defamation and criminal threats by reference to concepts such as xenophobia and intolerance. For while these concepts are an important definitional component of hate speech, it makes little sense to relate them to the three crimes defined by the Penal Code. However, if we want to include in the category of hate speech the examples described of attacks on teachers or judges, it turns out that there is no clear line of demarcation between hate speech and insulting, defaming or threatening. This is particularly evident in the 30 cases treated by the ROP as the use of hate speech and is often far from xenophobia. In addition, it should be noted at this point that a threat is a common element that can be part of any of the behaviours in question; after all, it is easy to point out attacks built on the principle: insult plus threat, defamation plus threat, hateful message plus threat. However, from a linguistic point of view, a threat boils down to a performative act aimed at a specific perlocutionary action (causing fear in the recipient), as already mentioned. Thus, the main problem of distinction is drawn on the line: hate speech versus insult or defamation. In the analysis of hate speech, insult and defamation, the illocutionary content plays a primary role (what someone claims about someone, what predicates he attributes to him).³

If one can imagine defamation or insult caused by social considerations (e.g., insulting judges because of their immunity or teachers because they have long vacations), then, by the same token, one can see that it is difficult to clearly separate these concepts from hate speech related to social tensions. Therefore, it turns out that the latter intersects with the first two categories of the Penal Code. This may be the reason for the caution of lawmakers in introducing hate speech into the code’s provisions; these considerations may also have been dictated by the response of the government spokesman stating that Polish law already considers acts related to hate speech.

To outline the difference between hate speech and defamation and insult, it is necessary to focus on what is defined by these illocutionary acts; what the vocabulary used in them refer to; what they define. If we turn to typical examples of hate speech, the predominant reference will be to traits that a person is entitled to on the basis of biological inheritance or, which is unfortunately less precise, social.⁴ This refers to a person’s phenotypic traits (mainly skin color), followed by sexual orientation, national (ethnic) origin, religious beliefs (Adamczak-Krysztofowicz & Szczepaniak-Kozak, 2018). Instead of the term “phenotypic traits”, one could – so to speak, in accordance with cognitive definitions referring to the consciousness of the average language user – simply speak of appearance; after all, very often hate messages, in addition to skin colour, are based on the appearance of, for example, an Arab, a Jew, a Roma, and so on. Thus, one can observe a return to article 257 of the Penal Code, which covers the cases mentioned, as well as the issue of attacks caused by irreligiousness. From a legal perspective, it can be considered that hate speech is related to the acts indicated in the cited article of the Penal Code. And, although it is possible to find examples of the behaviour in question that go beyond the scope set out in this provision (the aforementioned attacks on teachers or lawyers, economically disadvantaged social groups), the prototype of the concept of hate speech is associated with attacking someone with a message based on the characteristics of the person being addressed⁵, based on properties

---

³ Here, of course, the term “predicate” means neither a sentence-forming functor nor a sentence function of name arguments, but an expression describing certain properties or relations.


⁵ It seems that in the case of hate speech it is better to talk about the addressee rather than the recipient – although these expressions are used interchangeably, “recipient” suggests reception, decoding, while “addressee”
perceptible to the senses and arousing hatred (possibly contempt, disdain, etc.) in the sender. In defining this prototype, reference was made to the colloquial knowledge of language users; therefore, for example, the expression “biologically or socially inherited traits” was considered not to belong to such an average conceptual distinction. Thus, a hater prototypically is simply guided by what appear to him to be striking physical features. Further on, the prototypical framing of hate speech will be analysed in detail.

A prototypical take on hate speech

The prototype of a hate speech act should be considered – based on examples – in the following situation: the sender (N) broadcasts a message (K) depicting the addressee (A)⁶ as someone contemptible due to the fact that this addressee is the bearer of a trait or traits (a, b, c) that N considers abhorrent (contempt, disdain, etc.) and which, according to N, belong to the essence of A, i.e., the life and functioning of A are inextricably linked in N’s opinion with the possession of this trait or traits. Usually, these traits are perceptible through the senses (black or swarthy skin, slanted eyes, facial appearance, etc.).

The described prototype does not cover other cases, which were cited earlier. To address them, it is necessary to resort to a cognitive tool referred to as an extension of the prototype. To do this, one looks for a conceptual scheme that allows one to capture other cases with the same prototype. This concept serves to organize the knowledge of individual concretisations of a given term (Taylor, 2007, pp. 559-560). It can be presented as follows: N expresses strongly negative feelings toward A in the message, because the latter is a non-occasional bearer of a trait or traits considered by N to arouse strong dislike (disgust). The analysed extension involves a meaning operation on the essential possession of the trait or traits in question (non-occasionality).

It seems reasonable to assume that the attribution of certain negative traits to someone within the framework of hate speech is associated – viewed from the perspective of the senders of such messages – with essentialism.⁷ According to the sender of such a message, the trait(s) indicated belong to the definition of the person attacked, to his immutable essence, and as such completely defines him. At this point, it is worth pointing out the difference between hate speech and insult or defamation. In the case of the latter, non-occasionality and essentialism usually do not appear, as they are referred to a specific situation that can change. After all, a headmistress accused of abusing her employees probably wouldn’t act in a sadistic manner if she didn’t hold that position; an insulted politician or public figure is attacked precisely by virtue of occupying a prominent place in social life. Besides, in the case of defamation, the sender is concerned with depriving the addressee of a certain function, while the occasional nature of the insult is particularly evident when it is directed at public officials (a court hearing, attacks on teachers or officials), in political life (elections, voting on emotionally stirring laws), during open meetings, exhibitions, etc., i.e.

⁶ To simplify this framing, we treat the person attacked by means of hate speech as the addressee, although of course it is possible and often happens that there is no return to the addressee at all; what appears instead is a scheme of the following type: the sender N broadcasts a message K in order to assign to the object O certain predicates, denoting the characteristics a, b, c that arouse disgust in N, with the object O being a person or a group of people.

⁷ We remain, of course, at the level of colloquial knowledge, since in modern methodology essentialism – except perhaps in phenomenology – is a marginal phenomenon.
in the context of some specific matters, characterised by variability against the background of the unchanging person of the addressee.

In the case of extending hate speech to non-prototypical phenomena, the indicated non-occasionality will play an important role. Already, the first such extension is suggested by article 257 of the Penal Code, speaking of religious beliefs or irreligiosity. Such traits tend to be more closely related to the person, penetrating more deeply into their life than a specific profession or social function.

You can see in the prototype expansion presented how the initial concept is becoming increasingly fuzzy and undefined. Can this be prevented? It seems that it can be done in two ways.

One can link acts of hate speech to article 257 of the Penal Code. Thus, the scope of meaning of this concept would extend from prototypical attacks caused by racism or nationalism (xenophobia) to manifestations of aggressive intolerance (religious or irreligious issues). In contrast, aggressiveness arising from, for example, economic reasons, related to certain professional groups, would be a manifestation of defamation. Defamation would not be considered in this context at all, due to the possibility of its being framed in terms of truth and falsehood. The proposed solution may offend with its formalism, but it allows us to maintain a clear distinction between the concepts of hate speech and insult.

Another solution would involve further linguistic analysis aimed at identifying at least one specific feature that characterises hate speech and differentiates it from the closest insult. At the level of illocution, there will be a subtle difference between the two, which will become apparent in the description. The description of hate speech may appear as follows: the sender speaks (communicates) in order to say that the addressee deserves contempt; the description of insult: the sender speaks (communicates) in order to let the addressee know that he deserves contempt.

The distinction between “to state” and “for the addressee to know” is linked to the fact that the insult is fully realised; that is, it becomes a fortuitous speech act when the addressee learns that someone has an extremely negative opinion of him or her. Hate speech messages, on the other hand, are primarily intended to maintain a radically negative attitude in a particular community toward a particular addressee, or even more often, addressees. In turn, the issue of whether the addressee or addressees themselves will learn about it does not come to the fore for the simple reason that hate messages may affect people who do not understand, for example, the language in which they are formulated (e.g., attacks on newly arrived refugees in a given country). Attacking using the partially universal language of gestures already falls into the category of active insult; therefore, applying the concept of hate speech in this case is not precise.

In the case of the addressee, on the other hand, one important distinguishing feature of hate speech emerges: the individual attacked in this way becomes the victim of such an attack because he is the bearer of certain characteristics (mainly skin colour, nationality, religion) that characterise the entire specified group. In this sense, the individuality of the addressee – a specific name, surname, etc. – recedes into the background, while the traits shared with a particular group come to the fore. This is why examples of hate speech in the media are usually pointed out by various types of moderators or third parties; in the case of insults, this makes little sense, since it is primarily the addressee inscribed in the message who is supposed to receive – with all the emotional effects – the message directed against him.

---

8 The reduction of the addressee to a bearer of the characteristics of a particular group is also pointed out by Calek et al., 2019, p. 30.
Thus, insult, being the speech act closest to hate speech, is distinguished from it in two respects. First, by the existence of a specified, individual addressee, who, at the level of perlocution, must understand that he has just been insulted. Secondly, this addressee has an individual and irreplaceable character, that is, in the nature of the communicative act is not a variable under which someone else can be substituted belonging to the set of people who are – as perceived by the sender – non-occasional bearers of a certain “inherent” characteristic (skin color, national origin, orientation – ciapaty, banderowiec, “gay”).

**Definition of hate speech**

Tables 1 and 2 are a summary of the considerations presented, but they only include hate speech and insult, since these – as indicated above – remain the communication activities closest to each other. In this view, hate speech is a communicative act attributing negative qualities to someone (more precisely – contemptible from the perspective of the sender), while the addressee himself does not need to understand it (he may be a refugee unfamiliar with the language of the hateful message). Thus, the sender does not have to directly affect the addressee (we can speak here of potential impressiveness, and not actualised as in the case of insult). Of course, the emergence of communicative intent – as in insult – is associated with the negative emotions of the sender.

**Table 1. Insult vs. hate speech in terms of speech act theory.**

<table>
<thead>
<tr>
<th>Type of act</th>
<th>Speech act element</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>illocution – intention, intent</td>
</tr>
<tr>
<td>Insult</td>
<td>informing the addressee of the sender’s negative perception of them</td>
</tr>
<tr>
<td></td>
<td>illocution – adjudication</td>
</tr>
<tr>
<td></td>
<td>predicates unrelated to any particular semantic field</td>
</tr>
<tr>
<td></td>
<td>perlocution</td>
</tr>
<tr>
<td></td>
<td>the addressee is supposed to feel humiliated, insulted, etc.</td>
</tr>
<tr>
<td>Hate speech</td>
<td>attributing negative characteristics to the addressee</td>
</tr>
<tr>
<td></td>
<td>predicates prototypically being an expression of xenophobia or intolerance</td>
</tr>
<tr>
<td></td>
<td>the addressee, the person being attacked, does not need to know about the resentment expressed against him or her</td>
</tr>
</tbody>
</table>

**Table 2. Insult vs. hate speech in relation to speech function and occasionality.**

<table>
<thead>
<tr>
<th>Type of act</th>
<th>The element of the communication situation and how it is implemented</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>emotionality</td>
</tr>
<tr>
<td>Insult</td>
<td>conditions the communicative intention</td>
</tr>
<tr>
<td>Hate speech</td>
<td>conditions the communicative intention</td>
</tr>
</tbody>
</table>

Hate speech can therefore be defined as follows: It is a type of speech act in which the sender, driven by negative emotions towards certain characteristics a, b or c, which he attributes as non-occasional to the addressee inscribed in the message, formulates a message P expressing contempt (dislike, hatred, etc.) towards the indicated characteristics a, b or c, towards the indicated characteristics a, b or c, while this message does not have to be formulated in a language that the
addressee understands, and the indicated non-occasional characteristic(s) are attributed to him as a typical representative of a specific collective prototype of an ethnic, national, religious or sexual nature (here already an extension supported, unfortunately, by observed communication practices). The definition uses the term “addressee inscribed in the message” rather than simply “addressee”, since the addressee does not need to understand what is being said to him.9 It is worth asking at this point, then, why someone formulates a hateful message when the recipient may not be able to respond to the attack.

An attempt to define the purpose of hate speech
As already indicated, in defamation, insult or criminal threat, the response of the addressee, i.e., the perlocutionary effect, turns out to be crucial. For this, one must feel – to put it mildly – discomfort and react somehow: point out the falsity of the accusation (defamation), feel offended or threatened and file a lawsuit or report the case to the police. In the case of hate speech, the perlocutionary effect turns out to be more complicated. Arguably, it is also about humiliating a person, but it seems that the social context of such messages should also be considered. Because of the already indicated characteristic of the addressee, who functions as a kind of *pars pro toto* of a specific group, acts of hate speech do not play out only between some specific sender and addressee, as they involve the collective.

The non-occasionality of the feature or features that evoke in the senders of hate messages a reaction of radical rejection, and the associated frequent pretextual treatment of a specific addressee (as *pars pro toto*) allow us to conclude that this kind of public communication is a denial of communication in its essence. After all, if this is supposed to be a connection, even on the basis of a “protocol of incompatibility”, then, in the case under analysis, it is more a matter of delimiting a space that, in the imagination of the sender of hate speech, is supposed to remain unaffected by the addressee or addressees of this type of communication. In the *Lebensraum* of the hateful sender – in this context, reaching for this infamous term seems particularly justified – the presence of what is imputed to the addressee is, as it were, *a priori* disallowed or abolished.10

Finally, it is worth adding that the messages that constitute hate speech are also of an identity nature, as they often form the ideological core of some social group, supporters of certain, usually extreme, views. It can be hypothesised that they constitute signals that gather supporters around some extreme idea. In this sense, they may resemble ancient tribal or ancestral calls, identifying members of such a group. This is observed in the activity of selected social media users. For example, in relation to some tension-provoking event or information, there is often a spontaneous polarisation; groups antagonistic to each other are formed; there is an exponential growth of comments around a selected radical comment. Only the intervention of a possible moderator can stop this process. For this reason, virtually all social media moderate users’ statements and have a policy of removing hateful messages (Derecho, 2021), and online portals suspend the ability to comment on articles (Onet is an example). This shows that hate speech has

---

9 In addition, hate speech often functions in closed communities, where it performs, among other things, the perfomatic function of affirming identity, constituting it in the face of shared “values”. One might be tempted to identify two subtypes of hate speech: explicitly expressed and derivative, but this is a topic for another article.

10 This is particularly evident in the slogan “Europe will be white or deserted” or the characterization of Ukrainians as bandits, resunas, etc. A similar expansion of one’s own space can be seen in the anti-Jewish slogans of 1968, the posters depicting the AKs as “backward dwarfs of reaction” and, of course, in all Nazi propaganda.
the character of actions that Gustav Le Bon associated with the crowd and the contagiousness characteristic of the spread of irrational information within it (Le Bon, 2009, p. 60 et seq.). The person employing for hate speech – consciously or not – seeks to make the idea of *Lebensraum* – a space impervious and impenetrable to the “bearers” of unoccasional and abhorrent qualities in the sender – spread within the reach of his message based exclusively on contagiousness. The intended effect is to resemble something that Thomas Mann portrayed with great insight, referring to the effects of Nazi hate speech:

“So much for the city’s image. But in its air there remained something of the state of minds in the last decades of the fifteenth century, something of the hysteria of the now extinct Middle Ages, something of a latent spiritual epidemic [...] and, however too bold it may sound, one can imagine that suddenly some St. Vitus dance will break out here. St. Vitus dance, a children’s crusade will set off, some simpleton’s visionary-communion sermons will begin, combined with burning at the stake the vanities of this world, stigma and mystical folk processions will appear. [...] Of course, nothing of the sort was happening [...] The police would not allow this to happen, in keeping with the era and its laws. And yet! What the police did not tolerate in silence in our days – and again in accordance with the era. [...] Our times themselves incline secretly, or by no means secretly [...] back toward those eras and enthusiastically repeat symbolic deeds that have something grim and insulting about them like a slap in the face to the spirit of modernity” (Mann, 2018, pp. 58-59).

Thus, in global terms, as an effect involving social relations, hate speech is “symbolic deeds” that reduce communication to tribal or identity exhortations, and thus imply the blocking (knockout or perhaps lockdown) of the possibility of communication.

**Bibliography**


