

CONVERSATIONAL IMPLICATURE IN THE DISCOURSE OF DIGITAL HATE TOWARDS BTS ON X

Irmala Sukendra^{1*}; Riska Khumairoh²

^{1,2}English Language Education Department, Faculty of Teacher Training and Education,
Syekh Yusuf Islamic University Tangerang
Banten, Indonesia 15118

¹isukendra@unis.ac.id; ²khumairohriska45@gmail.com

Received: 29th October 2025/Revised: 19th January 2026/Accepted: 18th February 2026

How to Cite: Sukendra, I., & Khumairoh, R. (2026). Conversational implicature in the discourse of digital hate towards BTS on X. *Lingua Cultura*, 20(1), 95-107. <https://doi.org/10.21512/lc.v20i1.14626>

ABSTRACT

This study aims to identify the predominant types, linguistic strategies, and perceived functions of hate speech targeted at the global music phenomenon BTS (Bangtan Sonyeondan) on X (formerly Twitter). This study used a descriptive qualitative approach, in which the data consisted of comments containing hate speech against BTS from January 2023 to March 2025, collected from various user accounts, both anonymous and public. Approximately 100 posts were analyzed to reflect the diversity of hate speech. Data collection was carried out using the documentation method, which involves gathering evidence in the form of text and images. We point out how online harassers use irony and metaphor as a shield, spreading hate while hiding in plain sight. The findings reveal that the most dominant form of hate speech is negative characterization, manifested through derogatory labels and stereotypes, followed by dehumanization and demonization. The most frequently deployed impoliteness strategies are positive impoliteness and bald-on-record direct attacks. The analysis indicates that conversational implicature is routinely exploited to veil hostile intent through sarcasm, metaphorical language, and irony. This result shows why platforms struggle to moderate implied malice, prompting a shift from policing words to understanding intent, which confirms that in high-profile, platform-mediated contexts, hate speech is not entirely explicit but rather systematically employs indirect methods. These results confirm that hate speech against BTS on X is not solely explicit but also operates through subtle, indirect mechanisms that function to damage reputations, foster exclusion, and circumvent platform moderation policies.

Keywords: BTS, conversational implicature, digital hate speech, impoliteness strategies, X (Twitter)

INTRODUCTION

Hate speech is defined as an expression that incites hatred, discrimination, or violence based on identity attributes like race, ethnicity, or gender, functions as a tool for ideological domination and social marginalization (Castaño-Pulgarín et al., 2021; Post, 2011, as cited in Siregar et al., 2023; Marques, 2023; and Brown, 2024). Its manifestations are diverse, ranging from direct attacks and stereotypes to dehumanizing rhetoric, which can trigger social exclusion and even violence (Maarouf et al., 2022; Schmitz et al., 2022; Fahdiansyah, 2024). Scholars

largely agree that this phenomenon is rooted in social bias and a desire to marginalize certain groups (Fonseca et al., 2024). Marques (2023) emphasizes the role of negative emotions in the formation of harmful social norms that govern the treatment of vulnerable groups. Thus, the consequences are severe, leading to the subordination and attack of individuals based on protected identities (Brown, 2024). Kang et al. (2023) assert that hate speech is not merely demeaning; it is shaped by historical and cultural factors that erode social cohesion. From an ideological perspective, Carlson (2021) frames it as a mechanism to maintain social hierarchy.

Studies on hate speech have focused on political issues (Faiq & Noori, 2022; Kaulika et al., 2024; Fauzi & Nursapia, 2024; and Sari et al., 2025), gender (Pasaribu, 2021; Godinez et al., 2022; Pornying & Tantiniranta, 2024), and religion (Prakoso & Runturabi, 2021; Sirait et al., 2025). The hate speech also tends to target specific groups or political figures (Pasaribu, 2021; Piazza & Van Dorren, 2023; Fauzi & Nursapia, 2024). Although in practice some popular personas may be targeted, this usually happens at a specific time rather than repeatedly. In the literature on hate speech on social media, various studies have explored its forms, strategies, and pragmatic meanings across diverse digital contexts. In relation to that, Bahador (2023) offers a classification of hate speech intensity, ranging from mild to extreme, dividing it into six categories across three typologies based on intensity. From a linguistic perspective, Culpeper (1996, as cited in Guillén-Nieto, 2023) explains that hate speech often employs impoliteness strategies to damage the interlocutor's face. To add on, Grice (1975, as cited in Guillén-Nieto, 2023) distinguishes between two types of implicature: conventional implicature, which arises from the fixed meaning of a word or phrase, such as the use of pejorative terms that directly imply hatred, and conversational implicature depends on context and the principle of cooperation, where hateful messages are conveyed implicitly through metaphors, irony, or sarcasm. Hate speech can also be used as a tool to dehumanize groups, normalizing violence against them by amplifying egregious claims through social media (Carlson, 2021).

Social networking sites gained popularity in the early 21st century and emerged to allow family, friends, and acquaintances to connect online. Initially, *Facebook* became one of the most popular social networking sites, with billions of users worldwide. Other social media platforms, such as X (formerly *Twitter*), *Instagram*, and *Pinterest*, also have numerous users. Diverse forms of social media also emerged for sharing specific types of content. For example, *YouTube* allows users to share videos, *LinkedIn* emphasizes professional relationships and offers pages structured like a resume, while *TikTok* is designed specifically for sharing short videos. In recent years, social media has become the primary space for interaction, enabling users to build identities, navigate social norms, and express emotions without geographical or time zone boundaries (Waziri, 2023; Wajahat, 2024).

One of the platforms most vulnerable to this phenomenon is X (formerly *Twitter*), which plays an important role in the spread of hate speech. The platform's open structure, lack of restrictions on friendship, and features such as hashtags, retweets, and quote tweets accelerate the circulation of provocative messages (Castaño-Pulgarín et al., 2021; Fonseca et al., 2024; Rodríguez-Peral et al., 2025). Although often praised for providing unprecedented freedom of expression, this freedom operates without control, leading to issues such as misunderstandings, shifts in emotional meaning, and the spread of aggressive

discourse (Singh, 2022; Candrasari, 2021). This situation is exacerbated by user anonymity, algorithmic bias, and the formation of echo chambers that make hate speech easier to produce and accept (Aljasir, 2023; Carlson, 2021). It shows that social media is not only a space for expression but also a fertile ground for the proliferation of hate speech.

Additionally, algorithms that prioritize emotional content, especially angry or divisive content, further amplify its viral impact and deepen digital polarization (Carlson, 2021). It makes X an essential arena for examining how hate speech operates. Hate speech on X often appears in multimodal forms, ranging from text to memes and images (Toktarova et al., 2023). Hate speech on X can also take the form of conversations (Fonseca et al., 2024) and hashtags (Ayalew & Ayalew, 2024; Orgeret et al., 2025). The rate of hate speech on X is also approximately 50 percent higher for several months after Elon Musk purchased the social media platform than in prior months (Hickey et al., 2025).

Although previous studies provide valuable insights, most of the research tends to focus on one dimension of hate speech, whether its types (Paasch-Colberg et al., 2021; Faiq and Noori, 2022), linguistic strategies (Pasaribu, 2021; García-Díaz et al., 2023; Karapetjana, 2023), or communicative functions (Barth et al., 2023; Sigalingging et al., 2025). Moreover, the object of hate speech generally focuses on political figures or public issues, as can be found in Faiq and Noori (2022), who focused on political comments on Facebook from English and Kurdish-speaking users, Kaulika et al. (2024) analyzed hate comments against Noah Schnapp's political views on Instagram, and Pasaribu (2021), who examined hate speech against President Joko Widodo on Facebook based on gender differences. Other studies focus on other aspects, such as gender (Pornying & Tantiniranat, 2024; Nascimento et al., 2025) and identity (Bianchi, 2024; Margono et al., 2024).

Futhermore, the phenomenon of hate speech on X becomes increasingly evident when directed at public figures, particularly celebrities with global popularity. Global entertainment figures, especially in the context of digital fan culture, are still rarely studied in depth. One of the most prominent examples is BTS (Bangtan Sonyeondan). This South Korean boy band gained worldwide fame and was formed in 2013 by Big Hit Entertainment with seven (7) members: Kim Nam Joon or RM (previously Rap Monster), Kim Seok Jin or Jin, Min Yong Gi or Suga, Jung Ho Seok or J-Hope, Park Ji Min, Kim Tae Hyung or V, and Jeon Jung Kook. Despite BTS's global success, including becoming the first Asian artist to top the Billboard Hot 100 and winning numerous international awards (Liang, 2023; Khilmia, 2024), they are often the primary targets of hate speech on social media. Hate speech directed at BTS usually stems from the competitive dynamics of the K-pop industry, which fosters constant comparisons with other artists, sparking digital conflicts and fan wars (Pita & Widiensyah, 2023). Letwory and Sihombing (2023) emphasize that these dynamics

can lead to flaming, disinformation, and systematic efforts to discredit competing artists. Furthermore, BTS is also the target of cultural prejudice and racism, where Asian cultural representation is often viewed as inferior to Western culture (Virginia et al., 2023).

Therefore, this study aims to fill this gap by presenting an integrated linguistic analysis of hate speech against BTS on X. This research combines three main analytical frameworks, namely the classification of hate speech types based on Bahador's typology (2023), impoliteness strategies according to Culpeper (1996, in Guillén-Nieto, 2023), and pragmatic functions based on Grice's implicature theory (1975, as cited in Bezugla et al., 2023). The data consists of posts targeting BTS, both individually and collectively. This study also intends to identify anti-fan users. This integrative approach is expected to provide a comprehensive understanding of how hate speech is constructed, conveyed, and interpreted in digital spaces, while enriching academic discourse on online hate speech, influenced by platform culture and the status of global celebrities.

METHODS

This research employs a descriptive qualitative design to understand hate speech in its natural context, specifically the types, strategies, and functions of hate speech directed towards BTS on X. This design was chosen because it allows us to deeply explore the meaning and complexity of hate speech (Furidha, 2024). The data consists of approximately 100 posts and comment threads containing hate speech directed at BTS, taken from various user accounts on X, both anonymous and public, from January 2023 to March 2025. Data collection was carried out using the documentation method, which involves gathering evidence in the form of text and images from relevant comments and uploads (Creswell, 2014, as cited in Daruhadi & Sopiaty, 2024). Each data point is recorded with contextual information, such as the upload time and account identity, to maintain the integrity of the analysis context. Data analysis for this study implemented the five stages outlined by Bingham (2023) as follows: (1) organizing raw data in the form of screenshots, (2) constructing a structured dataset, (3) performing theory-based open coding, (4) interpreting results through classification tables, and (5) explaining the findings within a theoretical framework.

This study adopts Bahador's (2023) Intense Hate Speech Scale to identify types of hate speech, Culpeper's (1996, as cited in Regaro, 2023) impoliteness strategies to analyze linguistic forms, and Grice's (1975, as cited in Bezugla et al., 2023) theory of implicature to determine communicative functions. The results of the analysis were later integrated into a comprehensive classification table, and the findings were explained within the theoretical framework to answer the research questions. To ensure research validity, source triangulation was

conducted by comparing hate speech across different accounts. In contrast, theory triangulation was achieved by applying the three analytical frameworks simultaneously (Denzin, 1978, as cited in Messaoudi & Sakale, 2023). This approach enhanced the reliability and accountability of the findings. In terms of user categorization, X accounts are classified based on usernames and profile pictures.

RESULTS AND DISCUSSIONS

The findings describe the various types, strategies, and pragmatic functions of hate speech practiced by social media users on X targeting BTS. A total of 100 English-language comments and posts were analyzed to reveal how hate speech is constructed, expressed, and interpreted in the digital world. Although delivered asynchronously and text-based, these utterances share the same communicative purpose as oral interactions: to reject, belittle, and socially exclude the target through language.

In this research, we find five (5) types of hate speech against BTS on X: (1) disagreement, (2) non-violent negative actions, (3) negative characterization or insults, (4) dehumanization and demonization, and (5) violence. Of the six categories proposed by Bahador (2023), this study fails to include the sixth category: Death. Disagreement emerged as mild rejection, without a hostile tone, indicating that users are more likely to express hatred offensively than to offer neutral criticism. Non-violent negative actions, such as calls for the group's disbandment, while not abusive, still aim to damage BTS's public reputation. Negative characterization or insults were the most dominant category, marked by demeaning labels and stereotypes directed at BTS's appearance and status. Table 1 depicts the types of hate speech:

Table 1 Types of Hate Speech Against BTS on X

Types of Hate Speech	Description	Frequency
Disagreement	Expressing rejection or denial without harsh language	3
Non-violent negative action	Accusing immoral or dishonest behavior without violence	14
Negative characterization or insult	Using derogatory labels or mocking language	57
Dehumanization or demonization	Compared to animals or unnatural entities	24
Violence	Calling for punishment or exclusion	2
TOTAL		100

In disagreement, hate speech is characterized by expressions of disagreement towards the target's actions or existence, without necessarily including direct insults or violent language. This type is classified by Bahador (2023) as an early warning sign of hate speech, in which a user expresses disagreement with the views, actions, or existence of a person or group without using insults, coarse language, or calls for violence. In Figure 1, the comment "Please, we don't want it, just stop!!" falls into the disagreement category. This comment rejects Taehyung's solo debut, with the phrases "don't want it" and "just stop!!" indicating that the user is pleading for the action not to be done or to be put on hold as a sign of disagreement. This statement remains a matter of difference of opinion, rather than a personal insult or incitement to violence.



Figure 1 Disagreement Expressing Rejection Without Insult

In the Non-Violent Negative Action category, hate speech is portrayed by accusations against negative actions without direct insults or violence. According to Bahador (2023), this type involves accusations against external groups related to non-violent negative actions, such as theft or betrayal, without any explicit call for violence. In Figure 2, the comment "Mass streaming, Mass buying, Not organic at all" is categorized as *Non-Violent Negative Action*. The phrase "Not organic at all" reinforces the view that their popularity is seen as impure. It accused BTS of having foul achievements and of buying their popularity and awards. It attributes BTS's success to manipulation through mass streaming and mass buying practices. This statement serves to delegitimize BTS's achievements and damage their credibility, thus falling under the category of Non-Violent Negative Action. This category includes accusations against other groups related to non-violent negative actions, such as manipulation or cheating, without any incitement to violence.



Figure 2 Non-Violent Negative Action Through the Manipulation Chart



Figure 3 Negative Characterization Implies Criticism Without Direct Aggression

The third category, Negative Characterization or Insult, is employed through derogatory labels, stereotypes, and direct insults as defined by Bahador (2023). This type involves using stereotypes or general insults against a specific group, such as calling them immoral or attacking the group's fundamental nature, rather than just individual behavior. In Figure 3, the comment "The obsession and desperation" falls under *Negative Characterization/Insult*, which refers to forms of insult or negative labeling of an individual's character, such as calling them desperate or immoral. The phrase "obsession" implies an excessive or

unreasonable trait, while “*desperation*” describes an emotional state that is considered low. Both phrases are directed at Jungkook to belittle his personal image, not to criticize his work or musical activities.

In Figure 4, the statement “*FINALLY PEACE, MUSIC TO MY EARS*” is categorized as Negative Characterization/Insult. The phrase “*finally peace*” implies that BTS's presence is a source of disturbance or noise, while “*music to my ears*” describes their absence as a relief. The sentence implies that the absence of BTS means they do not have to hear BTS music, and that BTS music is a nuisance for them. This negative statement portrays BTS not only as something disliked but also as something disturbing. In line with this, Faiq & Noori (2022) claim that this category is more harmful than simply negative, non-violent actions, which are destructive because they imply something fundamental about the entire group, not just an individual incident.



Figure 4 Negative Characterization Using Derogatory Remarks

Within dehumanization/demonization, hate speech labels the target as an evil entity or moral threat that must be eliminated, thus declaring their humanity lost. Referring to Bahador (2023), this category includes more severe forms of hate speech. Dehumanization depicts the target group as non-human entities, such as animals or diseases, thereby stripping away their humanity. Demonization, on the other hand, portrays the out-group as a significant threat or monster that must be eradicated for the in-group's survival. In Figure 5, the comment “*THE EVILS HAS BEEN DEFEATED*” falls under this category. The phrase “*the evils*” positions BTS as the embodiment of evil, while “*has been defeated*” implies that they are enemies who must be defeated. The statement describes the target as a non-human entity (dehumanization) or

as a symbol of a major threat/evil (demonization). The remark is a blunt accusation of the group and regards them as less than human. BTS is positioned as a demonic symbol that must be overpowered. Paasch-Colberg et al. (2021) proposed that in this category, the targets of hate speech are compared or referred to as non-human things or beings that are connoted as inferior, disgusting, or even dangerous. In accordance with this, Faiq & Noori (2022) claimed that in this category, hate speech is employed to view the target as a hated single creature that has been stripped of its own identity and humanity. Thus, this statement is not merely insulting, but labels BTS as dangerous moral enemies.



Figure 5 Dehumanization by Labeling BTS as ‘Evil’



Figure 6 Violence Through Attacks on Existence

In violence, hate speech is characterized by the normalization of harmful physical actions, death, or violence against individuals or groups. According to Bahador (2023), this type of hate speech is the form of hate speech that explicitly encourages discriminatory actions or physical violence against specific groups

and is usually couched in narratives of self-defense or justice. In Figure 6, the comment *“Fat, ugly and criminal. His mother should have aborted him, I’m sure she is very disappointed with him”* is categorized as violence, where the target is not only physically and morally insulted, but also denied the right to live. The phrase *“should have aborted him”* explicitly rejects the target’s existence, while the words *“fat, ugly, criminal”* attack physical and moral aspects.

In this study, violence is also achieved through aggression on the target’s existence or through demanding the target’s death. Bahador (2023) claimed this category comprises metaphorical or aspirational violence that is either non-lethal or lethal and that the responses require literal, non-lethal hostility against out-group members, such as assault. The comment *“I paid for a psychic on Etsy, and they said he will probably die on the frontlines”* in Figure 9 falls under this category. The phrase *“die on the frontlines”* normalizes the death of the artist as something to be expected, even though it is conveyed in a sarcastic or humorous tone. This comment clearly implies a desire for the target’s death.



Figure 7 Violence Through Death Wish

Although the sample in Figure 7 wishes the target of the hate speech to die, it is considered as violence and not on death, the sixth category in Bahador’s hate speech intensity scale (2023). The sentence only wishes or appeals to the target’s death and does not describe any physical action of killing or termination of the hate speech target. The *Death* category, as Bahador explained (2023), comprises the physical annihilation of a group, which evokes reactions favoring its complete annihilation. We failed to find any hate speech against BTS that suits this

category in this study. Thus, it can be concluded that for this study, only five (5) types of hate speech are found as death is considered serious and can lead to a potentially criminal investigation.

To answer the question what linguistic strategies of the hate speech practiced by users in comments and posts on X against BTS, this study applied the four types of impoliteness strategy from Culpeper (1996, in Guillén-Nieto, 2023): (1) Bald on Record Impoliteness, (2) Positive Impoliteness, (3) Negative Impoliteness, (4) Sarcasm or Mock Politeness, and (5) Withholding Politeness (see Table 2). Out of the five strategies, only **Withholding Politeness** was not identified in the data. This absence indicates that while hate speech on X often takes direct, sarcastic, and degrading forms, it rarely manifests through the subtler act of refusing expected politeness.

This study found that one post or comment may contain multiple strategies. The overall frequency of strategies exceeds the number of individual comments, showing the layered nature of online verbal hostility. The findings showed that bald-on-record impoliteness openly displays insults or offensive labels. In contrast, positive Impoliteness aims to damage a person’s positive face, which is their desire to be valued and socially accepted. Negative impoliteness attacks the target’s personal space or freedom through comments that violate privacy, whereas sarcasm or mock politeness uses irony or humor to disguise insulting intentions. Overall, these findings highlight the layered and complex nature of hate speech on digital platforms, where users employ multiple linguistic strategies to express hostility in both overt and covert ways. Table 2 depicts the linguistic strategies of hate speech against BTS found in X.

Table 2 Impoliteness Strategy in Hate Speech Against BTS on X

Strategy	Key features	Frequency
Bald on Record impoliteness	Direct insult without mitigation	42
Positive impoliteness	Attacking social image or using name-play	52
Negative impoliteness	Threatening or removing autonomy	12
Sarcasm or mock Politeness	Insincere compliments with hidden	19
Withholding politeness	Absence of expected empathy	-
Total		125

In the bald-on-record impoliteness category, social media users directly attacked BTS without attempting to soften or disguise their intentions.

According to Culpeper (1996, in Guillén-Nieto, 2023), this strategy is used directly and without mitigation, usually through open insults or explicit verbal attacks. In Figure 8, the comment “*THEY LOOK DISGUSTING*” uses the Bald on Record Impoliteness strategy. The phrase “*look disgusting*” is a crude expression of disgust aimed directly at BTS's appearance. The insult is conveyed bluntly without any attempt to soften its meaning.

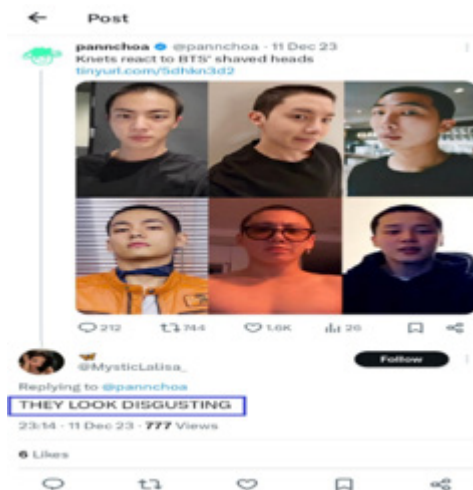


Figure 8 Bald on Record Impoliteness with Direct Insult

The comment “*nobody will ever want this pork face*” in Figure 9 also falls under the bald-on-record impoliteness. This strategy works through blunt insults, without any niceties or forms of politeness. The phrase “*pork face*” compares the target's face to a pig, which means fat or ugly. This phrase implies comparing a person's face to a pig's. This insult is crude and explicit, with no attempt to soften the meaning, thus fitting the bald-on-record impoliteness.

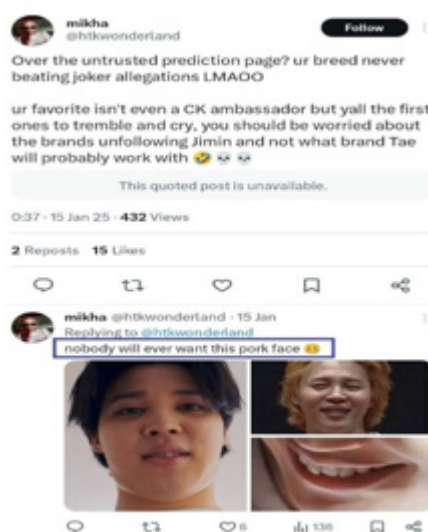


Figure 9 Bald on Record Impoliteness Using Crude Animal Comparison

In positive impoliteness, social media users damage BTS's positive image through mockery, insults, or identity harassment. Subyantoro & Apriyanto (2020) explain that in positive impoliteness, social media users attempt to undermine the target's positive face through derogatory and offensive expressions that question the target's social value or worth. This strategy can be achieved by ignoring, excluding, or showing disinterest. In Figure 10, the comment “*HagJIN extends his records of FAILING 90% of goals AGAIN with his debut album 'HAPPY' and title track 'Running Wild.'* He joins Ratmonster and JhOrse” uses the Positive Impoliteness strategy. This strategy aims to damage a person's positive face, the desire to be respected and socially accepted. The phrases “*HagJIN*,” “*Ratmonster*,” and “*JhOrse*” are derogatory nicknames that offend the members' personal identities. The addition of ‘*hag*’, ‘*rat*’, and ‘*horse*’ to the BTS members' names is intentional, meant to mock their names and make a pun. It is interesting as Jin is associated with a hag, while the other two members are associated with animals. Additionally, the clause “*FAILED 90% of goals*” mocks Jin's achievements by highlighting his failures. This statement demeans the dignity of the BTS members by insulting their abilities and identities, thus appropriately classified as positive impoliteness.



Figure 10 Positive Impoliteness, Mocking Failures, and Insulting with Derogatory Nicknames

In negative impoliteness, social media users attack basic rights or individual freedoms, such as autonomy, security, or personal space, to intimidate

or belittle the target. Regaro (2023) reveals that in negative impoliteness, online media users often employ linguistic strategies that aim to attack the autonomy, authority, and dignity of their interlocutors. The comment “*He should get life in prison*” in Figure 11 uses this strategy. This statement portrays BTS as criminals who do not deserve leniency. The phrase “*life in prison*” denies the target any possibility of freedom, as if they deserve the most severe punishment. This finding is in line with Manik et al. (2022), who found that, in negative impoliteness, speakers often attack the negative face of their target by depriving them of autonomy and freedom through demeaning and judgmental comments. Therefore, the comment in Figure 11 not only expresses verbal hatred but also negates the target's humanity and individual freedom, portraying them as guilty figures who do not deserve compassion or social respect.



Figure 11 Negative Impoliteness Threatens Freedom

In sarcasm or mock politeness, social media users indirectly convey insults or mockery through statements that appear neutral or positive (Culpeper, 1996, as cited in Guillén-Nieto, 2023). This strategy involves using insincere politeness, in which the speaker intentionally conveys polite expressions with an implied meaning that is sarcastic or insulting. In Figure 12, the comment “*He breaks the record for drunkest K-pop DUI case. Congratulations on your first ever record!*” also uses the sarcasm or Mock Politeness strategy. This strategy conveys praise or friendliness that appears positive, but the actual intention is to mock. The phrase “*Congratulations on your first-ever record*” seems like a congratulatory remark. However, in this context, it is actually used to mock negative behavior, namely the DUI (Drunk Under Influence) case. The difference between the form (congratulatory remark) and the actual intention

(mockery) indicates sarcasm, so this utterance is categorized as a Sarcasm/Mock Politeness strategy. Situmorang et al. (2024) argue that this form of impoliteness allows users to express hatred indirectly through irony, mockery, and insincere praise, often used to make the target appear foolish or inferior. This strategy demonstrates how netizens hide hatred behind humorous or sarcastic tones, reinforcing the impact of verbal attacks while appearing intelligent or harmless.



Figure 12 Sarcasm that Portrays Violations as Achievements

The functions of hate speech are understood through Grice's (1975, as cited in Guillén-Nieto, 2023) theory of implicature, namely (1) conventional implicature and (2) conversational implicature. Both of these implicatures are found in hate speech against BTS on X. Conversational implicature hides hostility behind seemingly neutral statements, depending on context, cultural references, or sarcastic tone, making it an indirect form of hostility that avoids detection. Conventional implicature arises when hatred is expressed directly through offensive language or overt insults, indicating an aggressive, confrontational style. These findings emphasize that users manipulate language to damage BTS's reputation and social image, both covertly and openly. In the conventional implicature category, social media users convey hatred through specific structures or word choices, allowing the meaning to be understood without additional context. According to Grice (1975, as cited in Guillén-Nieto, 2023), Conventional implicature arises from the natural or conventional meaning of the words and expressions used. In the case of hate speech, this type of implicature serves to identify hate speech that is delivered explicitly. Table 3 depicts the Communicative Functions of hate speech against BTS on X.

“*He can bark, he is talented.*” In Figure 13, it is categorized as conventional implicature, which arises from the natural meaning of the words/phrases used. The sentence “*He can bark*” assumes that the

target of the hate speech can bark (i.e., make a loud, abrupt noise like a dog). Alternatively, “to bark” can be defined as speaking loudly and usually belligerently without knowing what you are talking about.

Table 3 Communicative Function of Hate Speech Against BTS on X

Types of implicature	Description	Frequency
Conventional implicature	Meaning tied to word choice or sentence structure.	26
Conversational implicature	Meaning derived from context, irony, or shared knowledge.	74
TOTAL		100



Figure 13 Conventional Implicature with Sarcastic Dehumanization

This sentence can then be interpreted as a mockery of the target, allowing them to say whatever they please, since it will not mean anything. Additionally, the word ‘talented’ is used sarcastically to mock his talent. The hate speech is conveyed explicitly because this insult arises directly from the conventional meaning of the sentences in the comment.

Conventional implicature is likely to be associated with the meanings of particular words rather than with conversational outcomes (Bach, 1999, in Samitharathana, 2022). It is generally understood that everyone already knows the meaning of a certain thing. In Figure 14, the comment “Just disband that g4rbage group” also falls under conventional implicature. The hatred in this comment is expressed overtly in common, easily understood language. It employs descriptive words with negative connotations in certain contexts. The phrase “garbage group” directly labels BTS as trash, meaning worthless or unworthy. Additionally, the phrase “just disband” directly urges the band to break up.

In the conversational implicature category, hostility is conveyed indirectly through context, tone,

or implied meaning, rather than through explicit words. Referring to Grice (1975, as cited in Guillén-Nieto, 2023), conversational implicature refers to non-literal (non-conventional) meanings that arise from the context of the conversation and the principles of cooperation between speakers and listeners. This type of implicature is relevant to understanding hate speech delivered implicitly, i.e., when the meaning of hatred is not stated directly but implied through contextual clues, situational cues, or shared knowledge. Figure 15 has the comment “With what face they will return after all these dirty tricks came out of the company and nazist overnationalist racist misogynistic money-spoilt attitude of the group,” which refers to non-conventional meanings understood from the context of the conversation or shared knowledge, not just the literal meaning of the words. The phrase “dirty tricks came out of the company” refers to accusations that the record label artificially inflated BTS's achievements, despite BTS's global popularity and success, which surpass those of its peers. This allegation is widely recognized among dedicated K-pop fans but is not generally known to the broader public. Whereas the series of indictments, such as “Nazi, racist, misogynistic, corrupt,” implicitly formed a negative image. The meaning of hatred arises not solely from the words themselves, but from the context and shared assumptions. Thus, the comment in Figure 15 falls under the function of conversational implicature.



Figure 14 Conventional Implicature Demanding Disbandment and Labeling BTS as “A Garbage Group”



Figure 15 Conversational Implicature Delegitimizing BTS's Comeback Through Accusations of Corruption and Immorality

In this study, we found that hate speech against BTS on X appears as a form of social interaction in which users utilize language to express dislike, mockery, and belittlement of the target's social image. This kind of speech demonstrates complex linguistic patterns through various types, strategies, and functions used by social media users. At the type level, the most dominant form is Negative Characterization, in which users assign labels, stereotypes, or insults that directly highlight BTS members' identities and appearances. It shows that personal attacks are the most frequent form of hate speech. Dehumanization/demonization is present through metaphors that compare BTS to animals, serving to strip them of their humanity and create an image of inferiority. Finally, violence was found in several comments, often delivered with sarcasm, but implying legal or social consequences. The absence of death indicates that although hate speech against BTS on X is widespread and often severe, it does not escalate to the most extreme form of explicitly promoting deadly violence. Overall, these findings suggest that hate speech against BTS does not stop at mild criticism but escalates towards insults, dehumanization, and even calls for real-world consequences beyond social media. These findings are consistent with research by Faiq and Noori (2022), who also found that Negative Characterization is the most prevalent form of hate speech in English-language political comments on Facebook. This similarity shows that hate speech that attacks personal or group identities remains a major pattern across platforms, both in political and entertainment contexts.

The attacks were not only explicit but also covert through linguistic strategies. In terms of strategy, the most dominant form was Positive Impoliteness. This strategy targeted the positive face of the victim, namely the desire to be respected and recognized, by mocking, giving derogatory nicknames, or questioning their reputation. These findings differ from Pasaribu's (2021) research, which found Bald on Record Impoliteness to be the most dominant strategy in gender-related hate speech on Facebook. This difference shows that the choice of hate strategy is greatly influenced by the

social context, type of platform, and character of the target being attacked.

Furthermore, in terms of function, hate speech against BTS on X also demonstrates the complexity of language. The most dominant form is Conversational Implicature. This function allows users to convey hatred implicitly through sarcasm, insinuation, or rhetorical questions, so that the reader captures the negative meaning without explicit expression. This pattern is consistent with Sigalingging et al. (2025), who also found Conversational Implicature to be dominant in hate speech on TikTok. This similarity confirms that social media tends to facilitate subtle forms of hate speech, which are even more dangerous because they are difficult to identify and moderate.

Thus, the findings of this study confirm that hate speech against BTS on X reflects the dynamics of the digital world, where language is used as a tool to negotiate power, opinion, and social identity. The combination of Bahador, Culpeper, and Grice's perspectives shows that linguistic hatred is not only present in direct attacks but also in subtler forms that reinforce the polarization between pro- and anti-BTS sentiments in the digital space.

CONCLUSIONS

This research concludes that Hate speech against BTS on the X platform appears both explicitly and implicitly, through a combination of negative content, delivery strategies, and intentional meaning. The most dominant form of hate speech is negative characterization, where users provide labels, stereotypes, or insults that highlight the identity and appearance of BTS members, which is weaponized through a strategic fusion of conventional and conversational implicature. Linguistic strategies such as positive impoliteness, bald-on-record, sarcasm, negative impoliteness, and withholding politeness are used to reinforce messages of hate, both overtly and covertly. They are not random insults but calculated tools for social positioning. The function of hate speech also highlights the complexity of language: conversational implicature conveys hatred implicitly through sarcasm or ambiguous statements. In contrast, conventional implicature uses offensive expressions that are conventionally understood.

Metaphors, sarcasm, and irony are recurring linguistic devices that allow users to disguise hostility while simultaneously affirming solidarity with like-minded audiences. Metaphors compare BTS to demeaning or dehumanizing entities, sarcasm delivers false praise that conceals mockery, and irony feigns support but subtly belittles through contrast or hyperbole. This finding confirms that hate speech not only conveys dislike but also serves to negotiate opinions, social hierarchies, and public perception of the target. The combination of language types, strategies, and functions demonstrates how users manipulate language to influence BTS's social image,

while also affirming their position as public observers and commentators within the global entertainment ecosystem. This research, therefore, positions digital hate speech not as a mere emotional outburst but as a deliberate sociolinguistic strategy with significant implications for how public opinion is shaped, fandom is policed, and cultural power is contested in the digital sphere.

This study provides new contributions to understanding the linguistic and social dynamics of hate speech against public figures on digital platforms, particularly in the context of global music groups. However, this research is limited within a specific timeline. The integrity of the dataset is also limited by its source, as the dynamic nature of social media can lead to the deletion or alteration of content. Therefore, further research may focus on different subjects and would benefit from a longitudinal framework, secure data archiving, and an expanded scope encompassing a wider range of populations and contexts.

Author Contributions: Conceived and designed the analysis, I. S.; Collected the data, R. K.; Contributed data or analysis tools, I. S., and R. K.; Performed the analysis, I. S., and R. K.; Wrote the paper, I. S., and R. K.

Data Availability Statement: Data available on request from the authors. The data that support the findings of the research are available from the corresponding author, [I. S.], upon reasonable request. Reasons: 1. The data collection and categorization were undergone several steps so there are many files, 2. The files are large and compressing them may affect the quality of the files.

REFERENCES

- Aljasir, S. (2023). Effect of online civic intervention and online disinhibition on online hate speech among digital media users. *Online Journal of Communication and Media Technologies*, 13(4), e202334. <https://doi.org/10.30935/ojcm/13478>
- Ayalew, Y. E., & Ayalew, A. B. (2024). On selfies and hashtags: Disinformation during armed conflict in Ethiopia. In T. Roberts & G. H. Karekwaivanane (Eds.), *Digital disinformation in Africa: Hashtag politics, power and propaganda* (pp. 79–98). Zed Books.
- Bahador, B. (2023). Monitoring hate speech and the limits of current definition. In C. Strippel, S. Paasch-Colberg, M. Emmer, & J. Trebbe (Eds.), *Challenges and perspectives of hate speech research* (pp. 291–298). Digital communication Research. <https://doi.org/10.48541/dcr.v12.17>
- Barth, N., Wagner, E., Raab, P., & Wiegärtner, B. (2023). Contextures of hate: Towards a systems theory of hate communication on social media platforms. *The Communication Review*, 26(3), 209–252. <https://doi.org/10.1080/10714421.2023.2247455>
- Bezugla, L., Govorukha, N., Kryvoruchko, S., & Malaya, O. (2023). Conventional implicatures in Ukrainian discourse. *Theory and Practice in Language Studies*, 13(5), 1250–1258. <https://doi.org/10.17507/tpls.1305.20>
- Bianchi, C. (2024). Call me by my name: Hate speech and identity. In S. Cruschina & C. Gianollo (Eds.), *An investigation of hate speech in Italian: Use, identification, and perception*, 27–54. Helsinki University Press. <https://doi.org/10.33134/HUP-27-2>
- Bingham, A. J. (2023). From data management to actionable findings: A five-phase process of qualitative data analysis. *International Journal of Qualitative Methods*, 22, 1–11. <https://doi.org/10.1177/16094069231183620>
- Brown, A. (2024). Reverse hate speech, pragmatics, and the authority problem. *Philosophy and Social Criticism*, 52(5), 813–843. <https://doi.org/10.1177/01914537241308129>
- Candrasari, Y. (2021). Nonverbal communication reconstruction on Facebook. *Jurnal The Messenger*, 13(1), 94–107. <https://doi.org/10.26623/themessenger.v13i1.1000>
- Carlson, C. R. (2021). *Hate speech*. MIT Press.
- Castaño-Pulgarín, S. A., Suárez-Betancur, N., Vega, L. M. T., & López, H. M. H. (2021). Internet, social media, and online hate speech: Systematic review. *Aggression and Violent Behavior*, 58, 101608. <https://doi.org/10.1016/j.avb.2021.101608>
- Daruhadi, G., & Sopiati, P. (2024). Research Data Collection. *International Journal of Social Service Research*, 4(7), 1–18. <https://doi.org/10.46799/ijssr.v4i7.863>
- Fahdiansyah, M. (2024). Sociopragmatic analysis of hate speech in Hilarion Heagy's conversion news on Twitter. *Celtic: A Journal of Culture*, 11(1), 12–22. <https://doi.org/10.22219/celtic.v11i1.30342>
- Faiq, A. M., & Noori, M. S. (2022). Hate speech forms and implications in English and Kurdish social media. *Journal of University of Human Development*, 8(4), 80–88. <https://doi.org/10.21928/juhd.v8n4y2022>
- Fauzi, W., & Nursapia, H. (2024). Analysis of hate speech on social media (Twitter and Instagram) against the 2024 presidential candidate of The Republic of Indonesia. *International Journal of Humanities Education and Social Sciences*, 3(4), 1865–1876. <https://doi.org/10.55227/ijhess.v3i4.86>
- Fonseca, A., Pontes, C., Moro, S., Batista, F., Ribeiro, R., Guerra, R., Carvalho, P., Marques, C., & Silva, C. (2024). Analyzing hate speech dynamics on Twitter/X: Insights from conversational data and the impact of user interaction patterns. *Heliyon*, 10(11), e32246. <https://doi.org/10.1016/j.heliyon.2024.e32246>
- Furidha, B. W. (2024). Comprehension of the descriptive qualitative research method: A critical assessment of the literature. *Journal of Multidisciplinary Research*, 2(4), 1–8. <https://doi.org/10.56943/jmr.v2i4.443>
- García-Díaz, J. A., Jiménez-Zafra, S. M., García-Cumbreras, M. A., & Valencia-García, R. (2023).

- Evaluating feature combination strategies for hate-speech detection in Spanish using linguistic features and transformers. *Complex & Intelligent Systems*, 9(3), 2893–2914. <https://doi.org/10.1007/s40747-021-00590-9>
- Godinez, P., Rico, S., & Sarikakis, K. (2022). Gender-based hate speech: Contributions to the global policy debate from Latin America. *International Journal of Communication*, 16(2022), 4758–4778.
- Guillén-Nieto, V. (2023). *Hate speech: Linguistic perspectives*. De Gruyter.
- Hickey, D., Fessler, D. M. T., Lerman, K., & Burghardt, K. (2025). X under Musk's leadership: Substantial hate and no reduction in inauthentic activity. *PLoS ONE*, 20(2), e0313293. <https://doi.org/10.1371/journal.pone.0313293>
- Kang, M., Rivé-Lasan, M.-O., Kim, W., & Hall, P. (Eds.). (2023). *Hate speech in Asia and Europe: Beyond hate and fear* (1st Ed.). Routledge.
- Karapetjana, I., Roziņa, G., & Spirida, M. (2023). Critical discourse analysis of hate speech from a linguistic perspective. *Language: Meaning & Form/Valoda: Nozīme un Forma*, 14(6), 74–90. <https://doi.org/10.22364/vnf.14.06>
- Kaulika, A., Salim Mansyur, A., & Wardoyo, C. (2024). Impoliteness strategies in hate comments on Noah Schnapp's Instagram post. *Celtic: A Journal of Culture*, 11(2), 115–126. <https://doi.org/10.22219/celtic.v11i2.33866>
- Khilmia, M. N. (2024). The impact of BTS's victory at the Billboard Music Awards as the first K-pop singer in increasing South Korea's global soft power. *Budi Luhur Journal of Strategic & Global Studies*, 2(2), 1–14. <https://doi.org/10.36080/jsgs.v2i2.3424>
- Letwory, A. M., & Sihombing, L. H. (2023). Phenomenon of fandom war in K-pop community on social media. *Ekspresi dan Persepsi: Jurnal Ilmu Komunikasi*, 6(2), 229–238. <https://doi.org/10.33822/jep.v6i2.5808>
- Liang, R. (2023). Analysis of the influence of BTS on K-pop. *Communications in Humanities Research*, 14(1), 42–47. <https://doi.org/10.54254/2753-7064/14/20230403>
- Maarouf, A., Pröllochs, N., & Feuerriegel, S. (2022). The virality of hate speech on social media. *Proceedings of the ACM on Human-Computer Interaction*, 8 (CSCW1), 1–22. <https://doi.org/10.1145/3641025>
- Manik, S., Pardede, H., Franklin, T. N. D., & Pasaribu, T. K. A. (2022). Impoliteness in Indonesian hate speech on Basuki Tjahaja Purnama (BTP) as found in YouTube commentary. *Jurnal Education and Development*, 10(3), 420–429. <https://doi.org/10.37081/ed.v10i3.4128>
- Margono, H., Saud, M., & Ashfaq, A. (2024). Dynamics of hate speech in social media: Insights from Indonesia. *Global Knowledge, Memory and Communication. Advance online publication*, 75(5-6), 1784–1804. <https://doi.org/10.1108/GKMC-11-2023-0464>
- Marques, T. (2023). The expression of hate in hate speech. *Journal of Applied Philosophy*, 40(5), 769–787. <https://doi.org/10.1111/japp.12608>
- Messaoudi, I., & Sakale, S. (2023). Triangulation in research: Exploring Moroccan EFL university graduate students' knowledge and familiarity. *International Journal of Linguistics, Literature and Translation*, 6(11), 98–105. <https://doi.org/10.32996/ijl.2023.6.11.13>
- Nascimento, F. R. S., Cavalcanti, G. D. C., & Costa-Abreu, M. D. (2025). Gender bias detection on hate speech classification: An analysis at feature-level. *Neural Computing & Applications*, 37, 3887–3905. <https://doi.org/10.1007/s00521-024-10841-8>
- Orgeret, K. S., Mutsvairo, B., de Bruijn, M., Schroeder, D. T., Badji, S. D., & Moges, M. A. (2025). Hashtags, hatetags and social media campaigns in Ethiopia's Tigray conflict. *Information, Communication & Society*, 29(3), 1–19. <https://doi.org/10.1080/1369118X.2025.2533314>
- Paasch-Colberg, S., Strippel, C., Trebbe, J., & Emmer, M. (2021). From insult to hate speech: Mapping offensive language in German user comments on immigration. *Media and Communication*, 9(1), 171–180. <https://doi.org/10.17645/mac.v9i1.3399>
- Pasaribu, A. N. (2021). Hate speech on Joko Widodo's official Facebook: An analysis of impoliteness strategies used by different genders. *Eltin Journal: Journal of English Language Teaching in Indonesia*, 9(1), 56–64. <https://doi.org/10.22460/eltin.v9i1.p56-64>
- Piazza, J., & Van Doren, N. (2023). It's about hate: Approval of Donald Trump, racism, xenophobia and support for political violence. *American Politics Research*, 51(3), 299–314. <https://doi.org/10.1177/1532673x221131561>
- Pita, A. L., & Widiansyah, S. (2023). Fenomena fanwar dan fanatisme antar fandom K-pop di media sosial Twitter. *SEMAR: Jurnal Sosial dan Pengabdian Masyarakat*, 1(2), 1–11. <https://doi.org/10.59966/semar.v1i02.280>
- Pornyng, W., & Tantiniranat, S. (2024). Hate speech towards female game players in Valorant game. *Journal of English Language and Linguistics*, 5(3), 374–392. <https://doi.org/10.62819/jel.2024.657>
- Prakoso, G. R., & Runturambi, A. J. S. (2021). Strategic intelligence analysis of religion-based hate speech in social media: A case study of the Directorate Intelligence and Security at Polda Metro Jaya. *International Review of Humanities Studies*, 6(1), 57–71.
- Regaro, J. (2023). Linguistic features of multilingual hate speech in the online “Bardagulan.” *International Journal of English Language Studies*, 5(4), 120–139. <https://doi.org/10.32996/ijels.2023.5.4.10>
- Rodríguez-Peral, E. M., Franco, T. G., & Bustos, D. R. P. (2025). Propaganda of hate speech on social network X: Trends and approaches. *Social Inclusion*, 13, 1–21. <https://doi.org/10.17645/si.9317>
- Samitharathana, R. W. (2022). A contrastive analysis: What is the conventional implicature? Is the “bad” content of a slur conveyed as a conventional implicature? *European Journal of Language and Culture Studies*, 1(1), 1–4. <https://doi.org/10.24018/ejlang.2022.1.1.1>

- Sari, R. P., Anwar, M., & bin Abdul Hakim, M. K. (2025). Investigating hate speech comments: A forensic linguistic study. *Austronesian: Journal of Language Science & Literature*, 4(1), 47–64. <https://doi.org/10.59011/austronesian.4.1.2025.47-64>
- Schmitz, M., Burghardt, K., & Muric, G. (2022). Quantifying how hateful communities radicalize online users. In *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 139–146). <https://doi.org/10.1109/ASONAM55673.2022.10068644>
- Sigalingging, A., Girsang, M., David, C., & Silalahi, M. (2025). Types of implicatures: Hate speech, TikTok, social media. *Journal of Society Innovation and Development*, 7(1), 170–178. <https://doi.org/10.5281/zenodo.15277943>
- Singh, H. (2022). Impact of social media on interpersonal communication. *International Journal of Communication and Information Technology*, 3(2), 26–30. <https://doi.org/10.33545/2707661x.2022.v3.i2a.69>
- Sirait, N., Purba, R., Silitonga, M. U. A., & Putri, D. M. (2025). Hate speech in Gus Miftah Instagram comment section: An ice tea seller case study. *Linguistik, Terjemahan, Sastra (LINGTERSA)*, 6(1), 47–56. <https://doi.org/10.32734/lingtersa.v6i1.19205>
- Siregar, F. S., & Manurung, Y. H. (2023). Literasi digital sebagai upaya antisipasi ujaran kebencian di media sosial. *Jurnal SOMASI (Sosial Humaniora Komunikasi)*, 4(1), 68–76. <https://doi.org/10.53695/js.v4i1.929>
- Situmorang, Y. F., Lestari, F. D., & Sinaga, N. T. (2024). Impoliteness strategies of the hate comments on Twitter. *ALACRITY: Journal of Education*, 4(3), 220–228. <https://doi.org/10.52121/alacrity.v4i3.437>
- Subyantoro, S., & Apriyanto, S. (2020). Impoliteness in Indonesian language hate speech on social media contained in the Instagram account. *Journal of Advances in Linguistics*, 11(2), 36–46. <https://doi.org/10.24297/jal.v11i.8655>
- Toktarova, A., Syrlybay, D., Myrzakhmetova, B., Anuarbekova, G., Rakhimbayeva, G., Zhylanbaeva, B., & Kerimbekov, M. (2023). Hate speech detection in social networks using machine learning and deep learning methods. *International Journal of Advanced Computer Science and Applications*, 14(5), 396–406. <https://doi.org/10.14569/IJACSA.2023.0140542>
- Virginia, O., & Triandana, A. (2023). Analisis wacana kritis Theo van Leeuwen dalam pemberitaan mengenai isu rasisme terhadap boyband Korea BTS pada media pemberitaan daring Kompas, Kumparan, dan Republika. *Kajian Linguistik dan Sastra*, 2(2), 151–160. <https://doi.org/10.22437/kalistra.v2i2.23271>
- Wajahat, M. A. (2024). Social media and its impact on non-verbal communication. *International Journal for Multidisciplinary Research (IJFMR)*, 6(3), 1–15. <https://doi.org/10.36948/ijfmr.2024.v06i03.23102>
- Waziri, Z. Y. (2023). The role of politeness theory in resolving misunderstandings in social media communication. *Zamfara International Journal of Humanities*, 2(2), 36–42. <https://doi.org/10.36349/zamijoh.2023.v02i02.004>