

Hate Speech in Balinese Language on Social Media: A Forensic Linguistic Perspective

Putu Ary Sintya Dewi¹, I Wayan Pastika², I Gusti Ayu Gde Sosiowati³

^{1 2 3}Departement of Linguistics, Faculty of Humanities, Universitas Udayana, Indonesia

ABSTRACT: This study examines hostile expressions in Balinese language comments on social media concerning discussions of Ngaben and cremation practices from a forensic linguistic perspective, with particular attention to their potential relation to hate speech and their illocutionary functions. The study aims to identify forms of hostile or potentially hateful expressions found in Balinese language social media comments and to analyze the illocutionary acts performed through these expressions. A qualitative descriptive method was employed using Balinese language comments related to Ngaben discourse on social media as the data source. The data were collected through documentation and analyzed by considering their linguistic forms, targets, contextual meanings, and pragmatic functions. The analysis was informed by the Indonesian National Police Circular Letter No. SE/6/X/2015 concerning the handling of hate speech, supported by Brown (2015) and Waldron (2012), while illocutionary acts were analyzed based on Austin (1962) and Searle (1979). The findings indicate that the selected comments predominantly contain individual-directed hostile expressions rather than clearly established identity based hate speech. The expressions include implicit intimidation, personal reprimand, and hostile or mocking warnings. In terms of illocutionary acts, the comments perform primarily expressive and directive functions. Expressive acts are used to convey hostility, irritation, mockery, or disapproval, while directive acts are employed to influence the addressee's behavior, including telling the addressee to stop interfering in other people's affairs. The findings demonstrate that offensive or hostile language should not automatically be classified as hate speech. Instead, forensic linguistic analysis requires consideration of the target, contextual meaning, identity related elements, and communicative function of each utterance. The study contributes to the analysis of Balinese language digital discourse by demonstrating the importance of distinguishing interpersonal hostility from identity based hate speech through a combination of forensic linguistic and pragmatic analysis.

KEYWORDS: Forensic linguistics, Hate Speech, Illocutionary acts, Social Media

I. INTRODUCTION

The rapid development of social media has fundamentally transformed the way people communicate, express opinions, and respond to social and cultural issues. Platforms such as Instagram provide users with opportunities to participate in public discussions through comments, reactions, and other forms of digital interaction. Although these platforms facilitate freedom of expression and public participation, they also create spaces in which offensive, hostile, insulting, and hateful expressions can be produced and circulated rapidly. Social media communication is therefore not only a technological phenomenon but also an important linguistic domain in which social attitudes, identities, conflicts, and power relations are manifested through language. ElSherief et al. (2018) emphasize that social media can facilitate anti-social behavior, including online harassment, cyberbullying, and hate speech. Their study further demonstrates that the target of an utterance is an important factor in understanding online hate speech, particularly in distinguishing hate directed toward a specific individual from hate generalized toward a group sharing a protected characteristic. From a linguistic perspective, hate speech cannot be identified solely from the presence of rude, offensive, or emotionally charged words. The interpretation of an utterance requires attention to its linguistic form, communicative intention, target, and social context. Brown (2015) conceptualizes hate speech in relation to phenomena such as group defamation, negative stereotyping, stigmatization, and expressions that contribute to hostility toward particular groups. Similarly, Waldron (2012) emphasizes that hate speech may cause harm not merely because it offends individuals but because it can undermine the dignity and social standing of members of targeted groups. This distinction is important in social-media research because not every insulting or aggressive comment should automatically be categorized as hate speech. An expression may be offensive because it attacks an individual personally, whereas hate speech generally involves a broader identity-based dimension.

Hate Speech in Balinese Language on Social Media: A Forensic Linguistic Perspective

This distinction becomes particularly important when examining digital discourse involving cultural and religious practices. Social and cultural practices are often closely associated with collective identities, and disagreement about such practices can develop into linguistic expressions that target either an individual or a broader community. In this context, the target of an utterance becomes a significant analytical element. ElSherief et al. (2018) distinguish between directed hate speech, which targets a particular person or entity, and generalized hate speech, which targets a group associated with a protected characteristic. Their findings also indicate that generalized hate speech is frequently associated with religious hostility. Such a distinction provides an important analytical basis for determining whether hostile comments in a social-media discussion represent interpersonal aggression or identity based hate speech. The issue is particularly relevant in discussions of Ngaben, a cremation related ritual associated with Balinese Hindu cultural and religious practices. Ngaben is not merely an individual practice but is situated within broader cultural, religious, and communal contexts. Consequently, public discussions concerning changes or modernization of Ngaben practices may generate different linguistic responses, ranging from support and criticism to ridicule, hostility, and potentially identity based attacks. In digital environments, these responses are expressed through relatively short comments in which meanings may be highly implicit, context-dependent, and pragmatically complex. The linguistic characteristics of Balinese, including informal lexical choices, address forms, evaluative expressions, and culturally specific meanings, therefore require careful interpretation rather than direct translation alone.

Forensic linguistics provides an appropriate perspective for examining such discourse because it investigates the relationship between language and legal or evidential contexts. Coulthard and Johnson (2007) describe forensic linguistics as the study of language in relation to legal processes and evidence, while Olsson (2008) emphasizes the role of linguistic analysis in investigating language associated with crime and the law. Within digital communication, forensic linguistic analysis can contribute to the identification and interpretation of potentially problematic language by examining lexical choices, semantic meanings, pragmatic functions, contextual references, and the intended or perceived target of an utterance. Thus, forensic linguistic analysis does not simply identify whether a particular word is offensive; rather, it examines how an expression functions within a particular communicative context. In the Indonesian context, hate speech has also become an important issue in discussions of digital communication and law enforcement. The Indonesian National Police Chief Circular Letter No. SE/6/X/2015 concerning the handling of hate speech identifies several forms of problematic expression, including insult, defamation, blasphemy, unpleasant acts, provocation, incitement, and the dissemination of false information, particularly when related to identity categories such as ethnicity, religion, religious denomination, belief, race, intergroup relations, skin color, gender, disability, and sexual orientation. The circular letter is useful as an operational reference for examining potentially hateful expressions in Indonesian social media discourse. However, its application to linguistic data requires contextual and linguistic interpretation because the same lexical expression may perform different functions depending on its target, context, and communicative purpose. Previous studies have demonstrated the usefulness of forensic linguistic approaches in analyzing hate speech on social media. Ramadani (2021), for example, examined hate speech in the Instagram comments of Indonesian socialmedia users and demonstrated the relevance of forensic linguistic analysis to digitally mediated expressions. Permatasari and Subyantoro (2020) similarly investigated hate speech on Facebook and identified several forms of problematic expression, including provocation, incitement, insult, blasphemy, defamation, and the dissemination of false information. These studies demonstrate that social media comments can provide valuable linguistic evidence for identifying different forms of hostile and potentially hateful discourse.

Research focusing specifically on Instagram has also continued to develop. Putra et al. (2024) examined hate speech in comments on the Instagram account of the Indonesian House of Representatives (DPR RI) using forensic linguistic analysis, lexical semantic analysis, speech-act theory, and politeness analysis. The study demonstrates that Instagram comments can be analyzed not only in terms of their lexical characteristics but also through their pragmatic and legal implications. Katuuk et al. (2024) likewise examined digital traces of hate speech on Instagram from a forensic linguistic perspective, demonstrating the importance of documenting and linguistically interpreting hostile expressions in digital communication. More recently, Hayati et al. (2025) analyzed hate speech on Willie Salim's Instagram account and identified forms such as insult, blasphemy, and the dissemination of false information through lexical, semantic, and pragmatic analysis.

The relationship between hate speech and speech acts is another important area of research. Hate speech does not only communicate negative propositional content it also performs communicative actions. A speaker may insult, warn, threaten, criticize, ridicule, command, or express anger through a particular utterance. Austin (1962) established the foundational concept of speech acts by demonstrating that utterances can perform actions, while Searle (1979) further developed the classification of illocutionary acts into representative/assertive, directive, commissive, expressive, and declarative categories. These categories are useful for examining what speakers are doing through their utterances rather than focusing only on what the utterances literally say. The pragmatic dimension is particularly relevant to social media hate speech because hostile meanings are not always expressed explicitly. A comment may appear grammatically simple but perform a warning, insult, threat, command, or expression of contempt depending on the context. Ubaidillah and Wijana (2021), for instance, demonstrate that hate speech on Indonesian social media may perform various directive functions, including incitement, invitation, ordering, prohibition, criticism, and warning. Their findings

Hate Speech in Balinese Language on Social Media: A Forensic Linguistic Perspective

indicate that identifying the illocutionary force of hate related expressions can provide a deeper understanding of how hostility is communicated through language.

Recent studies have increasingly combined forensic linguistic and speech act approaches. Prameswari et al. (2025), for example, investigated hate speech in Instagram comments using forensic linguistic theory, speech act analysis, and the Indonesian National Police Chief Circular Letter No. SE/06/X/2015. Their findings demonstrate the coexistence of different illocutionary functions and forms of hate speech in Instagram comments, including provocation, incitement, defamation, unpleasant acts, insult, blasphemy, and false information. Similarly, a recent study by Elisabeth Dwi Clara et al. (2026) explicitly examined online hate speech on Instagram through forensic linguistics and illocutionary acts, demonstrating the continuing relevance of integrating linguistic and pragmatic perspectives in the analysis of digital hate speech. Studies concerning Indonesian and Balinese cultural contexts also indicate the importance of examining hate speech in relation to religious and cultural identity. Maharani et al. (2022), for example, investigated hate speech in comments concerning Desak Made Darmawati and a discourse involving Hindu religious identity in Bali.

This research is particularly relevant because it demonstrates how social media discourse may involve linguistic expressions that extend beyond criticism of an individual statement and become associated with broader religious and cultural identities. Such research provides an important foundation for examining hate speech in Balinese related social media discourse. However, previous studies have generally focused on Indonesian language social media data, political figures, government institutions, celebrities, or broad categories of online hate speech. Although these studies have contributed significantly to the understanding of digital hate speech, comparatively less attention has been given to hate speech expressed specifically in the Balinese language and situated within a culturally specific discussion of Ngaben. This gap is significant because language cannot always be interpreted independently of cultural context. Balinese expressions may contain lexical forms, address terms, humor markers, implicit meanings, and culturally specific references that are not fully represented through direct English or Indonesian translation. Consequently, the analysis of Balinese language hate speech requires attention to the original linguistic form as well as its pragmatic and cultural context.

Another gap concerns the relationship between the classification of hate speech and the illocutionary function of an utterance. Previous studies have shown that hate speech can perform different speech acts, but the two dimensions should not be treated as identical. Hate speech classification concerns the nature of the target, the identity dimension, and the social or discriminatory content of an expression, whereas illocutionary analysis concerns the communicative action performed by the speaker. An utterance may therefore be expressive without necessarily constituting hate speech, or it may function simultaneously as an expressive act and an identity based hateful expression. Maintaining this distinction is particularly important for forensic linguistic analysis because it prevents every hostile or insulting comment from being automatically categorized as hate speech. The present study addresses this gap by examining Balinese language comments on social media concerning discussions of Ngaben and cremation practices from a forensic linguistic perspective. The study focuses on two interconnected but analytically distinct dimensions the identification of hate speech and the analysis of illocutionary acts. The first dimension examines whether an expression constitutes hate speech by considering its linguistic form, target, context, and identity related characteristics. The second dimension examines the communicative function performed by the expression, particularly whether it functions as an assertive, directive, expressive, commissive, or other illocutionary act.

This study therefore contributes to the growing body of research on digital hate speech in two ways. First, it provides a culturally and linguistically specific analysis of hate related discourse expressed in the Balinese language, an area that remains comparatively underexplored in forensic linguistic research. Second, it demonstrates how hate speech classification can be complemented by illocutionary analysis without conflating the two analytical dimensions. By retaining the original Balinese expressions and providing English translations for international readers, the study also seeks to preserve the linguistic characteristics of the source data while making the findings accessible to a wider academic audience. Based on this background, the study addresses two research questions: (1) What forms of hate speech are found in Balinese language comments on social media? and (2) What illocutionary acts are used in Balinese language comments containing hate speech on social media? Through these questions, the study seeks to demonstrate how forensic linguistic analysis can distinguish identity based hate speech from general interpersonal hostility while simultaneously revealing the pragmatic actions performed through digital language.

II. THEORITICAL FRAMEWORK

Forensic linguistics is an interdisciplinary field that applies linguistic knowledge to legal and investigative contexts. Olsson (2008) describes forensic linguistics as the application of linguistic analysis to issues involving law, while Coulthard and Johnson (2007) emphasize the role of linguistic evidence in legal and investigative processes. In this perspective, language can be examined not only as a means of communication but also as evidence that may reveal meaning, intention, interaction, and potentially harmful communicative behavior. The development of social media has expanded the relevance of forensic linguistics because written comments, posts, and other digital texts can become objects of linguistic investigation. Previous studies have demonstrated that social media comments can be analyzed from a forensic linguistic perspective by examining their linguistic forms, meanings, speech

Hate Speech in Balinese Language on Social Media: A Forensic Linguistic Perspective

acts, and potential legal implications. In the present study, forensic linguistics provides the overarching perspective for examining Balinese language comments on social media, particularly comments associated with the discourse surrounding Ngaben. Within forensic linguistic analysis, one important issue is the identification of hate speech. Hate speech generally refers to expressions that attack, demean, or promote hostility toward individuals or groups on the basis of characteristics associated with their identity. Brown (2015) emphasizes that hate speech involves communication directed against individuals or groups because of characteristics such as race, ethnicity, religion, nationality, or other forms of group identity. Waldron (2012) further argues that hateful expression can affect the dignity and social standing of targeted individuals or groups. These perspectives indicate that hate speech cannot be identified solely from the presence of offensive or emotionally negative vocabulary. The target of the expression, the basis of the attack, and the surrounding context must also be considered.

The distinction between hate speech and offensive language is therefore important in the present study. Allan and Burridge (2006) explain that offensive language may involve expressions that violate social norms, convey negative attitudes, or cause offense. However, an insulting, mocking, or rude expression directed toward an individual does not automatically constitute hate speech. Hate speech requires a more specific examination of the target and the social or identity-based dimension of the expression. In the Indonesian context, the identification of hate-related expressions can also be considered in relation to Indonesian legal and institutional frameworks, including Indonesian National Police Circular Letter No. SE/06/X/2015 concerning hate speech. Accordingly, this study examines hate speech by considering the linguistic form of an expression, its target, its contextual meaning, the identity or group characteristic involved, and its potential to demean, intimidate, provoke hostility, or discriminate. This distinction is particularly important for forensic linguistic research because it prevents all forms of insult, mockery, or hostility from being automatically categorized as hate speech. In addition to examining the nature of potentially hateful expressions, this study analyzes the illocutionary acts performed through the comments. Speech act theory originates from Austin (1962), who distinguishes among locutionary, illocutionary, and perlocutionary acts. An illocutionary act concerns the communicative action performed through an utterance, such as stating, requesting, warning, criticizing, or expressing an attitude. Searle (1979) subsequently developed a classification of illocutionary acts into five major categories: assertives, directives, commissives, expressives, and declarations. Assertives commit the speaker to the truth of a proposition; directives attempt to influence the hearer to perform an action; commissives commit the speaker to a future action; expressives communicate the speaker's psychological or emotional state; and declarations bring about a change in a particular state of affairs through the utterance itself. Searle's classification has been widely applied in the analysis of social-media discourse, including studies examining hate and provocative speech.

The analysis of illocutionary acts is particularly relevant to social media comments because users may communicate hostility indirectly through different speech functions. A comment may appear to be a statement on the surface but function as a warning, threat, ridicule, or criticism in its particular context. Likewise, an expression containing laughter markers or apparently humorous language may perform an expressive act of mockery or contempt. Therefore, identifying an illocutionary act requires consideration of the linguistic form, contextual meaning, communicative intention, and relationship between the participants. In this study, Searle's classification is used to determine whether Balinese language comments function primarily as assertives, directives, expressives, commissives, or declarations. Hate speech analysis and illocutionary act analysis are treated as two related but distinct analytical dimensions. Hate speech analysis addresses the character of the expression, its target, the basis of the attack, and its potential social harm, whereas illocutionary analysis identifies the communicative action performed through the utterance. Consequently, an illocutionary act does not automatically constitute hate speech. For example, a directive may function as an ordinary request or criticism, while an expressive act may contain personal ridicule without targeting a social group on the basis of identity. Conversely, an identity-based hateful expression may perform an expressive act of contempt or a directive act intended to encourage hostility. Recent research on Instagram comments similarly demonstrates the value of examining hate speech together with illocutionary categories while maintaining attention to the distinction between linguistic function and hate speech classification. Based on this theoretical foundation, the present study examines Balinese language social media comments through a forensic linguistic perspective by integrating hate speech analysis and illocutionary act analysis. The first dimension determines whether an expression represents hate speech or another form of offensive or hostile language by considering its linguistic characteristics, target, context, and identity based dimension. The second dimension identifies the illocutionary function of the expression based on Searle's classification. This framework enables the study to examine both the nature of potentially harmful language and the communicative action performed through that language. Such an integrated approach is expected to provide a more comprehensive understanding of how Balinese is used in hostile or controversial social media discourse and to demonstrate the relevance of local language data to forensic linguistic research.

III. METHOD

This study employed a qualitative descriptive approach from a forensic linguistic perspective to analyze hate related and hostile expressions in Balinese language comments on social media. A qualitative approach was considered appropriate because the study focused on interpreting linguistic forms, contextual meanings, communicative intentions, and the functions of utterances rather than quantifying linguistic phenomena. Forensic linguistics provides the main perspective for examining the linguistic evidence contained

Hate Speech in Balinese Language on Social Media: A Forensic Linguistic Perspective

in social media comments, particularly expressions that may involve insults, mockery, intimidation, threats, or hate speech. The analysis was conducted by considering both the linguistic characteristics of the expressions and the contexts in which they occurred. The data consisted of Balinese language comments collected from Instagram in discussions concerning Ngaben cremation in Bali. The comments were selected from social media discussions in which users expressed different opinions, reactions, criticism, and disagreement concerning the topic. The data were not limited from the outset to expressions that had already been determined as hate speech. Instead, potentially offensive, hostile, threatening, mocking, or hate-related expressions were selected for further linguistic examination. This procedure was adopted to allow the analysis to distinguish between hate speech and other forms of problematic language, since an offensive or hostile expression does not necessarily constitute hate speech. The original Balinese expressions were retained as the primary data, while English translations were provided to facilitate interpretation for international readers.

The data were collected using documentation and note taking techniques. Relevant Instagram posts and their comment sections were identified, and comments written in Balinese that were relevant to the research objectives were documented. The selected comments were then transcribed and organized for analysis. Each datum was assigned a code to facilitate systematic presentation and discussion. The selection of data considered several criteria, namely that the comment (1) was expressed predominantly in Balinese, (2) was related to the Ngaben discourse under investigation, and (3) contained linguistic features that indicated criticism, ridicule, insult, hostility, intimidation, or other potentially harmful expressions. The analysis was conducted in two main stages. First, the data were examined to determine the nature of the linguistic expressions from a hate-speech perspective. The analysis considered the linguistic form, meaning, target, contextual background, and communicative implications of each comment. Particular attention was given to whether the expression attacked an individual or a group on the basis of identity related characteristics, such as ethnicity, religion, race, or other group identities. Expressions that were offensive, insulting, mocking, or hostile but lacked a clear identity based dimension were distinguished from hate speech. This distinction was necessary to avoid categorizing every form of negative or aggressive language as hate speech.

Second, the selected utterances were analyzed in terms of their illocutionary acts based on Austin's (1962) speech act theory and Searle's (1979) classification. The analysis identified the communicative function performed by each utterance, including assertive, directive, commissive, expressive, and declarative acts where applicable. The interpretation considered not only the grammatical form of the utterance but also its contextual meaning and communicative intention. For instance, an expression that takes the form of a warning may function as an attempt to intimidate the target, while an apparently humorous expression may perform an expressive act of ridicule or contempt. The two stages of analysis were treated as distinct but complementary. Hate speech analysis was used to determine the character and potential social harm of an expression, including whether it involved an identity based attack. Illocutionary analysis, on the other hand, was used to identify the communicative action performed through the utterance. Therefore, the identification of an expressive, directive, or assertive act did not automatically indicate that the utterance constituted hate speech. The final classification was based on the combination of linguistic evidence and contextual interpretation. The findings were presented using a summary table followed by qualitative descriptive analysis. The table presents the original Balinese comment, its English translation, the identified illocutionary act, and the classification of the expression. Each datum is subsequently discussed in detail to explain its linguistic indicators, contextual meaning, communicative function, and relevance to hate speech classification. The use of the original Balinese data alongside English translations ensures that the linguistic characteristics of the source expressions remain visible while allowing the findings to be understood by an international readership.

IV. ANALYSIS

The analysis of Balinese language comments in social-media discourse concerning Ngaben reveals various forms of negative, hostile, and potentially hateful expressions. The comments demonstrate that not every offensive or impolite utterance can be categorized as hate speech. Some expressions are directed toward individuals and function as personal criticism, mockery, intimidation, or warnings, whereas other expressions construct negative representations of a particular social or religious group. In this study, hate speech is identified by considering the categories and contextual characteristics outlined in Indonesian National Police Circular Letter No. SE/6/X/2015 concerning the Handling of Hate Speech. The analysis focuses on the linguistic form of the expression, the target of the utterance, the identity associated with the target, and the relationship between the expression and the potential construction of hatred, discrimination, violence, or social conflict. This framework allows the analysis to distinguish identity based hate speech from other forms of offensive or hostile language. The analysis of illocutionary acts is conducted as a separate analytical dimension using the speech act framework of Austin (1962) and Searle (1979). The focus is on the communicative function performed by the speaker, including assertive, directive, and expressive acts. Consequently, the classification of an utterance as hate speech does not automatically determine its illocutionary category. The two dimensions complement one another but address different aspects of meaning.

Table 1. Balinese Language Social Media Comments

No	Original Comment	BalineseEnglish Translation	Target	Hate Speech	Illocutionary Act
1	Seme'e angge nanem cai ane konten...wkwwk..	"The one who made this content should be buried, lol."	Content creator	Personal hostile expression	Expressive + Directive
2	URUS IBE-IBE GEN SING PERLU NGURUS	"Just mind your own business; you don't need to concern yourself with other people's affairs."	Individual	Personal criticism	Directive + Expressive
3	De mekejang urus pak, urus iban ci gen pedidi."	"Do not concern yourself with everything; just mind your own business."	Individual	Not established as hate speech	Directive + Expressive
4	"AWAS mekeplag lengar ee puakkk 🤔🤔🤔"	"Watch out, your half bald head might explode, bro!"	Individual	Personal mockery	Directive + Expressive

DATA 1

"Seme'e angge nanem cai ane ngae konten...wkwwk.."
 ("The one who made this content should be buried, lol.")

The first datum contains an expression of intense hostility directed toward an individual, namely the person who created the content being discussed. The expression *seme'e angge nanem cai* is used to express the speaker's hostility toward the content creator. The phrase *nanem cai*, literally associated with the act of burying someone, produces a strong negative meaning when directed at a person. The expression is followed by *wkwwk*, a written laughter marker that reinforces the mocking and derisive tone of the comment. From a hate-speech perspective, this datum should not automatically be classified as hate speech. The target is an individual content creator rather than a group identified through religion, ethnicity, race, or another social identity. Although the utterance is highly hostile and can be interpreted as an implicit threat or death related expression, the available linguistic evidence does not establish that the hostility is motivated by the target's membership in a particular identity group. This distinction is important because offensive or threatening language and hate speech are not synonymous. Brown (2015) treats hate speech as a heterogeneous category and distinguishes forms such as group defamation and negative stereotyping from other forms of problematic expression. Similarly, Waldron's account focuses on attacks that degrade the dignity and social standing of members of a targeted group, rather than treating every offensive utterance as hate speech.

Accordingly, Data 1 is more appropriately classified as a personal hostile expression with an implicit intimidation or threatening dimension, rather than religion based hate speech. Its forensic relevance nevertheless remains significant because the utterance contains a potentially harmful expression directed at a specific person. From the perspective of illocutionary acts, the utterance primarily performs an expressive act. The speaker expresses anger, contempt, and hostility toward the content creator. The expression also has an indirect directive dimension, because the wording *nanem cai* can function pragmatically as an intimidating warning rather than merely as a description. The laughter marker *wkwwk* strengthens the expressive force by presenting the hostile statement in a mocking form.

Thus, Data 1 demonstrates that an utterance can be linguistically aggressive without necessarily qualifying as hate speech. Its significance lies in the distinction between individual directed hostility and identity based group hostility.

DATA 2

“*URUS IBE-IBE GEN SING PERLU NGURUS TIMPAL NE LENAN*”

(“Just mind your own business; you don't need to concern yourself with other people's affairs.”)

This expression can be interpreted as a direct interpersonal reprimand. The expression contains an imperative construction through *urus ibe-ibe gen*, which can be understood as an instruction for the addressee to deal with their own affairs. The subsequent expression *sing perlu ngurus timpal ne lenan* reinforces the speaker's rejection of the addressee's involvement in another person's affairs. Unlike Data 3, this expression does not explicitly construct a negative representation of a religious, ethnic, racial, or other identity-based group. The target appears to be an individual participant in the discussion. Consequently, the utterance is more appropriately classified as personal criticism or interpersonal reprimand rather than hate speech. The distinction is analytically important because the expression is certainly confrontational, but its hostility is directed toward the addressee's behavior rather than toward an identity category. Brown's framework is useful here because it cautions against treating all forms of offensive discourse as one homogeneous category; his analysis separates different forms and rationales associated with hate speech. The absence of an identifiable group based target in this datum therefore weakens the basis for classifying it as hate speech. The dominant illocutionary force is a directive act. Through the imperative construction, the speaker tells the addressee to stop involving themselves in another person's affairs. The utterance also carries an expressive function, because the directive communicates annoyance, rejection, and disapproval toward the addressee.

The pragmatic force of the utterance is therefore stronger than its literal informational content. The speaker is not merely providing advice to the addressee; the speaker is attempting to control the addressee's behavior within the interaction. The combination of directive and expressive functions demonstrates how interpersonal conflict can be constructed through ordinary Balinese linguistic forms without necessarily becoming identity based hate speech. From a forensic linguistic perspective, Data 2 is useful as a contrastive datum. Its inclusion demonstrates that the analysis does not label every negative or impolite social media comment as hate speech. Instead, the classification considers the relationship between linguistic expression, target, context, and social identity.

DATA 3

“*De mekejang urus pak, urus iban ci gen pedidi.*”

(“Do not concern yourself with everything; just mind your own business.”)

The utterance above is directed toward an individual who is perceived as interfering in another person's affairs. The expression “*de mekejang urus pak*” conveys a prohibition against involving oneself in matters that are considered unrelated to the speaker or the intended addressee. This is reinforced by the phrase “*urus iban ci gen pedidi,*” which instructs the addressee to focus on their own affairs. The lexical choices therefore construct a confrontational and admonishing tone. From the perspective of hate-speech classification, however, this expression cannot be sufficiently established as hate speech under the framework of Indonesian Police Circular Letter No. SE/6/X/2015. Although the comment conveys hostility and direct criticism, it does not explicitly target an individual or group on the basis of religion, ethnicity, race, or another identity category specified in the hate-speech framework. There is also no clear indication of provocation, incitement, or a call for discrimination or violence. Therefore, the expression is more appropriately classified as an individual-directed hostile expression or personal reprimand rather than identity-based hate speech.

From a pragmatic perspective, the utterance primarily performs a directive illocutionary act. The speaker attempts to influence the addressee's behavior by instructing them not to interfere and to attend to their own affairs. The expression is therefore not merely a statement of disagreement but functions as a direct behavioral instruction. At the same time, its confrontational wording conveys the speaker's negative attitude toward the addressee, giving the utterance a secondary expressive dimension. Thus, the dominant illocutionary force is directive, with an accompanying expressive element of irritation or disapproval. The data demonstrate the importance of distinguishing between hate speech and general hostile language in forensic linguistic analysis. While the comment contains a confrontational and potentially offensive expression, the absence of an identity-based target means that it should not automatically be categorized as hate speech. This distinction strengthens the analysis by showing that the classification of an utterance depends not only on its negative tone but also on its target, contextual meaning, and relationship to the characteristics of hate speech.

DATA 4

“*AWAS mekeplag nas lengar ee puakkk* 🤔🤔🤔”

(“Watch out, your half bald head might explode, bro!”)

Hate Speech in Balinese Language on Social Media: A Forensic Linguistic Perspective

The expression demonstrates that contains a combination of warning, mockery, and personal insult. The expression begins with *awas*, which functions as a warning or cautionary marker. This is followed by *mekeplag*, meaning to explode, and *lengar*, which refers to baldness or a half bald condition in the contextual use of the comment. The informal address *puakkk* adds an interpersonal and colloquial dimension, while the repeated laughing emojis reinforce the mocking tone. The comment is directed toward an individual rather than an identifiable social group. The negative evaluation concerns the addressee's physical appearance, particularly the reference to *lengar*. Consequently, the expression is best classified as personal mockery and hostile warning, rather than hate speech.

This distinction is consistent with the theoretical distinction between ordinary offensive expression and group directed hate speech. Waldron specifically emphasizes that causing offense is not by itself equivalent to an attack on a group's dignity. In this utterances, the available linguistic evidence indicates interpersonal ridicule rather than an attack based on religion, ethnicity, race, or another group identity. The dominant illocutionary force is a directive act because *awas* functions as a warning to the addressee. However, the warning is not entirely neutral. The reference to the addressee's head potentially exploding and the laughing emojis create an expressive function of ridicule and amusement. Therefore, the utterance can be analyzed as a combination of directive and expressive acts. The emojis are particularly relevant to the pragmatic interpretation. Without the laughing emojis, *awas mekeplag* could potentially be interpreted more literally as a warning. The repeated laughing emojis, however, provide contextual cues that the speaker is mocking the addressee rather than seriously warning them about a physical event. This illustrates how multimodal elements in social media discourse can contribute to the interpretation of an utterance's illocutionary force. From a forensic linguistic perspective, a significant feature of this expression is demonstrates that social media meaning is not determined solely by lexical content. Informal address forms, expressive vocabulary, and emojis contribute to the overall pragmatic interpretation. Nevertheless, because the target is an individual and the attack is based on a personal physical characteristic rather than group identity, the expression should not be classified as hate speech on the evidence available.

The four expressions demonstrate that negative social media discourse may take several forms. Some comments involve interpersonal hostility, reprimand, mockery, or intimidation, whereas one expression contains a stronger identity based dimension. The first expression is directed toward an individual content creator and contains an implicit intimidating meaning. The second functions primarily as a personal reprimand. The fourth combines a warning with mockery directed toward an individual's physical appearance. None of these expressions provides sufficient evidence that the target is attacked because of membership in a particular religious, ethnic, racial, or other identity group. The third expression differs because it constructs a collective target through *kelompok* and attaches negative characteristics to that group. Within the context of the Ngaben discussion, the target is understood as Hindu devotees who choose cremation as a form of Ngaben. The combination of collective targeting, religious context, and derogatory characterization gives this expression a stronger identity based dimension. The comparison demonstrates that hate speech and illocutionary acts operate at different analytical levels. The hate speech analysis concerns the target, identity dimension, and nature of the negative expression, whereas the illocutionary analysis concerns the communicative action performed by the speaker. An important finding is that the most clearly identifiable hate-speech expression does not function primarily as a directive. The expression concerning *kelompok radikal* performs an assertive function by constructing a negative characterization and an expressive function by communicating condemnation. This finding indicates that hate speech can operate through negative representation and stigmatization rather than only through explicit threats or commands.

The other expressions provide useful contrastive evidence. Their inclusion demonstrates that the analysis does not automatically equate offensive or hostile language with hate speech. Such differentiation is particularly relevant to forensic linguistics because the interpretation of potentially harmful language requires attention to the target, context, linguistic form, and communicative function. Overall, the findings indicate that Balinese language social-media discourse concerning Ngaben contains a range of hostile communicative practices. The most significant identity based expression is the derogatory characterization of the religiously identifiable group associated with cremation. At the pragmatic level, the comments demonstrate expressive, directive, and assertive functions, showing that the social meaning of an utterance cannot be adequately understood through either hate-speech classification or speech act analysis alone.

CONCLUSIONS

This study examined Balinese language comments on social media concerning discussions of Ngaben and cremation practices from a forensic linguistic perspective, focusing on two main dimensions hate speech and illocutionary acts. The analysis demonstrates that hostile or offensive language should not automatically be classified as hate speech. The classification requires consideration of the linguistic form, communicative context, target of the utterance, and whether the expression is directed toward an individual or a group based on a particular social identity.

The analysis of the selected comments shows that the expressions primarily represent individual directed hostility, personal reprimand, and hostile or mocking warnings. The first expression, "*Seme'e angge nanem cai ane ngae konten...wkwwk...*", conveys hostility and implicit intimidation toward an individual, while the second expression, "*URUS IBE-IBE GEN SING PERLU NGURUS TIMPAL NE LENAN,*" functions as a personal reprimand directed at an individual. Similarly, "*De mekejang urus pak, urus iban ci*

Hate Speech in Balinese Language on Social Media: A Forensic Linguistic Perspective

gen pedidi” instructs the addressee not to interfere in other people's affairs and therefore represents a directive form of personal reprimand. The fourth expression, “*AWAS mekeplag nas lengar ee puakkk 🤔🤔🤔*,” conveys a hostile warning accompanied by mockery. None of these expressions provides sufficient evidence of an identity based attack against a group based on religion, ethnicity, race, or another protected characteristic. Therefore, they are more appropriately interpreted as hostile or offensive expressions rather than hate speech under the analytical framework applied in this study.

In terms of illocutionary acts, the findings indicate that the comments perform different pragmatic functions. The first and fourth expressions primarily perform expressive acts accompanied by directive force, as they convey negative attitudes while implicitly attempting to influence or intimidate the addressee. The second and third expressions are predominantly directive, as they attempt to regulate the addressee's behavior, particularly by telling the individual to stop interfering in other people's affairs. These findings demonstrate that the illocutionary function of an utterance cannot be equated directly with its classification as hate speech. An utterance may perform an expressive or directive act while remaining an interpersonal hostile expression rather than an identity based hateful expression.

From a forensic linguistic perspective, the findings highlight the importance of examining linguistic evidence in relation to context, target, and pragmatic function. The analysis of Balinese language social media comments requires attention to the meanings of the original expressions rather than relying solely on translated forms. The study therefore demonstrates that forensic linguistic analysis can help distinguish between general hostility, personal offense, and hate speech while also revealing the communicative actions performed through digital language.

Overall, the study contributes to the understanding of Balinese language discourse on social media by showing how hostile expressions surrounding culturally and religiously significant issues such as Ngaben can be examined through forensic linguistic and pragmatic perspectives. However, the limited number of data and the absence of sufficiently clear identity based hate speech in the selected comments indicate the need for further research using a larger dataset and more varied social media contexts. Future studies should examine a broader range of Balinese language comments containing explicit attacks against social or religious groups in order to provide a more comprehensive account of hate speech and its illocutionary functions in Balinese digital discourse.

REFERENCES

- 1) Akbar, Amin & Nia N. (2019). Tantangan dan Solusi Dalam Perkembangan Teknologi Pendidikan di Indonesia. Prosiding Seminar Nasional Pendidikan, Univeristas PGRI Palembang.
- 2) Allan, K., & Burridge, K. (2006). *Forbidden words: Taboo and the censoring of language*. Cambridge University Press.
- 3) Austin, J. L. (1962). *How to do things with words*. Oxford University Press.
- 4) Bousfield, D., & Locher, M. A. (Eds.). (2008). *Impoliteness in language: Studies on its interplay with power in theory and practice*. De Gruyter Mouton. <https://doi.org/10.1515/9783110208344>
- 5) Brown, A. (2015). *Hate speech law: A philosophical examination*. Routledge
- 6) Coulthard, M., & Johnson, A. (2007). *An introduction to forensic linguistics: Language in evidence*. Routledge.
- 7) ElSherief, M., Kulkarni, V., Nguyen, D., Wang, W. Y., & Belding, E. (2018). Hate Lingo: A target-based linguistic analysis of hate speech in social media. *Proceedings of the International AAAI Conference on Web and Social Media*, 12(1). <https://doi.org/10.1609/icwsm.v12i1.15041>
- 8) Fahmi, A. R., Ranu, A., & Khoiriyah, H. D. (2025). Illocutionary acts and hate speech on Gus Miftah: Analyzing pragmatic implications in social media discourse. *KLAUSA: Kajian Linguistik, Pembelajaran Bahasa, dan Sastra*, 9(1). <https://doi.org/10.33479/klaus.v9i1.1164>
- 9) Ferdiansa, D., Darwis, M., & Said, I. M. (2022). Types of hate speech in Jokowi's Instagram comment column during the PPKM period: Forensic linguistic analysis. *Jurnal Indonesia Sosial Teknologi*, 3(5), 621–631. <https://doi.org/10.59141/jist.v3i05.467>
- 10) Floranti, A. D. (2022). Racism toward Chinese ethnic group in Indonesian social media: Hate speeches analysis from forensic linguistic perspective. *JOMANTARA: Journal of Multicultural Studies in Social Sciences*, 2(2). <https://doi.org/10.23969/ijjac.v2i2.5753>
- 11) Katuuk, N. N. C., Wantasen, I. L., & Imbang, D. (2024). Analysis digital track of hate speech on Instagram using a forensic linguistic. *Jurnal Onoma: Pendidikan, Bahasa, dan Sastra*, 10(4), 3847–3854. <https://doi.org/10.30605/onoma.v10i4.4439>
- 12) Kepolisian Negara Republik Indonesia. (2015). *Surat Edaran Kapolri Nomor SE/6/X/2015 tentang Penanganan Ujaran Kebencian (Hate Speech)*.
- 13) Maharani, P. D., Candra, K. D. P., & Pebriyani, N. P. R. (2022). Kajian ujaran kebencian dalam komentar video pidato Desak Made Darmawati. *Prosiding Seminar Nasional Linguistik dan Sastra*.
- 14) Olsson, J. (2008). *Forensic linguistics* (2nd ed.). Continuum International Publishing Group.
- 15) Permatasari, D. I., & Subyantoro, S. (2020). Ujaran kebencian Facebook tahun 2017–2019. *Jurnal Sastra Indonesia*, 9(1), 62–70. <https://doi.org/10.15294/jsi.v9i1.33020>

Hate Speech in Balinese Language on Social Media: A Forensic Linguistic Perspective

- 16) Putra, V. G. R., Febriyatko, A., & Busri, H. (2024). Hate speech in Instagram comments of DPR RI: A forensic linguistic analysis. *KEMBARA: Jurnal Keilmuan Bahasa, Sastra, dan Pengajarannya*, 10(1). <https://doi.org/10.22219/kembara.v10i1.28128>
- 17) Ramadani, F. S. (2021). Ujaran kebencian netizen Indonesia dalam kolom komentar Instagram selebgram Indonesia: Sebuah kajian linguistik forensik. *Aksara: Jurnal Bahasa dan Sastra*, 22(1).
- 18) Searle, J. R. (1979). *Expression and meaning: Studies in the theory of speech acts*. Cambridge University Press. 19)
- Waldron, J. (2012). *The harm in hate speech*. Harvard University Press. <https://doi.org/10.4159/harvard.9780674065086>
- 20) Tiersma, P. M. (1999). *Legal language*. University of Chicago Press.



There is an Open Access article, distributed under the term of the Creative Commons Attribution – Non-Commercial 4.0 International (CC BY-NC 4.0) (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits remixing, adapting and building upon the work for non-commercial use, provided the original work is properly cited.