

## A Comprehensive Pragma-Linguistic Analysis of 'Imminent Harm' in Social Media Discourse: Speech Act Theory, Pragmatic Inference, and the Limits of Free Speech in Online Contexts

**Rusul Abdulkareem Ato**

University of Misan / English Department

Email: [rusl\\_abdulkareem@uomisan.edu.iq](mailto:rusl_abdulkareem@uomisan.edu.iq)

**Abstract:** *This research explores the way in which social media narratives construct "imminent harm" through a pragma-linguistic lens. It argues that harmful meanings are rarely directly articulated and can often only be interpreted through indirect means (i.e., via context-sensitive means). The researcher will utilize Austin and Searle's Speech Act Theory and Grice's Pragmatic Implicature Theory to develop a comprehensive analytic framework for analyzing how illocutionary intent is encoded and understood within online contexts, integrating elements of Speech Act Analysis with Pragmatics and Selected Aspects of Critical Discourse Analysis. A corpus of 120 social media posts sampled from 3 different platforms (Twitter (X), Facebook, and Reddit) were analysed using a mixed-methods research design to determine what types of speech acts are represented, what pragmatic strategies are used to convey implicit meanings in the speech acts and how discourse-level features can be used to classify an individual speech act as being harmful. Indirect speech acts (i.e., implicit threats/directives) were found to be more harmful than, for example, direct speech acts due to their use of: implication (implicature), ambiguous interpretations and collective framing; as these factors often generated an excessive level of emotional intensity across a large number of people who perceived such speech acts as imminent. The findings illustrate that the distinction between an individual's right to free speech and the individual's right to not be harmed are not always clear-cut, but rather are formed at the intersection of language form, pragmatic inference and social context. Contributions from these findings are made to linguistic theory by establishing the complex nature of harmful discourse detection and regulation, which outlays the need for approaches to be pragmatically informed and context sensitive to capture the implicit meaning of deemed harmful speech and the scope of complexity associated with the very structure of digital communication.*

**Key words:** *Pragmatics, Speech Act Theory, Imminent Harm, Social Media Discourse, Implicature, Indirect Speech Acts, Free Speech, Critical Discourse Analysis, Online Communication, Harmful Discourse*

### 1. Introduction

The way that we communicate and interact as human beings has changed very quickly because of social media. People are now using social media to express their thoughts and views and find an audience for their idea, as well as create conversations that have never occurred before. Twitter, Facebook, and Reddit (among other platforms) have created a new type of place for people to interact by using language as a tool to both inform one another as well as perform social behaviour with the potential for broader consequences. Moreover, with the explosion of new forms of digital discourse, it has become increasingly difficult to define the line between appropriate language and harmful language.

One issue that continues to arise within the realm of contemporary digital discourse pertains to the ability of people to identify language that may lead to harm; this is especially true in situations where harm has not been expressed physically. Historically, researchers studying harmful forms of

speech have primarily focused upon linguistic forms that communicate overt and direct indicatives of harmful behaviour; however, a growing body of scholarly evidence demonstrates that people also convey harmful intent through both indirect and implicit ways, as well as through forms of contextually defined language. As a result, there are many unanswered questions surrounding how meaning is created and conveyed in online settings.

The concept of "imminent harm" is a key element in the continuing debate over freedom of expression in the courts as well as in philosophical discussions. One of the most cited cases which established this rule is *Brandenburg v. Ohio*; whereby limiting free speech can only occur if the speech is likely to bring about immediate unlawful action. But it is questionable if these ideas apply to digital communications yet? In an online environment an individual statement does not always result in harm. The combination of conversations, emotional triggers, and shared conversations all create a cumulative effect that may ultimately result in harm.

How do we analyze harmful meanings from a language perspective? It is necessary to analysis and move beyond the literal meaning of language to analyze how meaningful utterances (i.e., speech acts) perform actions which include threatening, directing, and persuading individuals to perform acts. J. L. Austin's[1] work on Speech Act Theory and John Searle's[2] development of the Speech Act Theory provide an analytic framework that can be used for understanding how utterances perform actions. Another area that can be useful in understanding the construction of meaning within culture and language of an online context is via pragmatic theory, to include H.P. Grice's[3] work regarding implicature, which discusses how meaning can be implicit rather than explicit. Both these viewpoints have strong implications for how an individual communicates within the digital environment, as many individuals use imprecise and/or indirect ways to communicate verbally.

Many researchers have focused their investigations on the explicit use of harmful language online. However, the implicit nature of certain harmful discourse and the use of pragmatics to infer meaning from discourse has received far less attention. The notion of "imminent harm" has also not been examined using an integrated pragma-linguistic approach that incorporates speech act analysis, discourse analysis and pragmatics. The lack of research in this area prevents us from gaining a full understanding of how harmful meanings operate in our current communication environment.

For the purpose of this research project, the focus will be the pragma-linguistic analysis of the construction of "imminent harm" in social media discourse. It will investigate how harmful meanings are constructed through speech acts and implicitly communicated using pragmatic strategies, as well as how these meanings are shaped by larger discourse patterns. The project will integrate various qualitative and quantitative research methodologies in order to provide a systematic description of the linguistic and contextual characteristics that signal harm.

The importance of this research comes from making contributions and adding value to both the theoretical and practical aspects of language use. By providing evidence for how language can be used in digital environments, this research provides additional validation to existing models of language through its study of speech act theory and pragmatic theory within digital contexts. In addition, the research results provide information about detecting and regulating harmful speech; particularly when harm is implied (e.g., through text), rather than assumed. This research demonstrates how the understanding of harmful language should be expanded beyond direct expression, and instead understood as arising from inferences of the context, social interactions, etc., that are highly contextualized in the digital world.

### **Speech Act Theory and Online Discourse**

The Speech Act Theory is a foundational theory that theorizes how language acts as a social activity through both traditional and digital means of communication. As originally theorized by J. L. Austin, speech acts are viewed as not just propositional but as doing something (a speech act), through the delivery of the utterance made. Austin based his understanding of this process on the three types of acts that occur in every utterance – locutionary, illocutionary, and perlocutionary acts. In this way, all utterances have an illocutionary force that represents the speaker's conception of meaning and

intention; therefore, the meaning and intention behind an illocution are both incorporated within the illocution. This is further supported by J. Searle, who categorizes the forms of illocutionary acts within the speech act framework into four major categories: assertive, directive, commissive, and expressive. Therefore, speech act theory provides the groundwork for understanding how speech acts occur.

As far as digital communication is concerned, the ways that the speech act framework is applied have changed considerably. The following elements of digital communication – anonymity, asynchronicity, and reduced context – all play a significant role in how users engage and create speech acts. The ability of users to utilize these three elements has resulted in a heavy reliance on indirect forms of speech acts, especially when attempting to communicate complex or sensitive meanings. The Illocutionary force behind uttering speech acts can be defined by J. Searle[4] as the use of indirect speech acts when meaning is not explicitly encoded in the form of the language but instead must be derived through understanding the context within the digital environment.

In studies conducted recently that examined how people engage with each other via social media platforms, it was discovered that users very often use strategic vagueness or ambiguity to carry out actions like making threats against others, criticizing others, as well as inciting others to take action[5][6]. The use of vague or ambiguous language gives the speaker the ability to maintain plausible deniability while still accomplishing their communicative goal/error within social media platforms. Additionally, due to the nature of being an online environment, there are no physical social constraints (e.g., body language, facial expressions, etc.) that would normally limit how flexible or extreme a speaker can be with their language; thus, the speaker is able to use more flexible and sometimes more extreme types of language as compared to if they were engaged in face-to-face communication[7]. The perlocutionary (or effect) aspect of speech acts is important when examining how speech acts are created in the digital environment. Most online speech acts were designed (or intended) to produce some type of effect on the audience (e.g., persuasion, mobilization, emotional response, etc.). The viral nature of social media enables the amplification of perlocutionary effects and allows a single speech act to impact a large audience relatively quickly[8].

Therefore, although Speech Act Theory remains relevant in examining online speech acts, it does require some adaptation in order to adequately understand and explain the implicit nature, context dependent nature and socially mediated experiences present in digital communication. Because online communication is continuous in nature, and due to the dynamic nature of communication, each speech act is not an isolated event; rather, each speech act is part of a larger dynamic interaction process where meaning is negotiated between the speaker (linguistic), the intention of the speaker (user), and the interpretation of the audience (receiver).

### **Pragmatics and Implicit Meaning**

Pragmatics is essentially about meaning and how we construct and interpret it in context of use: beyond the literal semantic meaning of utterances. defines pragmatics as the relationship between language and users of language, mainly how speakers indicate their intended meaning and how listeners infer such meaning, according to the context in which the communication occurs. This perspective is very important for identifying implicit meaning, especially when direct expression is avoided or strategically minimized within an environment.

Implicature is one of the most significant aspects of pragmatic theory proposed by H.P. Grice as a principle of communication based on the Co-operative Principle, which is defined by a set of conversational maxims that allow speakers to communicate meaning that is additional to what they have explicitly stated. Implicature produces meaning that is not present within the linguistic form but results from a process of inferencing based on a common ground between interlocutors, the context, and their assumptions[9][10]. Implicature is becoming more salient in digital communication, especially on social media sites, where users frequently utilise indirect/implicit modes of expression to discuss sensitive topics, controversial subjects and meanings that could potentially be damaging either to themselves or to others. This use of indirect/implicit forms of expression is in part due to the moderation of platforms by discouraging explicit harmful language while allowing for ambiguity or indirectness. Therefore, often the construction of meaning comes from contextual cues, shared

ideologies and audience inferences as opposed to explicit linguistic markers.

In addition, indirectness, ambiguity, and the use of figurative language (e.g., irony or sarcasm) are very much connected to implicit meaning in online communication. This is because the speaker uses these strategies when encoding their intended meanings in a form that can be interpreted by an audience, but can be denied if the meaning is examined in a more global way[11][12]. Consequently, the communicative dimensions of strategic usage and manipulation of the relationship between what was said and what was intended by the speaker is an important aspect of pragmatic use, according to the Relevance Theory proposed by Sperber and Wilson[13]. A significant component of implicit meaning, therefore, is the degree to which it relies on context to provide additional meaning to the original intended meaning. Through this theory of Relevance, interpretation has been seen as the selection and use of the most relevant meanings based on cognitive and contextual factors. Thus, because online communication often has fragmented context or is distributed across multiple posts, the interpretation process in online environments is complicated to the extent that the users will need to reconstruct the intended meaning from very few (or no) direct contextual cues.

Studying pragmatics and implicit meaning is essential for the analysis of contemporary digital communication. This shows that communication can be made up of more than just what is plainly stated (i.e., the linguistic form), it is also about how we are able to understand things in their context and build understanding of what they mean (i.e., our inferential reasoning process). Harmful discourse provides an example of how potentially harmful meanings can exist in what might appear to be non-threatening or ambiguous contexts, and, therefore, are much more difficult to detect when restricting your analysis to the surface level.

### **Free Speech and Harm in Digital Contexts**

The way that harmful speech and freedom of expression relate to one another is very complex now, especially because so much expression takes place through digital media. Freedom of speech is a basic right throughout democracies. Nonetheless, both society and authorities believe that limits on freedom of speech should be applied, based on various kinds of expression's danger of causing various types of harm. The traditional legal framework focuses, for example, on the *Brandenburg v. Ohio* case where courts have ruled that speech is limited if it is intended to create or incite unlawful actions and there is a valid reason to believe that such unlawful action will occur. The *Brandenburg* standard focuses on the intent of the speaker and the immediacy of any resulting harm, meaning that a clear link must exist between the speech and any resulting harm.

At the same time, some scholars argue that the legal definition of speech causing harm may not fully represent what constitutes harm in a digital world. For example, Waldron[14] suggests that harmful speech has the potential to produce social and psychological harm even when the harmful speech does not produce physical harm immediately. Likewise, Benesch[15] created the term "dangerous speech," which means speech causing increased risk of violence by influencing or changing people's attitudes toward one another, and thereby increasing acceptance of violence as a means of expression over an extended period of time.

Harmful discourse in the online world typically occurs via indirect, cumulative, and/or networked means rather than through explicit incitement to harm others. Instead, users may unintentionally contribute to harm through repeated non-verbal expressions of hatred, dehumanizing or excluding others, etc. "None of the behaviours above would occur without some previous expression of hostility or dehumanisation of those individuals/their group; therefore, the harm that results from these behaviours may not come from any one individual's action at one moment, but rather may accumulate over time from multiple individuals expressing hostility or dehumanising behaviour."

Digital platforms are also changing how speech operates or is shared due to characteristics they exhibit, such as the anonymity; virality and algorithmic amplification of speech content. The more emotional engagement a piece of content elicits in an audience member; the more likely that content is to be shared and, as a result; the more likely that content will have an impact on others[16]. Therefore, harmful speech is more dependent upon its circulation and how it is received than on the

form of the language alone.

The challenge is that moderation systems, which generally use explicit (words) to determine whether a post conveys meaningful harm, do not typically account for the meaning of harm conveyed indirectly or implicitly. Users are more likely to adopt ambiguous or coded language when they want to express harmful ideas without being detected[17]. As a result, the line between free speech and harmful speech is constantly changing and dependent on the context in digital media because it can't be decided based only on the literal meaning of the words. To draw the line requires consideration of pragmatic inferences, conversational patterns, and social contexts to interpret what is meant in each case by determining whether someone will experience a degree of harm by receiving that message. Understanding harmful communications in general will involve understanding not only implicit meanings (what is implied) but also collective aspects of each communication and differences among societies related to technology/distribution.

### **Theoretical Framework**

The study on how "imminent harm" is formed through social media shows that language is not something separate from our actions but rather a type of social behavior that happens in unique situations and has an actual impact. Therefore, the research will look beyond the literal meaning of what was said to see how the types of words used, the meaning inferred from context, and social influences share effects on the way harmful messages are expressed. The basic theory for an understanding of speech as an action is called Speech Act Theory and it was originally outlined as well as expanded. The distinction made by Austin between locutionary, illocutionary and perlocutionary acts emphasises the fact that meaning in language is not only found when something is said but also when the speaker attempts to convey something through their choice of words and what the receiver actually hears after listening to the speaker say something. The focus will primarily be on the illocutionary force of speech acts related to "imminent harm" (for example: threat, directive and expressive evaluations).

Acts in digital contexts are often executed indirectly, thus requiring the analyst to make inferences about the meaning of an act or event based on contextual information rather than explicit linguistic markers. The second element of the model is rooted in pragmatics with reference to H. P. Grice's theory of implicature. Grice theorised that speakers are able to communicate meaning in addition to the literal meaning of their utterance through the use of inferencing structures derived from shared assumptions and contextual knowledge. Implicature is important in social media communications because it permits the user to express meaning that is potentially harmful or controversial in an indirect and ambiguous manner. The reliance on implicit meaning represents a shift from explicit forms of communication to more contextually dependent forms of expression whereby the act of interpretation becomes an essential part of understanding the intent of the speaker. The implications of relevance theory provide additional clarity regarding the way in which an audience interprets these meanings by selecting the most contextually relevant interpretation of the act based on cognitive and contextual resources that are available.

This project incorporates a social and ideological view on discourse by applying Critical Discourse Analysis, focusing upon Fairclough's[18][19] approach to CDA as a means of engaging with advanced complexities of discourse in a holistic manner. By examining language as a form of social practice reflecting & creating pattern of group identities, power relations and ideological position. Many times, online forms of harmful discourse are rooted in collective boundaries that separate in-group members from out-group members, thereby legitimating hostile behaviour and reinforcing social division. The creation of meaning is thus seen through the use of various discourse networks in the creation of meaning over time; and therefore will provide a social framework for constructing meaning based on wider discourses rather than just through an individual's cognitive ways of creating meaning.

When taken together these perspectives (SAT, pragmatics, CDA) provide a comprehensive view of how "imminent harm" is socially constructed through the use of online forms of communication. The SAT allows us to examine the functional characteristics of language, while pragmatics enables

us to determine the explicit and implicit meanings expressed in those languages and the meaning assigned to those languages by society's social dynamics and through the use of wider socio-ideological systems. All three forms of analysis create a singular and integrated analysis structure that show how harming individuals through digital communication occurs through the interplay of: 1) the linguistic form (the words that were used), 2) the use of context to infer meaning, and 3) the influence of society's social dynamics on how to understand and respond to language.

## **2. Methodology**

### **Research Design**

The goal of this study is to investigate the nature of 'imminent harm' in relation to social media through both quantitative and qualitative contextual data gathered through the use of a mixed-methods pragmatic-linguistic methodology. This compares both of these forms of data to determine how they are used to express harmful intent (indirectly) in a particularized context. The qualitative data examines how people express their intent indirectly through language that is determined by context and the quantitative data looks at how frequently and how to characterize entities using rare language features. The use of the two types of data provides a better understanding of how meaning is made and then derived from digital communications.

### **Data Source and Corpus Construction**

This data set contains 120 public social media posts from Twitter, Facebook, and Reddit; It contains a combination and variety of social media outlets that have the highest level of user engagement; a variety of discourse styles; and a high incidence of controversial or potentially damaging content. The corpus was developed to provide a variety of communicative contexts for a better understanding of what online discourse and discourse practices look like.

### **Sampling Strategy**

In selecting relevant posts to research purposes using purposive sampling, posts that contained a form of language that was perceived to have harmful intent or to be threatening or aggressive were selected as long as those posts appeared to have an illocutionary force being directed at another person with the possible intention of causing harm and contained adequate contextual information for the purposes of performing a pragmatic analysis. Conversely, posts that contained only informational content, did not have enough context for interpretive analysis or were automatically generated were not included in the study sample. The focus of this action was to ensure that the analyses would ultimately utilize linguistically and pragmatically meaningful examples.

### **Data Collection Procedure**

The method used to collect the data was performed in a structured manner. The first task in this process was to create a keyword list of harmful and threatening keywords (such as attack, destroy, they deserve) that could be utilized for identifying and collecting the posts of interest. After locating the identified keywords, the next step was to search for the respective posts using specific search tools for each platform. After the posts were initially retrieved, the final step was to apply preset inclusion criteria to each of the identified posts to determine if they were eligible to be included in the final sample. The final sample contained a wide range of variability in terms of both the number of posts and the level of balance between platforms.

### **Analytical Framework**

Written in this way, the following passage illustrates how I applied an integrated theoretical framework of Speech Act Theory, Pragmatics, and Critical Discourse Analysis. Each post's dominant act type (i.e., direct (command), commissive (promise), assertive (statement), expressive/emotive (emotion)) was coded using John Searle's Speech Act Theory (1969, 1979). Given the prevalence of indirect speech acts (i.e., statements made in such a way that the meaning being conveyed is not stated plainly, but instead is conveyed indirectly through inference from other statements), a major focus of

my analysis was on the concept of implicature, proposed by H.P. Grice (1975). Finally, at the level of the discourse, the emotional tone of each post; alignment to a group; targeting of an audience; and relative time urgency were the main features coded.

#### Coding Scheme

To create a systematic coding framework, we first developed a set of coding rules and guidelines that could standardize how coding would be performed across all samples. Each sample received a code based on multiple variables, including speech act type, explicitness of the speech act, pragmatic strategy, presence of a target, emotional tone and level of harm. As such, this type of coding methodology was used to ensure both the quality of qualitative and quantitative data through the consistent application of a coding process to the entire dataset.

#### Reliability and Validity

To check reliability, a second coder conducted an independent analysis on a limited number of data samples, after which inter-coder agreement was assessed using Cohen's Kappa. The results indicated that there was a high level of agreement amongst coders. Theoretical triangulation, data triangulation through multiple platforms, and the combination of qualitative and quantitative methodologies helped to strengthen validity. All of these steps enhanced the credibility and consistency of the study's findings.

#### Ethical Considerations

In keeping with accepted ethical guidelines for conducting research using digital media, this study only used data publicly available; thus, there was no use of personally identifiable information to analyze or present data (user anonymity will be protected) that was collected as an academic study; data is only used for academic purposes.

#### Limitation

There are limitations of this study. The data set consists of only the posts written in English; thus, results from this study cannot be used for generalizing beyond this sample size. The coding for implicit meaning was necessarily done subjectively; however, the researchers made great efforts to minimize subjectivity through the use of a systematic coding process and checks for reliability. In addition to the above, each social network uses their own algorithm to determine which posts to send users; thus the algorithm used by each social network may have influenced which posts in this study and were included in the data that were then included in the study database.

### 3. Results and Discussion

**Table 1.** Speech Act Types in Harm-Related Posts

| Speech Act Type         | Frequency | Percentage (%) | Pragmatic Function        |
|-------------------------|-----------|----------------|---------------------------|
| <b>Direct Threats</b>   | 22        | 18%            | Explicit harm intention   |
| <b>Indirect Threats</b> | 38        | 32%            | Implied harmful intent    |
| <b>Directives</b>       | 30        | 25%            | Mobilizing others         |
| <b>Expressives</b>      | 18        | 15%            | Emotional intensification |
| <b>Assertives</b>       | 12        | 10%            | Justifying harm           |

#### Interpretation

The majority of harmful speech acts consist of both direct and indirect threats; nevertheless when looking at the results overall, it can be seen that indirect threats (32%) greatly surpass direct threats (18%) within a considerable margin. The evidence shows that people use indirect illocutionary strategies on social media to create harmful discourse much more than explicit illocutionary performatives. From a pragmatic perspective, this implies that users purposefully avoid using direct performatives (for example, "I will do harm") in favour of contextualised utterances that produce an

equivalent perlocutionary effect but which remain ambiguous. This also illustrates the separation of illocutionary force from surface structure as a common feature of digital discourse. The remaining speech acts are ways of framing harm as a collective action and thus diffuse responsibility to others, thereby enhancing the persuasive power of the discourse; specifically, directives represent twenty-five percent of total speech acts. Expressive acts, at fifteen percent, serve as placative expressions of emotion and support hostile and aggressive attitudes. Collectively, the data suggest that imminent harm will not reflect a single linguistic category but instead reflect an aggregate pragmatic effect of multiple speech act types.

**Table 2.** Pragmatic Strategies Used to Encode Harm

| Strategy                  | Frequency | Percentage (%) | Function              |
|---------------------------|-----------|----------------|-----------------------|
| <b>Implicature</b>        | 42        | 35%            | Indirect meaning      |
| <b>Collective Framing</b> | 32        | 27%            | Shared responsibility |
| <b>Metaphor</b>           | 24        | 20%            | Symbolic harm         |
| <b>Humor/Sarcasm</b>      | 22        | 18%            | Masking intent        |

#### Interpretation

According to the data in Table 2, implicature (35%) is the primary means by which users of social media convey harmful meaning. This illustrates that social media users rely much more on pragmatic inferences rather than on specific or explicit linguistic expressions. Grice's theory of implicature suggests that meaning is developed through shared beliefs/assumptions rather than purely directly expressed. In this case, the use of implicature enables speakers to express harmful intentions without being held to direct accountability and thus, serves as a significant strategic use of language. Furthermore, collective framing (27%) is instrumental in altering individual views into group positions which create an increased level of legitimacy in the use of harmful discourse. The use of inclusive pronouns such as "we" creates a sense of urgency and moral obligation amongst members of the discourse as well as demonstrating a commonality of interest or purpose. In addition, metaphor (20%) and sarcasm (18%) serve as discursive barriers to presenting harmful content in less direct or confrontational ways. This serves to diminish the seriousness of the content, while allowing for the author's intent or objective to be preserved. Overall, the way that harmful discourse is expressed/encoded in an indirect and socially constructed manner demonstrates that these techniques are prevalent throughout the use of language.

**Table 3.** Pragmatic Features Signaling Imminent Harm

| Feature              | Frequency | Percentage (%) |
|----------------------|-----------|----------------|
| Emotional Intensity  | 84        | 70%            |
| Group Alignment      | 72        | 60%            |
| Target Specification | 66        | 55%            |
| Temporal Urgency     | 48        | 40%            |

#### Interpretation

The data suggests in Table 3 that emotion (70%) is the strongest predictor of imminent harm and that threat perception closely correlates to escalation of emotion, as opposed to simply being based on an explicit that has been made. It also supports the notion that practical cues such as tone, volume, and emotion are more frequently used to identify harm than the literal meaning of a word. Furthermore, emotion can enhance the seriousness of an utterance, as well as increase its perlocutionary effect.

Group alignment (60%) also contributes to harm construction by positioning all individual statements within a unified ideological framework, thus providing legitimacy to a given message while simultaneously enhancing its persuasiveness. Identifying a specific target or group (55%) also

provides important context for defining potential harm because it shifts the way we conceptualize hostility from abstract to specific.

Temporal urgency (40%) is also necessary for demonstrating immediacy; however, it is considerably less common than the factors previously discussed. The results indicate that imminence is a multidimensional pragmatic concept and, as such, is characterized by emotional, social and contextual factors, not just temporal characteristics.

**Table 4. Degree of Explicitness in Harmful Posts**

| Type                 | Frequency | Percentage (%) |
|----------------------|-----------|----------------|
| <b>Explicit Harm</b> | 34        | 28%            |
| <b>Implicit Harm</b> | 86        | 72%            |

### Interpretation

Table 4 shows that implicit harm is represented through the dominant type of implicit overt (i.e. covert or incidental) and thus most of the harmful online discourse is likely to be covert as opposed to overt. This highlights the importance of pragmatics in interpreting online discourse. Since many forms of harmful meaning do not have an explicit presence, the identification of harmful discourse relies upon contextualisation, common ground, and inferences. The preponderance of implicit harm also indicates that users have knowledge about the platform's regulations and therefore utilize certain forms of language that enable them to express harm without having their expression caught by the platform's moderation systems. Thus, from a theoretical perspective, this reinforces the idea of meaning in digital discourse is at the level of illocutionary intentions as opposed to locutionary content. Therefore, the distinction between free speech and harmful discourse becomes more unstable due to the necessity of interpretation versus having specific linguistic markers.

**Table 5. Combined Pragma-Linguistic Features**

| Category           | Sub-Feature         | Frequency (%) | Function              |
|--------------------|---------------------|---------------|-----------------------|
| <b>Speech Acts</b> | Indirect Threats    | 32%           | Implicit harm         |
| <b>Speech Acts</b> | Directives          | 25%           | Mobilization          |
| <b>Pragmatics</b>  | Implicature         | 35%           | Hidden meaning        |
| <b>Pragmatics</b>  | Collective Framing  | 27%           | Shared responsibility |
| <b>Discourse</b>   | Emotional Intensity | 70%           | Escalation            |
| <b>Structure</b>   | Implicit Harm       | 72%           | Covert aggression     |

### Interpretation

The construction of imminent harm linguistically occurs along various level as indicated in Table 5, which illustrates a comprehensive representation of this. The data presented show that an independent feature cannot define harmful discourse; thus, the parameters for defining harm are founded on the interaction of speech acts, pragmatic strategies and the discourse features of language. Harm is represented through both indirect threats and implicature which is central to their structural make-up; therefore, meaning is nearly always inferred rather than being direct. Thus, this indicates a movement towards communicative practices being covert, with speakers encoding harmful intentions into socially acceptable forms of language without overtly stating their harmfulness. At the discourse level, emotional severity (70%) and group cohesion both enhance the affective and ideological aspects of harm and convert single instances of harm into a collective experience. The interplay between emotion and identity has significantly increased the perception of urgency and legitimacy attached to the harmful content. Additionally, the predominate presence of implicit harm (72%) further supports the notion of digital discourse existing in an environment of restricted expression, whereby users adjust their language usage in order that they do not incur moderation while still being able to effectively communicate. Thus, the finding from this study suggests that the best approach to

understanding "imminent harm" is as a multi-layer pragmatic-linguistic construction emerging through the interaction of indirectness, escalation of emotion and collective experience.

**Table 6.** Examples of Harm Construction

| Example                             | Type     | Speech Act | Strategy           |
|-------------------------------------|----------|------------|--------------------|
| <b>They will regret this soon</b>   | Implicit | Threat     | Implicature        |
| <b>Someone needs to stop them</b>   | Implicit | Directive  | Collective framing |
| <b>Let's show them what happens</b> | Implicit | Directive  | Group alignment    |
| <b>I will attack them</b>           | Explicit | Threat     | Direct             |
| <b>This won't end peacefully</b>    | Implicit | Assertive  | Ambiguity          |

#### Interpretation

Table 6 shows how implicit harm depends on a shared understanding of context in which the message is interpreted by the audience rather than being encoded directly by the speaker. Indirect or future-oriented language, as demonstrated in indirect threats such as "They will regret this later," serves as a pragmatic marker of harm without committing to a specific act of harm. Other examples include direct requests framed in the form of suggestions ("Someone must prevent them from doing so") to blur the line between providing an opinion and promoting an incitement. Group-based language ("Let's show them") illustrates how the collective identity of a group has the potential to compound the harmful implications of an act because of the dispersal of responsibility among all individuals in the group involved. On the contrary, explicit threats are infrequent and straightforward to communicate (in a linguistic sense) because they do not have the same higher level of complexity (i.e., the strategic ambiguity) that characterizes the implicit threat. Collectively speaking, these examples show that inference rather than declaration is the most common method for constructing harmful discourse, and thus a pragmatic analysis would be necessary.

## 4. Discussion

The results of this research suggest that constructing the concept of imminent harm using language on social media is not simply a linguistic matter, but instead is a context-dependent, pragmatically negotiated process. In addition to relying on specific words with straightforward meanings, users construct various indirect, contextually ambiguous linguistic structures, which causes the interpretation of potential harm to shift from the textual level at the end of the communication process, to the inferential level as it is processed by the recipient. Using John Searle's speech act theory, this study's findings indicate that illocutionary force is realized through a different mechanism in the digital environment compared to the traditional face-to-face environment. As opposed to explicitly threatening or injuring someone through a direct and unequivocal act of speech (i.e., sending a threatening text message), the findings show that individuals speaking in the digital realm are increasingly relying on indirect forms of illocution, where the intended message is understood by inference rather than through grammar. Increasingly, this method of communicating and interacting will enable individuals to pursue socially risky activities such as threatening or encouraging harm while retaining a level of linguistic deniability.

The importance given to inferred meaning provides further support for H.P. Grice's theory of implicature. Previously, implicature was just the cooperative conversational strategy but now, there is a more strategic use of implicature whereby speakers encode sensitive or possibly harmful messages. In doing so, they rely on mutual knowledge, shared ideology, and contextual clues to ensure that the audience understands the speaker's intention but there is no direct statement of that intention. This change in understanding of implicature provides insight into the development of pragmatic principles used in mediated interaction through digital technology.

Finally, the findings also highlight the important role of collective identity development in the creation of harmful discourse. The use of plural pronouns and group-oriented expressions often

frames language in ways that transform individual opinions into collective views; in line with broader discourse analytic models (e.g., those developed by Norman Fairclough). This suggests that language is used as a resource for creating social reality and supporting systems of power. Under these conditions, harmful acts are not singular actions; rather, they are part of collective narratives that provide justification for and amplify the expression of hostility. A final emerging point from this analysis relates to emotional intensity as a measure of harm; data suggests that the amount of perceived urgency is directly related to affective phenomena, such as anger, urgency, and hostility.

The findings show that expressing harm is not only about the meaning of a specific statement but also incorporates how someone feels about it emotionally and how their interaction impacts the meaning of what was said. Consequently, many scholars will need to change how they think about these issues and include all the various emotional/contextual factors when they conduct research on whether or not something is harmful.

At the same time, the way implicit meaning is often used in contrast to explicit means there is an ongoing conflict between a person's right to express themselves freely and digital platforms regulating their expression. Based on how people adapt their language in response to digital platforms' expressed limitations on the form and content of their language, it is clear that there will still be harmful discourse occurring; however, rather than using an explicit/directly apparent form of language, the people will restructure what would have otherwise been straightforward harmful discourse into something that is implicit, ambiguous, and/or contextually dependent. As such, automated systems that restrict harmful discourse based on the lexical content of language could potentially miss such indirect/circumlocutory/harmful discourse because they are only looking at the surface (i.e., the lexical characteristics) of the language that is being used by the user within a close time period before/after the content was posted. The integrative analysis also shows that the "imminent" quality of harmful discourse cannot be attributed to one linguistic feature but instead occurs through the interaction of several layers of language (e.g. speech act type, pragmatic strategy, emotional tone, social symmetry).

Multi-dimensionality is indicative of the necessity of a comprehensive analytical framework for capturing the complex nature of digital communication. In conclusion, this study adds to the expanding corpus of literature that reconceptualises language from the standpoint of social action within online environments. The findings show that harmful discourse cannot solely be viewed as content, but also should be viewed in terms of implications, the speaker, and sociocultural and ideological contexts of interpretation. Thus, in order to either regulate or analyse this type of discourse, one must go beyond a literal understanding of discourse and explore the deeper pragmatic structures which underlie communication in today's digital world.

## 5. Conclusion

From a pragma-linguistic point of view, the aim of this study was to explore how 'imminent harm' is constructed during social media conversation. The data demonstrate that harm is rarely expressed with an explicit linguistic form; rather, the majority of harmful meanings are derived from indirect (non-linguistic), implicit, and contextually specific strategies that place reliance on the interpretative abilities of audiences. Furthermore, speakers appear to consistently use indirect speech acts to convey the intention to cause harm, most notably by using indirect threat and/or directive speech; the result is that they remain ambiguous as they do this, creating an evident difference in how illocutionary force is realised, where meaning is no longer evident through the literal structures of utterances, but instead is derived from pragmatic inferences and contextual interpretations.

This research provides evidence that implicature and collective framing can exert a disproportionate effect on the creation and maintenance of harmful communication. Speakers use these strategies to create encoded messages with sensitive meanings that are effective socially and subtle linguistically. The combination of social and linguistic properties of discourse (e.g., emotionality/intensity; alignment) magnifies the perceived urgency or seriousness of an utterance,

creating an event(s) or experience(s) representative of the construct of "imminent" as an event(s) and/or experience(s) through the interpretation of meaning. This study contributes to the understanding that the boundary between speech that is free and speech that is harmful is an unstable division based on the interdependence of language, context, and the interpretation of language as opposed to the independent use of explicit linguistic markers or defined criteria in the identification of speech that is harmful. As a result, the use of surface text analysis as the primary method for identifying harmful speech provides little value in determining the content of harmful speech.

This research will greatly impact both theory and practice. Theoretically speaking, it further develops Pragma-Linguistics by illustrating how Speech Act Theory and Pragmatics operate in online places, such as when communicating through sponsorship. Practically speaking, it indicates that more sophisticated methods for moderating content should be developed, which would consider situational relevance, implied meaning, and pragmatic inference. There are limitations to this research, as it consists solely of an English dataset and only contains a small number of total posts. Future research may look into cultural differences between different groups of individuals or communities as well as multilingual discourse, and also include the use of multimodal elements (e.g., photo or meme) to communicate harmful meaning.

## References

- [1] J. L. Austin, *How to Do Things with Words*. Oxford, U.K.: Oxford University Press, 1962.
- [2] J. R. Searle, *Speech Acts*. Cambridge, U.K.: Cambridge University Press, 1969.
- [3] H. P. Grice, "Logic and conversation," in *Syntax and Semantics*, vol. 3. New York, NY, USA: Academic Press, 1975.
- [4] J. R. Searle, *Expression and Meaning*. Cambridge, U.K.: Cambridge University Press, 1979.
- [5] M. Dynel, "'I has seen image macros!'," *Discourse & Society*, vol. 27, no. 6, pp. 660–688, 2016.
- [6] M. Kissine, *From Utterances to Speech Acts*. Cambridge, U.K.: Cambridge University Press, 2013.
- [7] S. C. Herring, "A faceted classification scheme for computer-mediated discourse," *Language@Internet*, vol. 4, 2007.
- [8] A. Marwick and R. Lewis, *Media Manipulation and Disinformation Online*. New York, NY, USA: Data & Society, 2017.
- [9] S. C. Levinson, *Pragmatics*. Cambridge, U.K.: Cambridge University Press, 1983.
- [10] J. Thomas, *Meaning in Interaction*. London, U.K.: Longman, 1995.
- [11] S. Attardo, "Irony as relevant inappropriateness," *Journal of Pragmatics*, vol. 32, no. 6, pp. 793–826, 2000.
- [12] M. Dynel, "The landscape of impoliteness research," *Journal of Politeness Research*, vol. 11, no. 2, pp. 329–354, 2015.
- [13] D. Sperber and D. Wilson, *Relevance: Communication and Cognition*. Oxford, U.K.: Blackwell, 1995.
- [14] J. Waldron, *The Harm in Hate Speech*. Cambridge, MA, USA: Harvard University Press, 2012.
- [15] S. Benesch, "Countering dangerous speech to prevent mass violence," *Genocide Studies and Prevention*, vol. 9, no. 3, pp. 45–55, 2014.
- [16] W. J. Brady, J. A. Wills, J. T. Jost, J. A. Tucker, and J. J. Van Bavel, "Emotion shapes the diffusion of moralized content," *Proceedings of the National Academy of Sciences (PNAS)*, vol. 114, no. 28, pp. 7313–7318, 2017.
- [17] E. A. Jane, "Flaming? What flaming?" *Continuum*, vol. 29, no. 1, pp. 65–77, 2015.
- [18] N. Fairclough, *Critical Discourse Analysis: The Critical Study of Language*. London, U.K.: Longman, 1995.
- [19] N. Fairclough, *Critical Discourse Analysis: The Critical Study of Language*, 2nd ed. London, U.K.: Routledge, 2013.