**PUBLIC COMMENT ON THE STATEMENTS ABOUT THE JAPANESE PRIME MINISTER**

**Ayako Hatano, University of Oxford**

The deletion of the phrase "くたばれ (*kutabare*)" (a phrase that translates roughly to "drop dead" or “go to hell” or “f\*ck”) in the Japanese context touches on a variety of sensitive issues surrounding online threats, freedom of expression, and the balance between protecting individuals and maintaining open political discourse. Here are some comments on the case, considering the sociopolitical landscape in Japan and Meta's policies:

**Sociopolitical/Legal Context in Japan**

First of all, it is important to consider the nature of political expression, which is more protected under international human rights law and the Japanese constitution. The International Covenant on Civil and Political Rights (ICCPR), to which Japan is a party, protects freedom of expression, including political speech. Article 19 of the ICCPR states that everyone shall have the right to hold opinions without interference and the right to freedom of expression, which includes the freedom to seek, receive, and impart information and ideas of all kinds.

Similarly, the Japanese Constitution safeguards freedom of expression. Article 21 explicitly guarantees freedom of assembly and association as well as speech, press, and all other forms of expression. This protection extends to political speech, which is considered vital for a functioning democracy. Political expression is given a higher level of protection because it contributes to public discourse and the democratic process.

**Online Threats Against Politicians**

Japan, like many countries, has seen a rise in online threats against politicians. High-profile incidents, such as the assassination of former Prime Minister Shinzo Abe, have heightened awareness and sensitivity toward online rhetoric that could incite violence. Politicians in Japan often face harassment and threats on social media, reflecting broader global trends.

**Limitations on Freedom of Expression**

Japan has laws that can restrict freedom of expression, especially when it comes to defamatory or threatening language. Article 230 of the Penal Code criminalizes defamation, and threats of violence can lead to criminal charges under various statutes. These legal frameworks can sometimes be seen as limiting critical speech, but they are often justified by the need to maintain public order and protect individuals from harm.

Cf. **Revision of the Penal Code: raise the statutory penalty for insult**

In June 2022, Japan’s Penal Code was revised concerning Article 231, which addresses public insults. The amendment raised the statutory penalty for insults from less than 30 days' detention to up to one year's imprisonment, and elevated the fine from less than 10,000 yen to up to 300,000 yen (around $2,200). Additionally, the statute of limitations for insults was extended from one year to three years, allowing a longer window for legal action to be taken against perpetrators.

The primary motivation behind this revision was to address the escalating issue of cyberbullying and its detrimental impact on individuals and society. The tragic suicide of professional wrestler and reality TV star Hana Kimura, who was subjected to relentless online harassment, highlighted the severe consequences of cyberbullying and spurred these legislative changes.

(See other developments of laws at <https://futurefreespeech.org/japan/>)

**Rhetorical Threats in the Japanese language and political discourse**

**Prevalence and Distinction from Credible Threats**

Rhetorical threats or language, like "くたばれ (*kutabare*)," are not uncommon in political discourse and online interactions in Japan. However, distinguishing between hyperbolic statements and credible threats can be challenging. Cultural nuances and the context in which statements are made play significant roles in this distinction. While hyperbolic threats are often seen as part of heated political debates, credible threats are taken seriously by authorities and platforms alike.

**Rabat Plan of Action and the Case of "くたばれ** (***kutabare***)**"**

The six-part threshold test in the Rabat Plan of Action, developed by the United Nations, is designed to assess whether expressions of hatred constitute incitement to discrimination, hostility, or violence, which would justify restrictions on freedom of expression under international human rights law. Here are the six criteria:

1. **Context**: The social and political context within which the speech occurs. This includes factors such as the political environment, existing tensions, and the potential for the speech to incite harmful actions.
2. **Speaker**: The position or status of the speaker. The influence or authority of the speaker can affect the potential impact of the speech. High-profile speakers such as politicians, religious leaders, or media figures can have a greater potential to incite violence or discrimination.
3. **Intent**: The intention behind the speech. Whether the speaker intended to incite discrimination, hostility, or violence is a crucial factor. This can be difficult to determine but is key to distinguishing incitement from other forms of expression.
4. **Content and Form**: The content of the speech and the form it takes. This includes the degree to which the speech is provocative, inflammatory, or violent in its language and imagery. The form (e.g., speech, written, broadcast) can also affect its impact.
5. **Extent of the Speech**: The extent and reach of the speech act. This includes the size of the audience and the means of dissemination (e.g., social media, mass media, small gatherings). Wider dissemination increases the potential impact.
6. **Likelihood of Harm**: The potential for the speech to lead to actual harm. This involves assessing whether the speech is likely to incite real-world discrimination, hostility, or violence. The probability and immediacy of harm are key considerations.

These criteria are used collectively to assess whether restrictions on freedom of expression are warranted in a given case. They help ensure that limitations on speech are applied carefully and only in cases where there is a clear and present danger of incitement to serious harm. The Rabat Plan of Action addresses issues related to hate speech and incitement to violence. While the content in this case does not constitute hate speech (as it specifically targets Prime Minister Kishida), this six-part test is useful for understanding the socio-cultural context of the speech. Here, I apply this framework to the case involving the use of "くたばれ (*kutabare*)" directed at Japanese Prime Minister Fumio Kishida:

**1. Context**

The context involves a social media post where a user replied to a news article about an alleged political funding scandal involving Prime Minister Kishida and the LDP. The comment contained the phrase "くたばれ (*kutabare*) " directed at Kishida. However, given that the political funding scandal is widely associated with many LDP politicians, this "くたばれ (*kutabare*) " can be considered a political opposition towards the LDP involved in the scandal, or against corrupt politicians, essentially criticizing the bad political deeds of the LDP. Thus, the user argued, this can be considered a form of political expression criticizing alleged misconduct.

**2. Speaker**

The anonymity of the speaker reduces the perceived threat level, as it's often associated with less accountability and influence compared to a high-profile figure. This anonymity may offer a sense of freedom to express opinions or share content without fear of personal repercussions or intense scrutiny. However, it also raises concerns about credibility and accountability, as there's typically no way to verify the identity or intentions of the speaker.

**3. Intent**

Determining intent is crucial, particularly in the context of online communication. The user likely intended to express anger and dissatisfaction with the political scandal rather than to incite actual violence. However, the phrase "くたばれ (*kutabare*)" can be interpreted as a death wish, which complicates the assessment of intent.

It is important to note the significant difference between “くたばれ (*kutabare*)” and "殺す (*korosu*)." The former expression does not convey an intention to harm or kill the target, whereas the latter does. Literally, the former expression may imply the speaker’s hope that the target will die on their own, without the speaker needing to kill it purposefully.

Furthermore, it should also be noted that the phrase “くたばれ (*kutabare*)” has lost its literal meaning and has become slang used towards someone who is annoying to the speaker. It corresponds to “drop dead” in that the expression may seem to imply the death of the target in the literal sense but does not actually do so. Unfortunately, among non-native Japanese speakers, the actual meaning and nuance of “くたばれ (*kutabare*)” is often misunderstood.

In this case, verifying the exact intent of the speaker is challenging due to their anonymity and the absence of social context regarding the author. However, it's worth noting that “くたばれ (*kutabare*)” is often casually used, for example in a feminist campaign named “くたばれ家父長制 (*kutabare kafuchosei*)”, meaning “f\*\*k off patriarchy”, as shown in the graphic below.

Cf.<https://www.facebook.com/marieillustration/photos/a.503694680120303/1249842542172176/?type=3>



The expression “くたばれ (*kutabare*)” and its variant “くたばっちまえ (*kutabatchimae*)” can also be found in lyrics of Japanese pop songs sung by girls’ groups, which suggests that “くたばれ (*kutabare*)” can be used casually and lightly as well.

For example, the phrase “くたばっちまえ (*kutabatchimae*)” is used in the song "Wedding Bell" (ウェディングベル), available here: [ウェディングベル](https://www.youtube.com/watch?v=TRzkoIdECJc&ab_channel=%284K%29%F0%9F%92%BFMVWithJP%28EN%29LyricsSubtitles%F0%9F%93%80)

**4. Content and Form**

The content is a derogatory expression directed at a politician. While it's inflammatory, it is also a common hyperbolic phrase used in political discourse, which often includes strong language. The form (a social media comment) typically has less weight and impact than more formal types of communication like speeches or articles.

**5. Extent of the Speech**

The extent of dissemination was limited; the comment received no likes or replies and was only seen by a small audience before being removed. This indicates limited potential impact or reach.

**6. Likelihood of Harm**

The likelihood of actual harm or incitement to violence appears low given the context and the nature of the speech. However, the phrase "くたばれ (*kutabare*)" directed at a high-profile figure like the Prime Minister can still be perceived as more serious, particularly given Meta’s policies aimed at protecting high-risk individuals from threats.

**Meta’s Policies on Bullying, Harassment, and Violence**

**Protecting High-Risk Individuals**

Meta’s policies on bullying, harassment, and violence are designed to protect high-risk individuals, including politicians and Heads of State, from credible threats. These policies aim to prevent content that can incite real-world violence or cause significant harm. The challenge lies in implementing these policies at scale, given the volume of content and the need to assess context and intent.

**Balancing Protection and Political Expression**

Meta must ensure that its policies do not stifle legitimate political expression, even when it includes rhetorical threats or strong language. Clear guidelines and consistent enforcement are crucial. Meta can use context-based moderation, where the intent, history of the user, and the target of the speech are considered before making content decisions.

**Meta’s Recommendation Policies and Political Speech**

**Impact on Access to Information**

Meta’s choice not to recommend political content on Threads and Instagram newsfeeds for pages not followed by users significantly impacts access to political information and discourse. This decision can reduce the visibility of diverse political opinions and limit the reach of political campaigns and discussions. While this may reduce the spread of harmful or misleading content, it also curtails the vibrancy of political debate on these platforms.

**Ensuring Political Expression**

To respect political expression, Meta can:

1. **Enhance Transparency:** Clearly communicate why certain content is not recommended and how users can access it if desired. It means Meta can provide clear explanations for why certain political content is not being recommended. This could be due to guidelines related to misinformation, hate speech, or other content policies; offer users clear options to access non-recommended content. For example, if a post is flagged and not promoted, users should still be able to view it by clicking a disclaimer or accessing it through direct links. Publishing regular transparency reports that detail the reasons behind content moderation decisions, the criteria used for recommendations, and the impact on political content visibility is helpful.
2. **Promote Civic Literacy:** Educate users on how to find and engage with credible political content outside of algorithmic recommendations. Provide users with educational resources that explain how to identify credible sources of political information. This could include guides on fact-checking, understanding bias, and evaluating the credibility of sources. Meta can even host workshops, webinars, and online courses on digital literacy, media literacy, and critical thinking, focusing on navigating political content, partnering with fact-checking organizations to offer in-platform tools that help users verify the accuracy of political information.
3. **Highlight Diverse Perspectives/ Avoid Echo Chambers**: Modify algorithms to ensure that users are exposed to a balanced mix of content, including perspectives they might not typically engage with. Avoid over-personalising content to the point where users only see viewpoints similar to their own.

**Conclusion**

Balancing the deletion of potentially harmful phrases like "くたばれ (*kutabare*)" with the need to protect political expression is complex. Meta's policies must navigate the thin line between preventing real threats and supporting robust political discourse. Ensuring clear guidelines, context-aware moderation, and user education can help Meta manage these challenges effectively.

----------------------