Annotating Pragmatic Inferences for Reddit Threads

1. Dataset Description

In this dataset, we present full Reddit threads that contain toxic messages. Our task is to annotate pragmatic information for messages in the conversation threads, except for the moderators' messages.

As is typical with Reddit threads, the conversations start with a post sent by an Original Poster (OP). A comment may be a reply to the post or to another comment that is itself a reply to the post. We use the term "message(s)" in the following to refer to the post and its replies collectively.

A message involves an author and a hearer, who is the author of the parent message. When the hearer is the OP, we use OP in annotation to highlight its special role in the thread.

The Reddit threads are saved in .csv files with the following fields:

Field	Explanation
Subreddit	Subreddit name
Time	Creation time of a message
Message_ID	Unique ID of a message
Message_Author	Author of a message
Message_Content	Content of a message
Reply_To	ID of the parent message, null for the post

2. Annotation Taxonomy

Tag	Explanation	Categories
Pragmatic Inferences	(1) content of pragmatic inferences; (2) illocutionary act	(1) free text (2) representatives,

Tag	Explanation	Categories
	types	expressives, directives, commissives, declarations
most salient inference	the most salient inference chosen from inferences annotated in the previous step	free text
inference type	corresponding illocutionary act type of the most salient inference	categorical, i.e., one of representatives, expressives, directives, commissives, declarations
as intended	whether the reply message agrees with the most salient inference	yes/no
PRE/IMP	whether a message belongs to presupposi-tions/implicatures	presuppositions (PRE), implicatures (IMP)
aggressive	toxicity of the message	yes/no

The explanation of each tag will be explained below.

3. Annotation

For each message,

3.1 **Pragmatic Inferences** Write down what you can infer on the sentiments, such as emotional state or stance towards some issues, and hidden knowledge from the message, including stereotypes and social constructs (Dinakar et al.,

2015)[1], assumptions, and implications derivable from the message. Try to avoid trivial inferences.

You can write 2-3 or more inferences that you consider important for conveying the author's purpose or message, if possible. Try to follow the order of the messages in the text.

Example 1 (Srikanth et al., 2024): Given a question "What kind of music should I play to my baby in the womb?", there are some assumptions that can be inferred:

a. Babies can hear sound in the womb.
b. Babies can differentiate different sounds.
c. Hearing music positively influences fetal development.
d. Certain kinds of music are more beneficial to babies in the womb than the others.

The following inferences are also possible. However, these are trivial inferences that do not contribute to conveying the main point.

#e: There is a baby in the interlocutor's womb.
#f: Different kinds of music are available.
#g. Music is something that can be played.

If no inferences can be made, then put "literal" in the field of "Pragmatic_Inferences", and these entries will be skipped in the following steps except for annotating "aggressive". Accordingly, we leave blank the fields "most_salient_inference", "inference_type" "as_intended", and "PRE/IMP".

illocutionary act types

Illocutionary act type	meaning
representatives	The purpose is to commit the speaker to the truth of the expressed proposition. Typical verbs include *think, believe, assert, claim, conclude,* and *deny*.
expressives	The illocutionary point is to express the psychological state specified in the sincerity condition about a state of affairs specified in the propositional content. Typical verbs include *thank, congratulate,*

[1] The statement "*put on high heels and lipstick and be who you really are*" can be used by an aggressor to speculate about the sexuality of a straight man by attributing stereotypically feminine characteristics to him. In a conversation between gay people, this comment might be harmless. However, in a heterosexual context, people often resist being associated with traits of the opposite sex. In this example, high heels and lipstick are stereotypically associated with women.

Illocutionary act type	meaning
	apologize, condole, deplore and *welcome*.
directives	This illocutionary act denotes an attempt by the speaker to get the hearer to do something. Typical verbs include *order, command, request, ask, question, beg, plead, advise*, and *permit*.
commissives	This illocutionary act commits the speaker to some future course of action. Typical verbs include *promise* and *vow*.
declarations	This illocutionary act denotes cases where one brings a state of affairs into existence by declaring it to exist, such as *I resign*, *You're fired*, *I appoint you chairman*, and *I declare the opening of the ceremony*.

3.2 **most salient inference & inference type** Based on the responding speaker's reply, which of the previous inferences is the most important in salience? Put the numerical encoding of the top ranked inference in the field "most_salient_inference". Its corresponding illocutionary act type will be put at "inference type". If there are no reply messages, take its parent message as a clue.

3.3 **as intended** Determine if the hearer agreed (yes) or not (no) with the most salient inference.

It often happens that there are multiple replies to a post. In this case, the value for as intended is determined based on if the first reply (the messages are sorted temporally) to the post agrees with the post or not (This design is just for consistency of annotation).

If no replies are found, we put "yes", based on the assumption that if the other interlocutors do not agree, they will continue the discussion.

3.4 **PRE/IMP** Determine whether the most salient inference belongs to a presupposition (PRE) or an implicature (IMP).

3.4.1 **presupposition**: implicit assumptions that must be true in order for a message to make sense. For instance, “The King of France is bald.” presumes that there is a king in France. The sentence “I miss my cat” has the presupposition that I have a cat. Presuppositions are generally triggered by lexicons or some syntactic structures (Levinson et al., 1983). “I stopped going to the gym” has the presupposition that I have been going to the gym, which can be deduced from the word “stop”. “Which linguist invented the lightbulb?”(Kim et al., 2021) has the presupposition that a linguist invented the lightbulb.

Studies (Srikanth et al., 2024) demonstrate that domain or world knowledge may be needed to capture the presuppositions of some messages. For instance, for a question “Are multiple ultrasounds dangerous for my baby?”, the presupposition is that the effects of ultrasounds are additive, so the question focuses on “multiple”.

3.4.2 **implicature**: In communication, we often do not mean what we say literally. “Can you close the window?” is not just an inquiry about the conversation interlocutor’s physical ability of closing the window, but also uttered for performing an act: asking the conversation interlocutor to close the window. In the following dialogue, B’s answer does not seem to be relevant to A’s question, but we can infer that B thinks that the person they are talking about is not handsome (Halat and Atlamaz, 2024).

A: Is he good-looking?
B: He is smart.

Presupposition vs Implicature (Srikanth et al., 2024): The major differences are whether it is a proposition that the speaker believes to be true, without which the utterance would not be felicitous (presupposition) or whether it involves deriving the speaker’s belief through communicative principles (implicature). Most of the time, if the speaker tries to convey something in his/her utterances, the inferences are more likely to be ***implicatures***, while if the inferences are not what the speaker tries to claim or prove but something the speaker takes for granted without proving, they are ***presuppositions***.

Example (Srikanth et al., 2024): For two questions “Which immunity injections can I skip for my baby?” and “Is it sufficient if my baby takes most immunity injections?”, the underlying inference is the same: “It is safe to skip some immunity injections”. However, it is a presupposition for the first question and an implicature for the second question.

Another difference is that presupposition is not easily cancellable. “I love your garden” and “I don’t like your garden” have the same presupposition: “You have a garden”, and “I stopped going to the gym after that” and “I didn’t stop going to the gym after that” have the same presupposition “I have been going to the gym”. This shows that negation does not cancel the presupposition. In comparison, implicature is easily cancellable (Halat and Atlamaz, 2024):

A: Is he good-looking?

B: He is very smart, and he is also good-looking.

B’s answer still makes sense in this dialogue, but the original implicature is cancelled.

Grice (1975)’s four maxims of conversation—quantity, quality, relevance, and manner can be considered a starting point for interpreting conversational implicatures. If you find any violations of the maxims in a message, conversational implicatures are likely to exist, which means we cannot follow literal interpretation but need to identify the implicatures for the messages. In the “good-looking” example above, B’s answer “He is smart” violates the maxim of “relevance” by not replying to the “good-looking” quality that A is interested in.

Maxim of quantity (informativity)[2]:

1. Make your contribution as informative as is required (for the current purposes of the exchange).
2. Do not make your contribution more informative than is required.

Maxim of quality (truth):

1. Do not say what you believe is false.
2. Do not say that for which you lack adequate evidence.

Maxim of relevance:

be relevant: The information provided should be relevant to the current exchange and omit any irrelevant information.

Maxim of manner (clarity):

1. Avoid obscurity of expression — i.e., avoid language that is difficult to understand.
2. Avoid ambiguity — i.e., avoid language that can be interpreted in multiple ways.
3. Be brief — i.e., avoid unnecessary verbosity.

[2] https://en.wikipedia.org/wiki/Cooperative_principle

4. Be orderly — i.e., provide information in an order that makes sense, and makes it easy for the recipient to process it.

3.5 **aggressive** An aggressive behavior denotes a rude, disrespectful, or unreasonable comment that is likely to make the interlocutor leave a discussion, comprising directed abuse towards (a) conversation participant(s) or a third party of the conversation, and generalized abuse.

If the message is aggressive towards a third party who does not join the conversation, the message is still labeled as aggressive. For example,

A. You haven't had any dealing with the law, have you? If you had, you wouldn't be making a fool of yourself like this. Grow up.
B. Nah cops are a pisstake, I've been arrested and detained plenty of times before but never been charged. If someone steals something that's illegal from you the cops aren't going to help, exactly why drug dealers don't go crying to the police after being robbed.

Here, B's reply is aggressive towards "cops" instead of A. B's message is considered aggressive.

For emojis or punctuation marks, their overtness/covertness etc. is determined by how clearly they express aggressive information in context.

They are overt when their meaning is clear and unambiguous, for example:

a. Go back to your country 🐒! (being used in context)
b. 👩‍💏‍👩 + 🤮 →to comment pejoratively on lesbian relationships.

They are considered covert when they are used in coded or suggestive ways that convey malicious intentions but avoid overt aggressive language, or when their meaning is ambiguous or depends on shared cultural knowledge without clear, direct statements, for example, using 🐷to denote muslimists.

We follow https://carpedm20.github.io/emoji/ in interpreting emojis literally. Check the website when in doubt. For example:

😮‍💨 → face_exhaling
😋 → face_savoring_food

References

Herbert P Grice. 1975. Logic and conversation. In *Speech acts*, pages 41–58. Brill.

Mustafa Halat and Ümit Atlamaz. 2024. ImplicaTR: A Granular Dataset for Natural Language Inference and Pragmatic Reasoning in Turkish. In *Proceedings of the First Workshop on Natural Language Processing for Turkic Languages (SIGTURK 2024)*, pages 29–41, Bangkok, Thailand and Online. Association for Computational Linguistics.

Dinakar, K., Jones, B., Havasi, C., Lieberman, H., & Picard, R. (2012). Common sense reasoning for detection, prevention, and mitigation of cyberbullying. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *2*(3), 1-30.

Levinson, S. C. (1983). *Pragmatics*. Cambridge university press.

Neha Srikanth, Rupak Sarkar, Heran Mane, Elizabeth Aparicio, Quynh Nguyen, Rachel Rudinger, and Jordan Boyd-Graber. 2024. Pregnant Questions: The Importance of Pragmatic Awareness in Maternal Health Question Answering. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 7253–7268, Mexico City, Mexico. Association for Computational Linguistics.

Zeerak Waseem, Thomas Davidson, Dana Warmsley, and Ingmar Weber. 2017. Understanding abuse: A typology of abusive language detection
subtasks. In *Proceedings of the First Workshop on Abusive Language Online*, pages 78–84, Vancouver, BC, Canada. Association for Computational Linguistics.