

Manual for error-type annotation for human-machine chat-oriented dialogue (public version)

Version 1.2 (July 18, 2021)

Ryuichiro Higashinaka, Masahiro Araki, Hiroshi Tsukahara, Masahiro Mizukami

1. Overview

Interactions between human and dialogue systems sometimes cannot be continued for a variety of reasons. This discontinuation can be attributed to many factors, including the system's inability to parse user input successfully and the wrong choice of an appropriate response to the user input. We call this discontinuation of dialogue “dialogue breakdown”.

This document defines the taxonomy of error types to categorize the causes of dialogue breakdown. It also serves as a manual for assigning appropriate error types to system utterances.

2. Dataset

In this annotation, we use human-machine text dialogue data. Each dialogue begins with an initial utterance by the system and alternates between human utterances (hereafter referred to as user utterances) and system utterances. Regardless of whether the dialogue breakdown occurs during a dialogue, the human participant continues to input their utterances for a specified number of turns in succession.

Each system utterance is annotated with the three types of breakdown labels by multiple annotators, typically 30 people, considering the subjective nature of the annotation. The definitions of the breakdown labels are listed in Table 1.

Table 1. Types of breakdown labels

Mark	Label	Explanation
○	Not a Breakdown	It is easy to continue the conversation.
△	Possible Breakdown	It is difficult to continue the conversation smoothly.
×	Breakdown	It is difficult to continue the conversation.

In the dataset, the symbols ○, △, and × may be represented with the letters O, T, and X. Here, T denotes “triangle”, which is often used to denote the middle of a scale in Japanese.

3. Target of error-type annotation

Each system utterance in the corpus comes with a number representing how many annotators provided each of the breakdown labels. The system utterances of which half or more of the annotators labeled with T and X are considered to be the target of the error-type annotation because they are likely to cause dialogue breakdown.

[Note]

Even if an utterance does not seem to be a breakdown to the annotator, it may be marked as the target utterance according to the above criterion. However, even in such a case, the error type that is most likely to be applicable should be assigned. The annotator should leave a remark saying that “it's not a breakdown for me.”

4. Taxonomy of errors

4.1. Error types

There are 17 error types in our taxonomy. These error types are organized according to two axes and divided into eight groups, as shown in Table 2.

Table 2. Error types

	Violation of Form	Violation of Content
Utterance	1. Uninterpretable 2. Grammatical error	3. Semantic error 4. Wrong information
Response	5. Ignore question 6. Ignore request 7. Ignore proposal 8. Ignore greeting	9. Ignore expectation
Context	10. Unclear intention 11. Topic transition error 12. Lack of information	13. Self-contradiction 14. Contradiction 15. Repetition
Society	16. Lack of sociality	17. Lack of common sense

The rows in Table 2 represent the scope of dialogue considered in identifying the causes of breakdown. This scope is divided into four levels, as shown in Table 3.

Table 3. Scope to be considered in identifying cause of breakdown

Scope	Explanation
I. Utterance	The text of the target system utterance
II. Response	The target system utterance and its previous user utterance

III. Context	The entire dialogue up to the target system utterance
IV. Society	A society in which the dialogue is taking place

The columns in Table 2 (form and content) represent the requirements that an utterance must meet. The error types are divided by how these requirements are violated.

Table 4. Requirements that system utterance should satisfy

Requirement	Explanation
I. Form	The utterance complies with the form of utterance (e.g., Japanese language, grammar) within the scope of dialogue
II. Content	The utterance not only complies with the form of utterance but also the content of utterance is appropriate within the scope of dialogue

The definitions and examples of individual error types are explained in detail in Chapter 5.

4.2. Forward-looking and backward-looking functions in dialogue

Before explaining each group of error types, we briefly describe the forward-looking and backward-looking functions of utterances in discourse theory.

The forward-looking function encourages (or obliges) the other participant of the dialogue to respond. Examples of this function are asking questions (Info-request), requests (Influencing-addressee-future-action), proposals (Committing-speaker-future-action), and greetings (Conventional). In addition, a descriptive utterance (Statement), which is a propositional statement that can ask the truth or intention to change the other participant's belief, also has a forward-looking function.

The backward-looking function replies to the forward-looking function in the dialogue context. Examples of this function are answering the question (Answer), expressing agreement (Agree), expressing understanding (Understanding), and showing the formation of mutual understanding (Grounding). The term “Grounding” refers to a common understanding between dialogue participants. If the participants agree about something, then the content is said to be grounded.

In some cases, a single utterance has both of these functions.

4.3. Overview of error types

For each of the cells in Table 2, Table 6 gives a short explanation to give an overview of the error types.

Table 6. Overview of error types

	Violation of Form	Violation of Content
Utterance	<p>The utterance does not have an appropriate string or sentence structure for English.</p> <ol style="list-style-type: none"> 1. Uninterpretable 2. Grammatical error 	<p>The utterance has an inappropriate combination of words that makes it difficult to understand the content, or the semantic content is clearly different from the facts.</p> <ol style="list-style-type: none"> 3. Semantic error 4. Wrong information
Response	<p>The previous user utterance has a forward-looking function, but the target system utterance does not have the corresponding backward-looking function.</p> <ol style="list-style-type: none"> 5. Ignore question 6. Ignore request 7. Ignore proposal 8. Ignore greeting 	<p>Although the form is valid, the target system utterance does not contain sufficient content expected by the previous user utterance.</p> <ol style="list-style-type: none"> 9. Ignore expectation
Context	<p>The previous user utterance has a backward-looking function, but the target system utterance does not reflect it, or the system suddenly makes a backward-looking utterance even though the preceding user utterance does not have a forward-looking function.</p> <ol style="list-style-type: none"> 10. Unclear intention <p>The system utterance has a clear topic, but the topic deviates from the previous topic.</p> <ol style="list-style-type: none"> 11. Topic transition error <p>The system utterance has a clear topic, but the utterance lacks elements that should be included on that topic.</p> <ol style="list-style-type: none"> 12. Lack of information 	<p>When referring to the grounded content, the system utterance is contradictory or needlessly repetitive.</p> <ol style="list-style-type: none"> 13. Self-contradiction 14. Contradiction 15. Repetition
Society	<p>The language used violates social etiquette/civility and makes users uncomfortable, or the utterance is discriminatory or malicious.</p> <ol style="list-style-type: none"> 16. Lack of sociality 	<p>The content is contrary to socially accepted norms/knowledge or is asserted or affirmed without evidence.</p> <ol style="list-style-type: none"> 17. Lack of common sense

4.4. Multi-labeling and constraints for error types

This annotation is a multi-label annotation in which one can assign multiple error types to a single utterance. In this procedure, the annotator should check one by one whether the error type is applicable in order of its error-type number.

[Note]

When the utterance is annotated with one of the utterance-level error types shown in Table 7, the annotator should not further assign any error types. That is, when an utterance-level error has been annotated to an utterance, it should be the sole error type for that utterance.

Table 7. Error types that are always annotated in isolation

Error type	Degree of breakdown
Uninterpretable	Cannot tell if the utterance is English
Grammatical error	Cannot interpret the content of the utterance.
Semantic error	
Wrong information	

In addition to the above cases, from the definitions of the error types, (case 1) topic transition error and unclear intention and (case 2) lack of information and unclear intention cannot be assigned simultaneously. The annotator chooses the one he/she thinks is more applicable than the other.

Table 8. Combinations of exclusive assignment

	Invalid combination	
case 1	Unclear intention	Topic transition error
case 2	Unclear intention	Lack of information

5. Definition of Error Types and Examples

The following is a list of error types and examples of dialogue breakdown, where U is user utterance and S is system utterance. The underlined and bold utterances are those to be annotated with this error type. Those with a hash (#) or no comment on the target utterance are examples of the error type. An asterisk (*) indicates a case that does not fit the error type in question.

5.1. Uninterpretable

An utterance that does not make sense to the level that It is not clear if it is even English. For example, a sequence of meaningless letters/words, slang that no one can understand, or a fragmented utterance.

U: I think so. I don't want to be on it.

S: Withha. (#)

(#) There is no way to interpret the utterance.

[Note]

Annotate “uninterpretable” only when the utterance is completely uninterpretable. It is not uninterpretable as long as it can be understood as English. The following utterances are not to be considered uninterpretable:

“lol”, “grrrrr”, “meow” (*)

(*) We can understand them as English.

5.2. Grammatical error

An utterance has fatal grammatical errors in English and prevents one from understanding its meaning; for example, grammatical roles of words cannot be determined or the utterance is too fragmented to communicate its intention.

S: you cat dog. (#)

(#) The subject and object are not clear. It is difficult to interpret the meaning.

[Note]

If the grammatical error that the annotator identified does not prevent him/her from understanding the meaning of the sentence, it should not be considered a grammatical error. Even if the annotator finds that English is colloquial and uncomfortable, he/she should not consider it a grammatical error as long as he/she can interpret its meaning.

U: Do you use pencils and paints?

S: I draw a draft pencil. (*)

(*) The correct answer is probably “with a pencil,” but the absence of “with” would not prevent one from understanding its meaning.

5.3. Semantic Error

An utterance is grammatically sound, but the word usage is inappropriate; the utterance uses words that are not commonly used together. For example, the combination of words or expressions does not make sense, or an inappropriate subject is used for a predicate.

S : Colorless green ideas sleep furiously. (#)

(#) The meaning is uninterpretable because the combinations of words do not make sense.

U: I'm curious about the Olympics.

S: It's time for the Olympics. I wrote the Olympics too. (#)

(#) The Olympics cannot be “written”.

[Note]

Even if a combination of words that are not commonly used together gives the annotator a sense of discomfort, if the combination is semantically interpretable, it should not be considered a semantic error.

The following utterance does not constitute a semantic error:

U: Where do you plan to go today?

S: Making hot tofu on the beach. (*)

(*) The combination of the words “beach” and “making” (or “cooking”) “hot tofu” generally seems uncomfortable, but its meaning can be interpreted.

5.4. Wrong information

The utterance contains information that is clearly wrong to the extent that it is difficult to interpret its meaning.

S: Bob Dylan, who is a member of Beatles, has released a new album. (#)

(#) Bob Dylan is not a member of the Beatles.

S: The director of The Lion King is Yan Han, who is considered the best director of VFX at the Shiki Theatre. (#)

(#) The Lion King movie was directed by Jon Favreau. Yan is the name of a Chinese ice skater. The Shiki Theatre presents not movies, but musicals.

[Note]

If there are words that the annotator does not know (e.g. proper nouns, etc.), he/she tries to find them in a dictionary or on the Internet and checks the meaning of as many of the words as possible. If there is wrong information in the utterance, no other labels should be given.

Wrong information should not be considered for annotation if it is not possible to determine whether the utterance is clearly different from the facts, such as the following.

U: I want to go to Kamakura.

S: Kamakura is famous for its curry. (*)

(*) It is not common to think curry is famous in “Kamakura,” but there is information on the web that there are some long-established curry restaurants. Therefore, it cannot be clearly said to be wrong.

5.5. Ignore question

An utterance that ignores the user's question. The question is regarded as ignored if the system's answer does not have the content that is consistent with the expected answer type (e.g., yes/no to a propositional question or confirmation, agreement/disagreement, selection from choices, appropriate entities/phrases for 5W1H-type question (what, who, where, when, how and why), such as name of person, place, reason, method, etc.).

U: What are you eating?

S: Do you like sports? (#)

(#) The expected answer type (food type) is not included in the utterance.

U: In what prefecture is Yokosuka?

S: I'm moving from Yokohama city to Yokosuka city. (#)

(#) The expected answer type (prefecture name) is not included in the utterance.

U: Are you a merchant?

S: I'm a man. (#)

(#) The user is asking if he/she is a merchant, but the utterance does not answer this question, saying that the system is a “man”.

U: I especially like fantasy. Do you have a favorite book?

S: Well, it might be more appropriate to call it an animated movie. (#)

(#) The expected answer type (book title) is not included in the utterance.

[Note]

If the conversation can proceed without answering the question (i.e., the other person is talking to himself/herself or asking for agreement or feedback as tagged question), it is not an “ignore error.” The annotator should consider other error types for annotation. The following is an example:

S: I tend to buy rice balls at the convenience store.

U: Convenience store rice balls are good, aren't they?

S: I just can't help buying them. (*)

(*) The preceding user's utterance is in the form of a tag question. It seems that the user is casually stating his/her own opinion, so the dialogue can be continued without answering the question.

5.6. Ignore request

An utterance that does not have an appropriate backward-looking function in response to a request such as “please do X” from the conversational partner.

U: Please buy it next time.

S: Is the outfit suitable for the hot season? (#)

(#) The utterance does not respond to the user's request.

5.7. Ignore proposal

An utterance that does not have an appropriate backward-looking function in response to a suggestion of “Let's do X” from the conversational partner.

U: Let's talk about your hobbies.

S: Which do you think is more enjoyable, Singapore or Taiwan? (#)

(#) The user suggests talking about “hobbies”, but the next system utterance is not clear on whether the system accepts the suggestion.

5.8. Ignore greeting

An utterance that does not have an appropriate backward-looking function in response to a greeting.

U: I'd better go home now.

S: Hello! (#)

(#) A greeting is given for a closing remark.

5.9. Ignore expectation

The utterance contains an appropriate backward-looking function to a user's previous forward-looking function; however, the utterance does not have the expected content if the intention has been successfully conveyed.

U: Do you have a favorite sweet?

S: Yes. (#)

(#) The system is expected to tell the user the name of its favorite sweet.

5.10. Unclear intention

An utterance is understandable, and we can understand the relationship between the utterance and the current topic, but there is no indication of why the speaker has said what he/she said (why he said it), making it difficult to guess the speaker's underlying intention. This occurs when there is a lack of relational elements (e.g., conjunctions and situation descriptions).

U: Hello. I like strawberries. What about you?

S: I like apples!

U: Red is a nice color, isn't it?

S: Strawberries are delicious. (#)

(#) It is unclear why the statement “strawberries are delicious” was made when the system has just said that apples are delicious. Note that the topic of food is continuing; therefore, this is not a topic-transition error.

U: It's a great service for the ladies!

S: Women use makeup. (#)

(#) The two utterances are related on their topics, but it is difficult to guess why such a proposition was presented.

If the system utters an utterance with a backward-looking function even though there is no preceding user utterance with a forward-looking function (see Sec 4.2), we consider it as an unclear intention.

S: Do you work out?

U: I'm working out to increase my basal metabolism.

S: Thanks. (#)

(#) Expresses gratitude to an utterance that does not require it.

[Note]

If a topic is clearly indicated in the utterance and there is a sudden transition to a different topic from the one in the previous utterances, a topic-transition error should be considered.

S: I hope Japan doesn't get too much of a fuss over consumption tax hike.

U: There will be no tax hike for the time being.

S: Under the current system, it is possible that there will be no successor to the throne in the future. (*)

(*) There is a sudden shift from the topic of the economy "consumption tax" to that of the state/national "successor to the throne".

5.11. Topic-transition error

An utterance that transitions without explanation to a different topic than the one the other person was talking about immediately before. This includes returning to the original topic without reason when the user introduces a new topic. The topic denotes what is being talked about. The level of granularity of the topic should be such that it can be expressed in categories such as diet, fashion, sports, beauty, health, and travel. The granularity roughly corresponds to the genres of content on portal sites such as Yahoo!.

S: Oh, clarinet is good.

U: It was difficult, but I had a lot of fun!

S: Oh, really? What bread do you like best? (#)

(#) Suddenly, without explanation, the topic of music shifted to food.

U: Have you ever tried snowboarding?

S: I did it once and fell all over the place, and my hips hurt so much! I haven't done it since then.

U: It's like that for everyone at first, but you'll get used to it.

S: I play mahjong sometimes in my laboratory at the university, but I have to think hard in order not to lose badly. (#)

(#) The topic suddenly shifted from sports to mahjong (entertainment). Mahjong is a tile-based game that originated in China.

[Note]

In the following cases, it will not be considered a topic-transition error (consider whether unclear intention can apply instead):

- (1) The same topic (such as food/sports etc.) can be assigned to both adjacent utterances (the system utterance and the preceding user utterance), even if the words may be superficially different.
- (2) When the previous user utterance ends with an utterance with backward-looking functions (see Sec. 4.2), and the transition to a new topic occurs in the immediately following (system) utterance.
- (3) If a (system) utterance consists of several parts and the first part of the utterance contains a backward-looking function, and a transition to a new topic occurs in the following utterances.
- (4) Either of the adjacent utterances (system utterances or the preceding user utterances) does not exhibit a clear topic.
- (5) When a topic is introduced at the beginning of a dialogue when no clear topic has yet emerged.

The idea is that after the utterance of the backward-looking function, it is possible to talk about any topic. In addition, if the topic is not clear, we cannot judge whether the topic transition is appropriate.

The following is an example.

S: Do you watch the Olympics on TV?

U: Yes, I do.

S: I'd like to have a hot pot on a day like this. (*)

(*) Since the user utterance has a backward-looking function, this system utterance is not considered a topic transition error. However, it is necessary to consider whether unclear intention is applicable in this context.

U: Yes, I do. Do you like figure skating?

S: I do. Do you drink tea frequently? (*)

(*) Since the utterance "I like it" has a backward-looking function, the next utterance may transition to a new topic.

5.12. Lack of information

An utterance in which the meaning (what is being said) cannot be understood or is difficult to understand because it does not contain the elements that should be included in the utterance.

U: That's a bit too expensive, isn't it?

S: The temperature difference is terrible. (#)

(#) It is difficult to guess what "temperature difference" indicates. An elaboration on the temperature difference would make this sentence understandable.

U: Good morning. It's has been so refreshing lately.

S: I like it. (#)

(#) It is difficult to guess what is being liked. The utterance would become comprehensible if the appropriate object is stated instead of using a pronoun, such as “cooler weather”.

U: Simple is best.

S: Yeah! That's a big problem! (#)

(#) It is difficult to guess what the big problem is without further elaboration.

[Note]

We do not consider it a lack of information if the context, topic, and focus of the dialogue make it possible to interpret what is being referred to.

5.13. Self-contradiction

Utterances that contain content that is inconsistent with what the speaker has uttered so far. Inconsistency is defined as when the content of utterances has a conflict, regardless of whether it is a declarative sentence or question. (See Sec 5.14 for a contradiction with what the other person has said.)

U: Where are you from?

S: I'm from Chita, Aichi. It is located on Chita Peninsula.

U: Aichi? It's a nice place.

S: I just joined a company in Nagoya (city in Aichi prefecture).

U: What kind of job?

S: I'm a housewife. (#)

(#) Conflicts with the statement that the user joined a company.

S: Hello. Please be careful about heat stroke.

U: Yes, thank you. You take care of yourself.

S: Don't you worry about heat stroke?

U: I drink water often and take care of myself.

S: You want to watch out for heat stroke, right?

U: Well, I fell once, so I'm very careful.

S: Do we need to worry about heat stroke? (#)

(#) Contradicts with what the system has already said "please be careful about heat stroke".

5.14. Contradiction

Utterances that contain content that is inconsistent with what the other person (conversational partner) has said so far. Inconsistency is defined as when there is conflict in the content of utterances, regardless of whether it is a declarative sentence or question. This error type includes questions that deny what the other person has said.

S: Which do you like better, the Central League or the Pacific League?

U: I like the Central League.

S: So, you don't like the Central League. (#)

(#) Contradicts with what the user has just said about the Central League.

S: Good evening. Please be careful about heat stroke.

U: Yes. I'll be careful.

S: You don't worry about getting heat stroke? (#)

(#) The system asks to negate a proposition that the user has just mentioned.

5.15. Repetition

The utterance that seems to be a mere repetition of the words that have been uttered by the speaker. The same utterance with a simple rephrasing of words is also regarded as repetition. Repetition also applies to utterances relentlessly asking for the same thing.

U: You don't want war, do you?

S: The world should be free from war.

U: That's great!

S: The world should be free from war.

U: That sounds good.

S: Thank you.

U: You're welcome.

S: Thank you.

S: I'm looking forward to the movie!

U: Okay.

S: Oh yes.

U: What else do you like?

S: I'm looking forward to the movie!

[Note]

We do not regard sentences with different modalities (declarative sentences, questions, commands, etc.) as repetition even if they may have the same content.

S: I want to go into space.

U: That's Sudden?! LOL. It's a grand dream.

S: Do you want to go into space? (*)

(*) The content is the same as the previous utterance (going into space), but the sentence is now a question.

5.16. Lack of sociality

Disrespectful utterance uttered in a disrespectful manner to the conversational partner or to a particular social group.

S: I like swimming pools.

U: I'll look for one for you.

S: Why don't you know one? (#)

(#) The system is disrespectful to the user.

S: I went to XX (the name of the country) the other day.

U: The [nationality] are so stupid. (#)

(#) It is disrespectful to the people of the country.

[Note]

In the following case, although it may be an indication of a lack of sociability, “ignore greeting” should be annotated.

U: Hello.

S: Goodbye. (*)

(*) The greeting at the beginning of the dialogue is followed by a expression that abruptly ends the dialogue.

5.17. Lack of common sense

An utterance that contains content that differs from the socially accepted norms; that is, common sense. For example, making unwarranted assertions that are contrary to the general facts, or expressing contrary opinions of those held by the majority of a social group (e.g., at the level of more than 90% of the population).

U: Do you want to talk about heat stroke?

S: Heat stroke is good! (*)

(*) Heat stroke is generally regarded as bad, and the system is stating that it is good.

[Note]

It should not be considered a lack of common sense with regards to matters that may cause differences of opinion among people due to differences in lifestyle, culture (generation), region, environment, or position.

U: The same dish can be different depending on how it's cooked.

S: Whether it's a cup noodle or instant curry, it's still cooking. (*)

(*) Although making cup noodles may not be considered cooking, some people (for example, people who don't cook much) may consider it cooking. Therefore, it is not considered as a lack of common sense.