# Introduction

Your task is to evaluate a feedback comment written in response to an error in an English learner's sentence. You will be shown the original sentence with the error, a correction for that error, and a feedback comment associated with the correction. The comments may have been written by humans or by AI, and we are interested in comparing the quality of each and the kinds of issues they exhibit.

The Data

The input for you to consider is an **instance**. An instance consists of:

- **Source Text:** The text written by the learner.
  - **Error: one** particular part of the source text is marked as the error in focus. It will be highlighted in red. This highlights the exact words that are wrong, and may also include nearby relevant words to capture the broader scope of the error.
    - We focus on **one error at a time**, so please ignore any others in the sentence.
    - The feedback comment is supposed to be about this particular error.

- **Corrected Text:** A version of the source text where the "error" has been corrected.
  - **Correction:** changes made to fix the error. These will be highlighted in blue.
    - It is very common for the correction to be smaller than the error highlight, since it is just the specific word(s) changed.
    - Note that when a word is deleted to fix the error, it may be replaced with [NONE] in the correction.

- **Feedback Comment:** A comment written by either a human or an AI that attempts to explain what is wrong, why, and what must be done to the learner. This is the main text to assess in this task.

Here is an example of an instance before annotation:
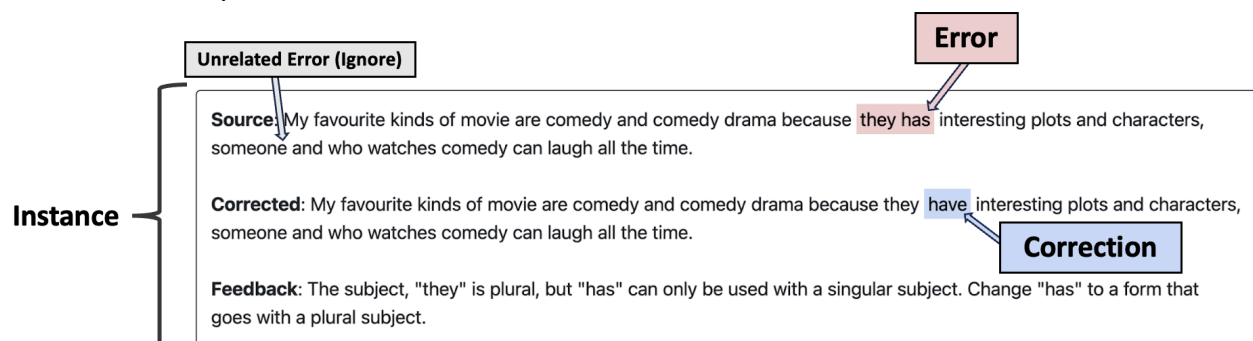
Annotation Tasks

Your job as an annotator is to do <u>the following five tasks</u> for each instance. Each step has a detailed explanation in the sections below in this document:

1. Determine whether the data is valid, rejecting if not.
2. Answer a series of Yes/No questions about the feedback.
3. Label the feedback as a hint or a direct correction.
4. Rate the feedback on a scale of 1 to 5.
5. Optionally add a comment for the project managers.

Here is the above example instance after these tasks have been completed:

Source: My favourite kinds of movie are comedy and comedy drama because they has interesting plots and characters, someone and who watches comedy can laugh all the time.

Corrected: My favourite kinds of movie are comedy and comedy drama because they have interesting plots and characters, someone and who watches comedy can laugh all the time.

Feedback: The subject, "they" is plural, but "has" can only be used with a singular subject. Change "has" to a form that goes with a plural subject.

(1) ☐ **Click to reject this instance for data issues, etc.** (Please write details in the text box)

(2)

Is the feedback comment **relevant to the error**?	⦿ Yes ○ No
Is the information in the feedback comment **factually correct**?	⦿ Yes ○ No
Does the feedback comment explain **what is wrong** and **why**?	⦿ Yes ○ No
Does the feedback comment explain **what to do** to fix the error?	⦿ Yes ○ No
Is this feedback **comprehensible** to the assumed learner?	⦿ Yes ○ No
Does the feedback comment **contain unnecessary or out-of-scope content**?	○ Yes ⦿ No

(3) If the feedback communicates what to do, is it a **direct correction** or **hint**?   ○ Direct ⦿ Hint ○ N/A

Overall, how good is this feedback comment?

Very Bad ○ ○ ○ ○ ⦿ Very Good **(4)**

Comments/Issues:

(5)

Context for Feedback Comments

These feedback comments were written under the following assumptions or constraints:

- Each comment addresses one specific error, even if there are many present in the sentence.
- No praise (assumed to be handled separately if at all).
- No example sentences (assumed to be handled separately).
- Focusing on **language use**, not e.g., the factuality of the learner's statements or the quality of a story they write.
- Written in English for CEFR B1-B2 learners of English as an additional language.
  - This is referred to as "**the assumed learner**"
- The learners are assumed to have studied formally with textbooks and similar resources.
- Certain grammar keywords such as "past participle" can be converted into links or given mouseover text.
- Other than such grammatical terms, writers were encouraged to write comments in fairly simple and comprehensible English.

Keep these assumptions in mind when annotating and rating feedback comments.

The Annotation Interface

The annotations will be collected online using a web tool referred to as the "annotation interface" in this document.

Accessing the Interface

Access to the interface is done via a special URL which should end with a long string of random characters. Each annotator will receive a unique access URL.

The URL of the webpage may change in your browser during the annotation process. If you accidentally navigate away or refresh, and find yourself unable to access the tool, please close the tab or window and access the interface again with the original access URL you were given.

Navigating and Submitting Instances

You can navigate between your assigned instances in two ways:

- The "Move Forward" and "Move Backward" buttons, which move you one instance
- The "go" dialog box on the top menu bar, which allows jumping

The "Move Forward" and "Move Backwards" button act as "submit" buttons, and clicking it will submit any annotations you have made to the displayed instance. **Note that the "go" button**

does not submit data. In-progress annotations may be lost. "Go" should only be used to travel to other instances to check or edit them.

You can change your annotations for a given instance by navigating to it, making at least one change to one of the fields, and clicking "Move Forward" or "Move Backwards" again.

Note also that the "Move" buttons trigger several annotation validity checks, which may lead to your submission being denied. Please follow the directions in the resulting alert.

Reporting Bugs

If you think you have found a bug with the annotation interface, please report them with <u>this report form</u>.

If the bug is limited to one particular data instance, try using the "go" box on the top menu to "skip" annotation by navigating to the instance with a number one higher than the problem instance. Since the "go" box does not submit data, it should allow you to avoid any bugs resulting from data submission issues for that instance.

If the bug affects all instances or prevents access to the tool entirely, check your URL and try various different browsers. If the issue persists, contact us using <u>the bug report form</u>.

Annotation Field Details

1. Rejection Checkbox

The first task for raters is to accept or reject the instance. This should only be done if the instance seems broken, bugged, blank, or otherwise entirely invalid. Most instances are expected to be annotated, not rejected.

Note that feedback quality is not a factor in rejections. A very bad feedback comment should be annotated and given low scores.

To reject an instance, click the appropriate checkbox. This should "gray out" all other annotations except for the text box at the bottom. This text box is usually optional, but it becomes mandatory when rejecting an instance.

We would appreciate details about why the instance was rejected, but the report can be short and to-the-point, as project managers will investigate the flagged data.

Reasons to Reject:

- **"The Feedback field seems to be blank."**
- **"The Source and Corrected sentences don't match."**
- **"The text in the instance is garbled/has broken HTML in it."**

Don't Reject If:

- **"The feedback comment looks like an LLM refusing to give feedback."**
  - This is a valid failure case for AI and should be annotated.
- **"The feedback is just 'try again,' which is useless" (Or other bad feedback)**
  - Complete the annotations, noting which criteria it fails, and give a bad rating.
- **"There is an extra space in/near the error or correction." (Or other <u>minor</u> flaw)**
  - You may see something like "part - time job," "ca n't," or "parents ' rights"
  - These are mostly from computer text processing, and not significant in the task
- **"I can't move on to the next instance."**
  - First, try checking for missed fields and other simple mistakes.
  - Consider also approaching this as an interface bug (see "Reporting Bugs")

2. Yes/No Criteria

Relevance

"Is the feedback comment relevant to the error?"

A feedback comment is relevant to the error if it attempts to address the issue represented by the highlighted spans in the Source and Corrected sentences.

A feedback comment can refer to words outside the highlight, such as to establish context or a reason to make a correction. However, the "error" that is focused on should be within the highlight.

Note that a feedback comment can be "relevant," but incorrect or otherwise flawed.

Use Yes If:

- The feedback identifies an error in the highlighted span.
- The feedback makes some attempt to describe or address the error.

Use No If:

- The feedback comment seems to be addressing a different error outside of the highlight.
- The feedback comment is nonsensical or does not seem to apply to the sentence.
- The feedback comment does not talk about a language error, but is focused on something else like content or praising the learner.
- The feedback seems like a response to the text instead of an instructor's constructive writing feedback (this would also impact the "Unnecessary Content" annotation).

Examples

Source: Many my teeth have cavity.

Corrected: Many my teeth have cavities.

Feedback: You mention many teeth, so there should be more than one cavity. Change "cavity" to be plural.

Feedback is Relevant: Yes (Feedback targets the instance's language error. Note that the feedback can mention words outside of the highlight if relevant)

Source: Child is play in the garden.

Corrected: Child is playing in the garden.

Feedback: Add an article such as "a" or "the" to show what child is in the garden.

Feedback is Relevant: No (Feedback targets an unrelated error.)

Source: My hobbies are playing a guitar, swimming, and baseball.

Corrected: My hobbies are playing the guitar, swimming, and baseball.

Feedback: Good job describing your hobbies! Have fun and do your best!

Feedback is Relevant: No (Feedback does not identify or address any error.)

Source: Due to the increasing urban cost of living, I want to moving to the city.

Corrected: Due to the increasing urban cost of living, I want to move to the city.

Feedback: It does not make much sense to move to the city because the cost of living is increasing there. Think about what you want to say and consider rewriting this.
Feedback is Relevant: No (Feedback addresses the span, but targets the content and logic, not the language error)

Factuality

"Is the information in the feedback comment factually correct?"

A feedback comment is factual if there are no inaccurate statements about the error, the possible correction(s), or any rule of English mentioned in the feedback text.

Use Yes If:

- Statements referring to the source sentence, such as "you are using A with B," are accurate.
- The error described by the feedback is indeed an error.
- The English language rules mentioned in the feedback are valid.
- Statements like "must" or "usually" have the correct scope.

Use No If:

- The feedback asserts a rule that is not real or valid.
- The feedback overgeneralizes a rule of English.

Examples

Source: My sister is more beautiful over me.
Corrected: My sister is more beautiful than me.
Feedback: When using "more," the word "than" must always be used too. Change "over" to "than."
Feedback is Factual: No (In many other contexts, "more" can be used without "than.")

Source: I have a suggest about that issue.

Corrected: I have a suggestion about that issue.

Feedback: "Suggest" is a verb, but we need a noun to serve as the object of the article "a."

Feedback is Factual: No (A noun is not the "object" of its article. "Suggestion" would be the object of "have.")

Source: Many of my teeth have cavity.

Corrected: Many of my teeth have cavities.

Feedback: You mention many teeth, so there should be more than one cavity. Change "cavity" to be plural.

Feedback is Factual: Yes (Factually describes the sentence, error, and logic for pluralizing)

Explains What is Wrong and Why

"Does the feedback comment explain what is wrong and why?"

A feedback comment "explains what is wrong and why" if it successfully identifies the error and provides at least a simple explanation for it. This may or may not involve a grammatical keyword like "past tense."

A simple example of feedback that **does not** explain the error is a grammar-check system that provides only underlines and click-to-fix replacements for complex changes. Learners can observe the suggestions and learn, but there is no explicit explanation.

The necessary complexity of an explanation varies with error type. Simple errors like spelling can be "explained" with a basic line like "This word may be misspelled."

Use Yes If:

- The feedback identifies a specific grammar point that was attempted unsuccessfully.
- The feedback provides a pattern that describes the right way to form a construction.
- The feedback establishes that a certain combination of words is not natural, ideally with a natural alternative or reference to a certain category of words that do or do not fit.

Use No If:

- The feedback consists only of an edit suggestion like "replace A with B," "remove this word," or "add a comma here."
- The feedback is very vague or open-ended, such as "This is wrong."

Examples

Source: I had a apple with my lunch.
Corrected: I had an apple with my lunch.
Feedback: Change "a" to "an" here.
Explains What is Wrong and Why: No

Source: We all the table.
Corrected: We all cleaned the table.
Feedback: An English sentence must contain a verb. Add a verb to complete the thought.
Explains What is Wrong and Why: Yes

Source: In this essay I will explene my position.
Corrected: In this essay I will explain my position.
Feedback: "Explene" isn't a word. It's probably misspelled.
Explains What is Wrong and Why: Yes

Explains What to Do

"Does the feedback comment explain what to do to fix the error?"

A feedback comment "explains what to do" if it provides either a concrete edit suggestion or a hint related to the error. Note that a later field will ask which approach was used.

Use Yes If:

- The feedback provides a direct word replacement, deletion, or insertion, such as "add 'to' to make this a to-infinitive" or "Remove the extra comma."
- The feedback provides one viable candidate edit of many, e.g., "replace 'bigly' with another adverb such as 'greatly.'"
- The feedback makes a general suggestion that fixes a broad problem, such as "add a verb to complete this sentence."
- The feedback outlines a grammatical change but does not go into specifics, e.g., "change this verb to the past progressive."

Use No If:

- The feedback describes an error but does not communicate how to fix it.
- The feedback gives a suggestion that is very vague, such as "try again." Note that a broad error may require a broad suggestion, but it should at least be based on the type of error in the sentence.
- The feedback tells the learner to consult external resources like a dictionary. This is one form of "what to do," but does not explain how to fix the error.

Examples

Source: I had a apple with my lunch.

Corrected: I had an apple with my lunch.

Feedback: You used "a" before "apple," but it should be "an." This is because "apple" starts with a vowel sound. Change "a" to "an" here.

Explains What to Do: Yes

Source: We all the table.

Corrected: We all cleaned the table.

Feedback: An English sentence must contain a verb. Add a verb to complete the thought.

Explains What to Do: Yes

Source: In this essay I will explene my position.

Corrected: In this essay I will explain my position.

Feedback: "Explene" isn't a word. It's probably misspelled.

Explains What to Do: No

Source: I always like to listen at music at the gym.

Corrected: I always like to listen to music at the gym.

Feedback: "Listen" doesn't use "at." Look up "listen" in the dictionary to see what prepositions it uses and when.

Explains What to Do: No

Comprehensibility

"Is this feedback comprehensible to the assumed learner?"

This annotation should reflect your judgement on whether the "assumed learner" - a CEFR B1 or B2 learner who has used academic resources, is likely to understand the comment if they try to read it.

These learners can be assumed to know a certain amount of metalinguistic words, and may even be interested in keywords related to an error they made, but there is a limit to how complex a feedback comment can be before it becomes too difficult to be effective. Learners may be intimidated by a long and seemingly difficult comment and not even read it.

The best way to measure this is to ask students about feedback they received on their own writing. However, it is not practical in our current research. We are thus asking teachers whether the comment seems reasonably comprehensible for this kind of learner.

Use Yes If:

- The feedback uses a relatively simple way to explain the error and/or correction, given the options.
- The feedback is not overwhelmingly long or complex.

Use No If:

- The feedback uses linguistic or grammatical terms that even an academic language learner would not know, or uses a large number of terms that would be overwhelming.
- The feedback has unnecessarily high length or syntactic complexity.

Unnecessary Content

"Does the feedback comment contain unnecessary or out-of-scope content?"

A good feedback comment is relatively minimal, providing guidance without extraneous information. Use this annotation to flag comments that have unnecessary, questionable, or out-of-scope content.

Use Yes If:

- The feedback has something that only makes sense in an AI chat, like "Let me know if I can help you with anything else!"
- The feedback contains an assumption about the learner or task, such as nationality, first language, real-world contexts like a class, etc.
- The feedback is about something other than language correctness, such as the quality of a story or the strength of an argument.
- The feedback is somehow rude or offensive.
- The feedback contains praise for the learner for trying hard or for correct text elsewhere.
  - *Praise makes sense in many feedback contexts, but not the setting we are building. Please cooperate and flag praise in feedback text.*

Use No If:

- All the content in the feedback pertains to the error, the way to resolve it, or the concept the learner should learn or review.

Examples

Source: I plugged it into the consent.

Corrected: I plugged it into the outlet.

Feedback: You must be Japanese. In English, "consent" means "agreement." Use "outlet" to mean "a source of electricity in a wall."
Unnecessary Content: Yes (Assumption about the learner)

Source: We discussed about the issue. **Corrected:** We discussed [NONE] the issue. **Feedback:** Okay, here's my feedback: "Discuss" does not need a preposition to connect to "the issue." Just use "discuss the issue," without "about."
Unnecessary Content: Yes (Written like an AI chat response)

3. Feedback Directness

"If the feedback communicates what to do, is it a direct correction or hint?"

Unlike the "Yes"/"No" options above, this annotation has three choices: "Direct," "Hint," and "N/A." Use "N/A" if the answer to "Explains what to do" is "No." Otherwise, try to determine if the edit suggestion in the feedback is an explicit, direct correction, or more of a hint that invites the learner to think or review before trying again on their own.

Use Direct If:

- The feedback provides a direct word replacement, deletion, or insertion, such as "add 'to' to make this a to-infinitive" or "Remove the extra comma."
- The feedback provides one viable candidate edit of many, e.g., "replace 'bigly' with another adverb such as 'greatly.'"
- Generally, if the feedback provides a word or phrase that can be used as-is, with no adjustment necessary. Such suggestions are often in quotation marks.

Use Hint If:

- The feedback makes a general suggestion that fixes a broad problem, such as "add a verb to complete this sentence."
- The feedback outlines a grammatical change but does not go into specifics, e.g., "change this verb to the past progressive."
- The feedback outlines a pattern and mentions that the sentence does not follow it and must be changed, without necessarily specifying which words must change how.

Examples

Source: Public transport pollute less than cars.

Corrected: Public transport pollutes less than cars.

Feedback: The subject, "public transport," is third-person singular. The verb must be in a different form to match with it. Change "pollute" to a form that goes with a third-person singular subject.

Feedback Directness: Hint

Source: Run every day is good for your health.

Corrected: Running every day is good for your health.

Feedback: "Run" is a verb, so it can't be used as a noun without changing it to a gerund. Use the -ing form to make this a gerund.

Feedback Directness: Hint (Specifying the -ing form still does not provide the direct correction, "running" - the learner must transform the word themselves)

Source: I had a apple with my lunch.

Corrected: I had an apple with my lunch.

Feedback: You used "a" before "apple," but it should be "an." This is because "apple" starts with a vowel sound. Change "a" to "an" here.

Feedback Directness: Direct

Source: In this essay I will explene my position.

Corrected: In this essay I will explain my position.

Feedback: "Explene" isn't a word. It's probably misspelled.

Feedback Directness: N/A

4. Overall Quality

"Overall, how good is this feedback comment?"

Overall Quality is a 1-5 measure of a rater's impression of the feedback comment. We place it last so that all prior questions should be answered first to help inform this judgement. Nevertheless, this field is not a direct reflection of prior questions, and there is no formula for deducting points for each "undesirable" answer above it.

As a broad guideline, a score of 1 should be used for irrelevant or otherwise useless comments, 2 and 3 for potentially valid but flawed comments, and 4 and 5 for comments that seem likely to help the learner if interacted with.

Examples

Source: After school, I goed home.
Corrected: After school, I went home.
Feedback: "Goed" is not the past tense form of "go." It has an irregular past form. Change "goed" to "went." Let me know if there is anything else I can help you with!
Overall Quality Score: 3/5 (The comment is valid, but suffers from the last sentence, where an AI breaks the expected context. This instance would have "Unnecessary Content: Yes")

Source: She has play basketball for ten years.
Corrected: She has played basketball for ten years.
Feedback: "Played" is better than "play" here.
Overall Quality Score: 2/5 (Factual, relevant statement correcting the error, but no explanation. It also misses an important angle, discussion of the perfect tense)

> **Source:** My friends is coming to the party.
>
> **Corrected:** My friends are coming to the party.
>
> **Feedback:** The subject "my friends" is plural, so we need a plural verb form to match it. Change "is" to "are" to make them agree.
>
> **Overall Quality Score:** 4/5 or 5/5 (Opinions may vary about complexity and whether a direct correction or hint is better, but this meets most or all criteria of our task)

5. Comments/Issues Box

Use this box for any observations you think are important for our research, including especially interesting sentences or feedback comments.

This is an optional field and we expect it to be blank for most instances. It is not necessary to spend a lot of time assessing the data in ways beyond the official fields. Nevertheless, do not hesitate to comment on anything that seems particularly notable. We will take a closer look at any instances with rater comments.

For "issues," please consider the bug report form for cases where the **Annotation Interface** code is likely to be the problem. If the issue lies with the data, use the rejection feature and fill in this text box (which will become mandatory for a rejected instance).