Annotation Protocol

Task Description

Together with this annotation protocol, you will receive a username and password to access the **Appraise Evaluation Framework.** After logging in, your task is to annotate the data contained in the platform according to the annotation rules provided in this protocol. There are 80 data points to annotate. We encourage you to use a computer or tablet—please avoid using a phone if possible. Take breaks as needed, but ensure that you carefully review the document throughout the annotation process.

Data

On the annotation page, you will see two fields: the English reference document (at the top) and its corresponding translated document (below), generated by an automatic translation system. The data includes translations for several African languages, such as Amharic, Hausa, Swahili, Yoruba, and Zulu and you will be assigned to the African language that you speak. This is the content you will be asked to annotate.

Your annotation task involves two parts:

1. **Identifying and marking error spans** in the translated document wherever you find issues.

2. **Providing a document translation score** between 0 and 100. A score of 0 indicates that the translation does not preserve the meaning of the English reference at all, while a score of 100 means the meaning is fully preserved.

Tasks

This is adapted from the Error Span Annotation paper

In the following section, you will find instructions related to each of the annotation fields. Please be precise in both your error annotations and your translation quality scoring. We have included test cases within the data to assess the consistency of your annotations. If the quality of your annotations does not meet the required standard, you may be asked to re-annotate the data before payment is issued. Below, we describe the expected annotation process in more detail.

1. **Error span annotation (ESA):** You are tasked with highlighting as many text fragments as you identify as containing translation errors (by clicking or dragging to select the start and end). You can click repeatedly on a highlighted fragment to increase its severity level or to remove the selection. There are two severity levels that you can use when highlighting errors. They are:

    I. Minor Severity: This is a case where the style, grammar, and lexical choices could be improved to sound more natural. Therefore, it is not considered a very severe issue.
    II. Major Severity: This is a more serious case where the translation changes the meaning, is difficult to read, and reduces usability—for example, through repetition of words or phrases.

It's sufficient if you highlight the word or rough area where the error appears. Each error should have a separate highlight.

If something is missing from the translated document, mark it as an error on the [MISSING] or [MISSING].

You are encouraged to annotate as many errors as you can find.

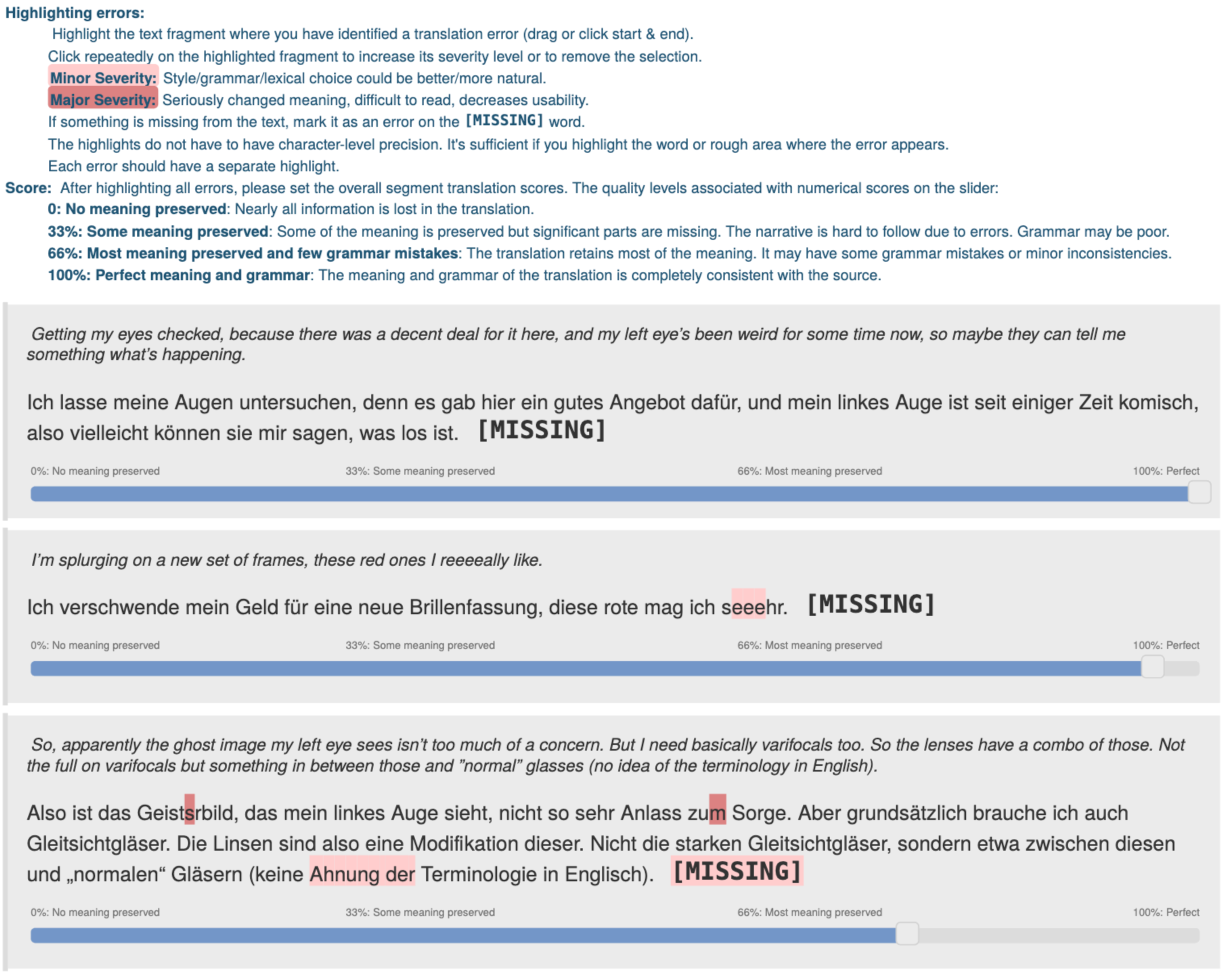
2. **Quality Score:** After the ESA, you are tasked with assigning a quality score using both a Likert scale and a 0–100 scale, where you select a number between 0 and 100 using a slider. This is a form of direct assessment that summarizes the overall quality of the translation. Below is a set of reference points that correspond to the quality levels associated with the numerical scores on the slider:

   0: No meaning preserved: Nearly all information is lost in the translation.
   33%: Some meaning preserved: Some of the meaning is preserved but significant parts are missing. The narrative is hard to follow due to errors. Grammar may be poor.
   66%: Most meaning preserved and few grammar mistakes: The translation retains most of the meaning. It may have some grammar mistakes or minor inconsistencies.
   100%: Perfect meaning and grammar: The meaning and grammar of the translation is completely consistent with the source.

Figure 1 below shows three example cases of English reference sentences and their German translations. In the first case, the German translation is perfect; in the second, it contains an error classified as minor severity; and in the third, there are major errors as well as missing words, so **MISSING** is annotated. Please also refer to Figure 2 below for a more detailed tutorial.

Note: The examples shown in Figures 1 and 2 illustrate sentence-level translations, but in this task, you will work with documents containing multiple sentences (up to 20 sentences per document).

If ou have an uestions or encounter any issues, please reach out to us via email: [redacted]



Highlighting errors:

Highlight the text fragment where you have identified a translation error (drag or click start & end).

Click repeatedly on the highlighted fragment to increase its severity level or to remove the selection.

Minor Severity: Style/grammar/lexical choice could be better/more natural.

Major Severity: Seriously changed meaning, difficult to read, decreases usability.

If something is missing from the text, mark it as an error on the `[MISSING]` word.

The highlights do not have to have character-level precision. It's sufficient if you highlight the word or rough area where the error appears.

Each error should have a separate highlight.

Score: After highlighting all errors, please set the overall segment translation scores. The quality levels associated with numerical scores on the slider:

0: No meaning preserved: Nearly all information is lost in the translation.

33%: Some meaning preserved: Some of the meaning is preserved but significant parts are missing. The narrative is hard to follow due to errors. Grammar may be poor.

66%: Most meaning preserved and few grammar mistakes: The translation retains most of the meaning. It may have some grammar mistakes or minor inconsistencies.

100%: Perfect meaning and grammar: The meaning and grammar of the translation is completely consistent with the source.

Getting my eyes checked, because there was a decent deal for it here, and my left eye's been weird for some time now, so maybe they can tell me something what's happening.

Ich lasse meine Augen untersuchen, denn es gab hier ein gutes Angebot dafür, und mein linkes Auge ist seit einiger Zeit komisch, also vielleicht können sie mir sagen, was los ist. **[MISSING]**

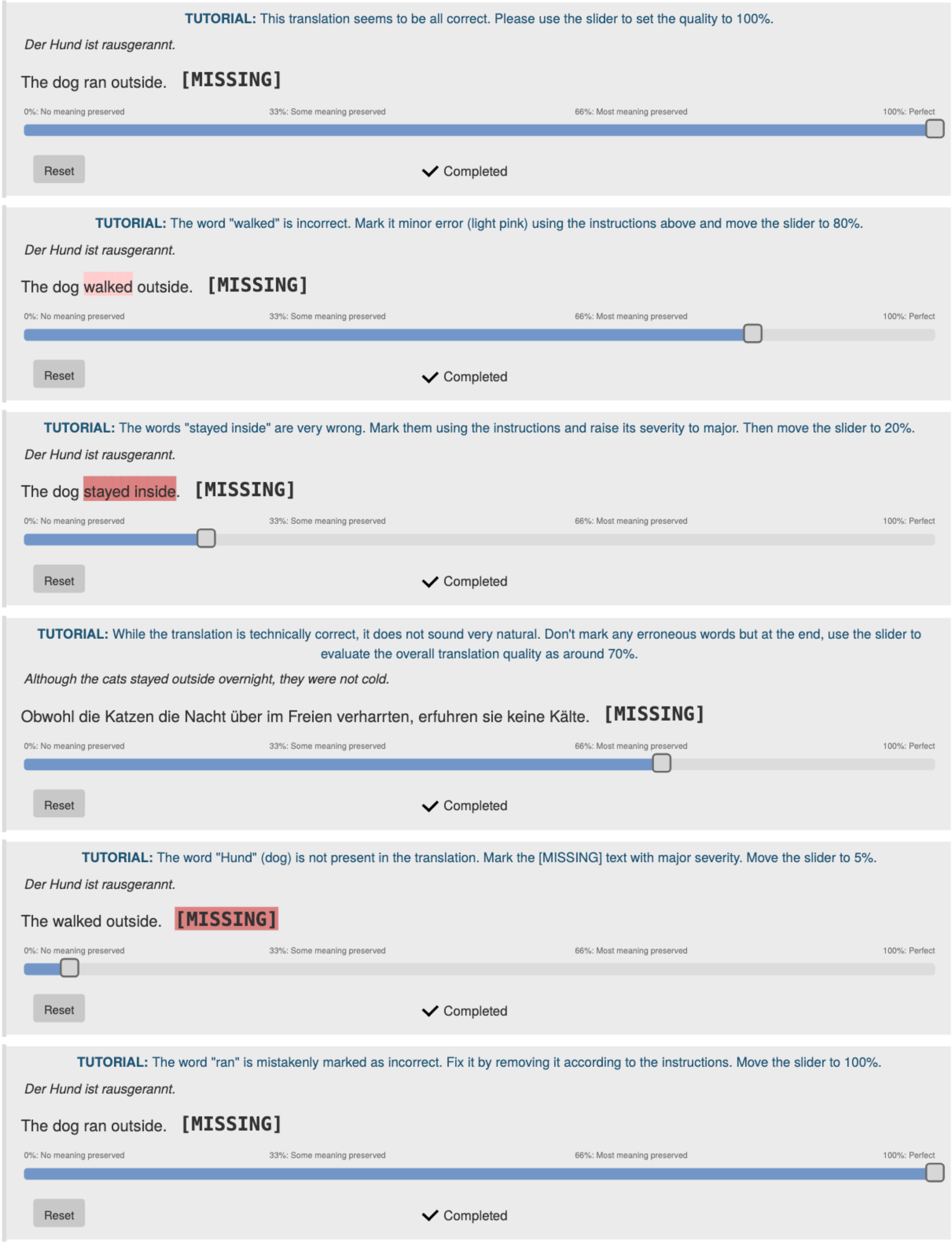
I'm splurging on a new set of frames, these red ones I reeeeally like.

Ich verschwende mein Geld für eine neue Brillenfassung, diese rote mag ich seeehr. **[MISSING]**

So, apparently the ghost image my left eye sees isn't too much of a concern. But I need basically varifocals too. So the lenses have a combo of those. Not the full on varifocals but something in between those and "normal" glasses (no idea of the terminology in English).

Also ist das Geistsrbild, das mein linkes Auge sieht, nicht so sehr Anlass zum Sorge. Aber grundsätzlich brauche ich auch Gleitsichtgläser. Die Linsen sind also eine Modifikation dieser. Nicht die starken Gleitsichtgläser, sondern etwa zwischen diesen und „normalen" Gläsern (keine Ahnung der Terminologie in Englisch). **[MISSING]**

Figure 1: Sample annotation procedure using the **Appraise Evaluation Framework.**



TUTORIAL: This translation seems to be all correct. Please use the slider to set the quality to 100%.

Der Hund ist rausgerannt.

The dog ran outside. [MISSING]

0%: No meaning preserved
33%: Some meaning preserved
66%: Most meaning preserved
100%: Perfect
Reset
✓ Completed

TUTORIAL: The word "walked" is incorrect. Mark it minor error (light pink) using the instructions above and move the slider to 80%.

Der Hund ist rausgerannt.

The dog walked outside. [MISSING]

0%: No meaning preserved
33%: Some meaning preserved
66%: Most meaning preserved
100%: Perfect
Reset
✓ Completed

TUTORIAL: The words "stayed inside" are very wrong. Mark them using the instructions and raise its severity to major. Then move the slider to 20%.

Der Hund ist rausgerannt.

The dog stayed inside. [MISSING]

TUTORIAL: While the translation is technically correct, it does not sound very natural. Don't mark any erroneous words but at the end, use the slider to evaluate the overall translation quality as around 70%.

Although the cats stayed outside overnight, they were not cold.

Obwohl die Katzen die Nacht über im Freien verharrten, erfuhren sie keine Kälte. [MISSING]

TUTORIAL: The word "Hund" (dog) is not present in the translation. Mark the [MISSING] text with major severity. Move the slider to 5%.

Der Hund ist rausgerannt.

The walked outside. [MISSING]

TUTORIAL: The word "ran" is mistakenly marked as incorrect. Fix it by removing it according to the instructions. Move the slider to 100%.

Der Hund ist rausgerannt.

The dog ran outside. [MISSING]

Figure 2: Annotation tutorial using the **Appraise Evaluation Framework.**