MT Quality Assessment Guidelines

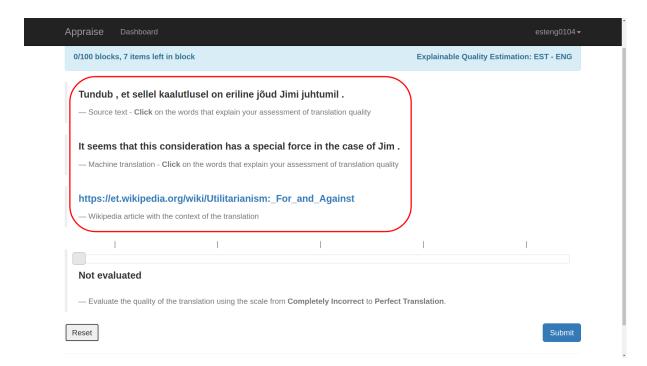
The goal of this task is to evaluate the quality of Machine Translation (MT). Given a source sentence and the MT, the annotators must **rate the overall quality** of the translated sentence and **highlight the words that justify their rating**. Annotation must be conducted **independently** by each annotator.

The annotation will be conducted using the Appraise annotation environment. Each annotator will be given an individual unique link to access their annotation task.

You will evaluate 1,000 sentences randomly extracted from Wikipedia and translated into English by a state-of-the-art neural MT system. For each sentence we provide the link to the corresponding Wikipedia article. This link must be consulted in case it is not clear what the original sentence is talking about.

For each pair of sentences you will have the following information:

- Original sentence
- Machine translation
- Translation context: a link to the Wikipedia article



Sentence-level evaluation

First, you need to rate the quality of the MT on a scale from low (left) to high (right). The following hints for the interpretation of the ranges of the quality scale will appear as you move the slider.

Completely incorrect. MT does not convey any part of the original sentence and is impossible to understand.

A few correct words, but the meaning is different or lost. MT contains a few correct keywords, but is impossible to understand or is very different from the original sentence.

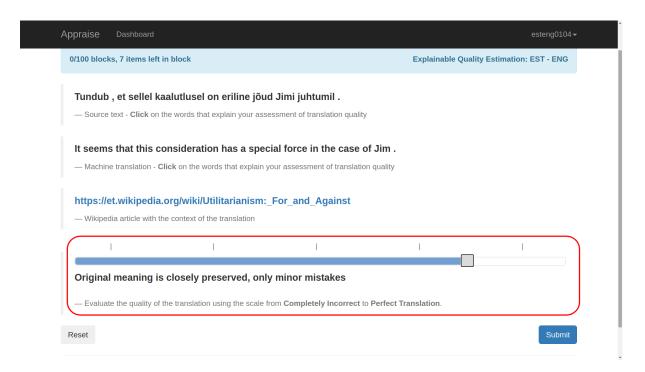
Only parts of the original meaning are preserved. MT conveys parts of the original sentence but it is difficult to recover the overall meaning due to major translation errors.

Translation is understandable, but contains a few mistakes. MT is understandable and conveys the overall meaning of the original sentence but contains a few translation errors.

Very good translation, only minor mistakes. MT closely preserves the meaning of the source but contains a few minor mistakes.

Near perfect or perfect translation

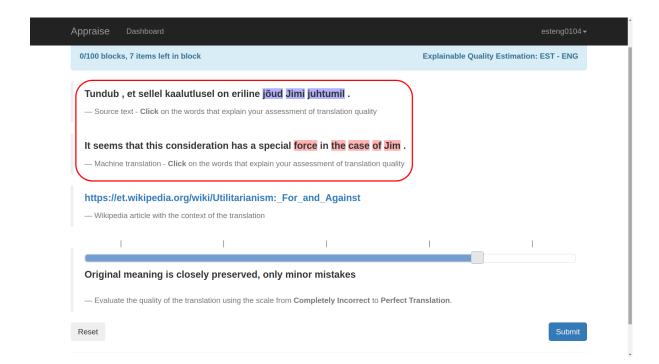
For the analysis purposes, this scale will be interpreted numerically as a 1-100 scale, but during the annotation process the specific numeric value corresponding to the position of the slider is irrelevant.



Word-level explanations

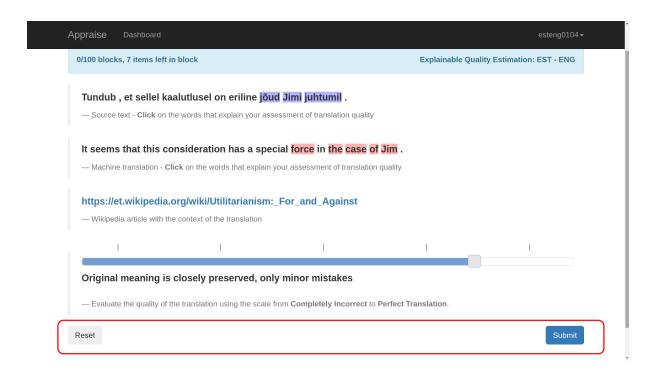
After rating the translation using the scale, you need to **click** on the words in the source and target sentences that justify your rating. Please, follow the rules specified below:

- If the MT output is considered perfect, no words need to be highlighted.
- Otherwise, you must highlight the words corresponding to the errors in translation that made you assign an imperfect score. The corresponding source words that caused the errors must also be highlighted.
- All and only the words necessary to justify the sentence score must be highlighted.
- If some content is missing from the MT output, highlight the source words that were not translated.
- If the MT output has some additional contents, highlight it only in the MT output.
- Original sentences and translations have been tokenized, which means there is a space before each punctuation mark. Please, do not consider these spaces as errors.



Saving your work

While you are working on the example you can click on the "reset" button if you need to correct your annotation. After you finish working on the annotation item, click "submit". Your work will be saved and you will access the next annotation item. Note that after clicking "submit" the annotation cannot be revised.



Examples

In the examples below, "Source" corresponds to the original sentence, "MT" is the Machine Translation (MT) output, "Context" is a link to the Wikipedia article from which the original sentence was extracted and "Score" is the sentence rating. Also, for informative purposes we provide the post-edited (PE) version of the MT output. This will not be available during annotation.

(1) Source: Algse päritoluga ukseava küljepõskedes on näha hästisäilinud riivpalgi avad MT: On the doorstep of the original origin, rice balls have evaporated.

PE: Well-preserved bolt beam holes can be seen on the sides of the original door opening .

Context: Eesti keskaegsed kivilinnused – Vikipeedia

Score: 7

In the case of a completely nonsensical translation, all the translated words must be highlighted. There is no need to highlight anything in the source sentence.

(2) Source: Peale selle nõuab Nicolaus üleüldist maarahu ja vaenuse lõppu.
MT: In addition, Nicolaus is calling for the end of the global rural population and hostility.

PE: In addition, Nicolaus calls for general land peace and end of hostilities.

Context: Nicolaus Cusanus - Vikipeedia

Score: 17

In this example, a few keywords are translated correctly, however the original meaning is completely lost due to a mistranslation error ("global rural population") and wrong word order ("end of peace and hostilities" instead of "peace and end of hostilities") arising from a syntactic ambiguity in the source sentence.

(3) Source: Pronksiajal võeti kasutusele pronksist tööriistad , ent käepidemed valmistati ikka puidust

MT: Bronking tools were introduced during the long term , but handholds were still made up of wood .

 $\it PE$: Bronze tools were introduced during the Bronze Age , but handles were still made of wood .

Context: Minose kultuur – Vikipeedia

Score: 48

In this example, translation conveys parts of the original sentence but the overall meaning is very difficult to recover due to various major errors (e.g. "Bronze Age" is translated as "long term"). All of the words that justify the low sentence score must be highlighted.

(4) Source: Ootamatult saabub koju ka Ursula , kes seekord jääb üpris viisakaks .

MT: Ursula , too , is coming home in anticipation of being quite polite this time .

PE: Unexpectedly , Ursula comes home as well , but this time she remains quite polite .

Context: "Onne 13" jaqude loend – Vikipeedia

Score: 60

In this example, the MT output conveys the overall meaning of the source sentence, but contains a major error ("in anticipation of") that makes it sound very strange in English.

It is not obvious which word in the source sentence corresponds to this error. In such cases, make your best guess by aligning source and target words. In this example, it appears that "in anticipation of" stands in place of the pronoun "kes" and we highlight this word accordingly.

Also, note that we highlight the word "ootamatult" ("unexpectedly") in the source sentence, as it is missing from the translation.

(5) Source: Tundub, et sellel kaalutlusel on eriline jõud Jimi juhtumil.
MT: It seems that this consideration has a special force in the case of Jim.
PE: It seems that this consideration has a special weight in Jim 's case.
Context: Utilitarianism: For and Against – Vikipeedia

Score: 89

In this example, MT output closely preserves the meaning of the source, but is not rendered perfectly in English. We need to highlight minor errors in this case, as they are the ones that justify a high, but imperfect score.

(6) Source: Mõnel juhul võib see printsiip tunduda liiga nõudlik . MT: In some cases , this principle may seem too demanding PE: In some cases , this principle may seem too demanding

Context: Practical Ethics - Vikipeedia

Score: 96

If the translation is considered perfect or near perfect, no words need to be highlighted.