

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

The screenshot shows the Appraise interface for MT quality assessment. At the top, there's a navigation bar with 'Appraise' and 'Dashboard' links, and a user identifier 'roneng0204'. Below this, a status bar indicates '0/100 blocks, 9 items left in block' and 'Explainable Quality Estimation: RON - ENG'. The main content area displays three items: the source text 'În zona Mediteraneană erau cultivate citricele și măslinii .', the machine translation 'Citric and olive trees were grown in the Mediterranean region .', and the Wikipedia context link 'https://ro.wikipedia.org/wiki/Evul_Mediu'. These items are enclosed in a red rounded rectangle. Below this, there's a horizontal slider for quality evaluation, currently set to 'Not evaluated'. A 'Reset' button is on the left and a 'Submit' button is on the right of the slider.

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.

- 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**.

Appraise
Dashboard
roneng0204

0/100 blocks, 9 items left in block
Explainable Quality Estimation: RON - ENG

În zona Mediteraneană erau cultivate **citricele** și măslinii .

— Source text - Click on the words that explain your assessment of translation quality

Citric and olive trees were grown in the Mediterranean region .

— Machine translation - Click on the words that explain your assessment of translation quality

https://ro.wikipedia.org/wiki/Evul_Mediu

— Wikipedia article with the context of the translation

Original meaning is closely preserved, only minor mistakes

— Evaluate the quality of the translation using the scale from Completely Incorrect to Perfect Translation.

Reset Submit

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.**

Appraise
Dashboard
roneng0204

0/100 blocks, 9 items left in block
Explainable Quality Estimation: RON - ENG

În zona Mediteraneană erau cultivate citricele și măslinii .

— Source text - Click on the words that explain your assessment of translation quality

Citric and olive trees were grown in the Mediterranean region .

— Machine translation - Click on the words that explain your assessment of translation quality

https://ro.wikipedia.org/wiki/Evul_Mediu

— Wikipedia article with the context of the translation

Original meaning is closely preserved, only minor mistakes

— Evaluate the quality of the translation using the scale from Completely Incorrect to Perfect Translation.

Reset
Submit

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*: 40 % dintre ei iau cursuri din motive profesionale .
MT: 40 % of them attend courses for professional reasons .
PE: 40 % of them attend courses for professional reasons .
Context: [Finlanda - Wikipedia](#)
Score: 99

In this example the MT is correct, so there is no need to highlight any words in the original or translated sentences.

- (2) *Source*: În zona Mediteraneană erau cultivate citricele și măslinii .
MT: Citric and olive trees were grown in the Mediterranean region .
PE: Citrus and olive trees were grown in the Mediterranean region .
Context: [Evul Mediu - Wikipedia](#)
Score: 82

This sentence contains a minor mistake. The word “citricele” is translated as “citric”, which is acceptable but not a perfectly correct way of saying it in English. We highlight the corresponding words in the original and translated sentences.

- (3) *Source*: Compoziția intitulată Înainte de 1907 , astăzi dispărută , a fost expusă în decembrie 1907 la Iași și mai apoi la București .
MT: The composition before 1907 , today disappeared , was exposed in December 1907 to Istees and later to Bucharest .
PE: The composition named Before 1907 , today disappeared , was exposed in December 1907 in Iași and later in Bucharest .
Context: [Octav Băncilă - Wikipedia](#)
Score: 57

In this example, the MT output is understandable but contains important errors: the name of the city “Iași” is mistranslated and a content word (“intitulată”) is missing in the beginning of the sentence.

We highlight the name of the city in the original sentence and its mistranslation in the MT. We also highlight the word “intitulată” in the source as it is missing in the translation.

- (4) *Source*: În fruntea societății era aristocrația formată din clerici și laici , alcătuind categoriile privilegiate .
MT: At the head of the company was aristocracy of clerics and lakes , making up privileged categories .
PE: At the head of the society was the aristocracy made of clergy and laity , making up the privileged categories .
Context: [Evul Mediu - Wikipedia](#)
Score: 40

In this example, the MT output conveys parts of the original sentence correctly but the overall meaning is difficult to recover due to major translation errors: “societății” is translated as “company” and “laici” is translated as “lakes”. We highlight these errors in the MT output and the corresponding words in the original in order to explain the quality score.

- (5) *Source*: Există un zvon căruia Jolly este zănișoara lui Roxy .
MT: There was a zvon to which Jolly would be endowed by Roxy .
PE: There was a rumor according to which Jolly would be the fairy of Roxy .
Context: [Winx Club - Wikipedia](#)
Score: 25

In this example, the MT output was assigned a very low score. A few keywords are translated correctly, e.g. “Jolly” and “Roxy”. But the word “zânișoara” is mistranslated and the word “zvon” is left untranslated, which makes the MT impossible to understand.

Note that it is difficult to assess the quality of this example without accessing the context provided by the Wikipedia article. We recommend checking the article from which the sentence was extracted whenever it is not clear what the original sentence is talking about.

(6) *Source*: 1926 — Profesor de drept la Școala Superioară de Comerț din Berlin .

MT: B34 : B34 : B34 : B34 : B34 : B34 : B34 : B34 : B34 :

PE: 1926 - Professor of Law at the Berlin School of Commerce .

Context: [Carl Schmitt](#)

Score: 1

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.