

Annotation Guidelines for MT Hallucinations

In this document, you will find guidelines to annotate translations on whether they represent severely incorrect, highly pathological translations.

Overall, you have to decide whether a translation is correct or not, and specify the type of mistake if it is not. See the details below.

1. Correctness

Indicate if the translation is correct or not.

- **Correct translation:** translation fully conveys the meaning of the source. It can contain minor mistakes that do not impact usability or understandability of the content (e.g.: wrong punctuation transfer)
- **Incorrect translation:** conveys meaning different from the source: e.g., parts of the translation are unrelated to the source, some content is omitted.

You should only proceed to Step 2 if the translation is incorrect.

2. Error type

If the translation is incorrect, answer a series of Yes/No questions that are respective to specific characteristics in the translations.

For any (source, translation) pair, you will be asked whether:

1. a translation erroneously contains **repetitions of words and phrases** (e.g.: single words or spans of words, symbols such as punctuation, or other non-words;

Source Sentence	Translation	Explanation
"Ihr Nachname wurde in der Episode nicht genannt, er war aber auf der Ehrenurkunde zu sehen, die Captain Picard von den Kindern geschenkt bekam. Laut Drehbuch ist Marissa neuneinhalb Jahre alt."	"Marissa was born in the United States , and was born in the United Kingdom , where he was born in the United Kingdom , where he was born in the United Kingdom."	The translation contains the repetition of the span of words "was born in the United Kingdom".
"Später easter einer der Kandidaten für den Posten des Kommandanten der neuen Enterprise (NX-01) und ist zugleich der Wunschkandidat des vulkanischen Botschafters Soval . Allerdings wählt Admiral Maxwell Forrest den erfahrenen Jonathan Archer für diese Position aus , um dadurch klarzustellen , dass das vulkanische Oberkommando nicht die Autorisation besitzt , sich in die Personalangelegenheiten der Sternenflotte einzumischen ."	"(TNG - 01) ===== =====	The translation contains the repetition of the symbol "=".
Aller guten Dinge sind Captain Picard bewegt sich schnell auf der Zeitachse , so	In 2372 , the USS Enterprise - D - D was assigned to the USS Enterprise - D - D - D . The USS Enterprise - D - D - D - D - D - D	The translation contains the repetition of the

[illegible]

Source Sentence	Translation	Explanation
"Unsere ausländische OMMELIFT Händler nehmen nach Rücksprache mit OMMELIFT Beratung und Service wahr."	"Our foreign OMMELunless agreed with OMMELB consultants and service providers."	The named-entity "OMMELIFT" is mistranslated.
"Eine Veranstaltung von Tesla in Zusammenarbeit mit dem Förderverein Klosterruine e.v."	"An event of Tesla in cooperation with the Association of Procurement Monsters."	The named-entity "Förderverein Klosterruine e.v" is mistranslated.
"So finden Sie uns in Saarbrücken und Mülheim."	"So you will find us in Saarbrücken and waste house."	The named-entity "Mülheim" is mistranslated.
"Private Parkplätze stehen für EUR 15,00 pro Tag in der Nähe zur Verfügung."	"Private parking is possible at a location nearby and costs EUR 11.00 per day."	The value "15,00" is mistranslated to "11.00"..

3. the translation is an instance of **omission**: (I) content of the source content is untranslated (there is no evidence of effort to translate parts of the source sentence), and (II) the content of the source sentence that was translated was done so correctly

Source Sentence	Translation	Explanation
"Der weltweit renommierte Ingenieur Jürg Conzett hat in Vals eine Brücke gebaut , die im allerbesten Sinne zeitlos ist - verbindet sie doch die schon Jahrtausende alte Technik des Steinbrückenbaus auf sehr subtile Weise mit der Stahlbeton- und Spannbeton - Moderne ."	"In Vals , the world - famous engineer"	There is no evidence of effort to translate a part of the content of the source sentence. The content from the source sentence that is translated is correctly translated. Thus, it is an instance of omission..
"Unter anderem werden die Mitgliedstaaten ersucht,"	"Among other things,"	There is no evidence of effort to translate a part of the content of the source sentence. The content from the source sentence that is translated is correctly translated. Thus, it is an instance of omission.

4. The translation is **strongly detached** from the source sentence content. In these incorrect translations, there is some content that is supported by the source sentence. However, there is a significant part of the translation that is not (it conveys different meaning to the meaning in the source sentence). Importantly, **the content that is hallucinatory does not reflect a genuine effort to translate any part of the source sentence**. If all tokens cannot be supported by the source, you should deem it as fully detached (next category).

Source Sentence	Translation	Explanation
"Bemüht, alles richtig zu machen scheitert er jedoch bei dem Spagat zwischen seinen Aufgaben als Baumarkt-Azubi, Schwiegersohn, Freund und Lover."	"However, he fails in the spaghetti between his tasks as a tree market trainee, a singer's son, a friend and a lover."	The translation content is strongly detached from the source sentence. Not only is a clause missing in the translation, there is a significant proportion of the translated content that conveys different meaning to the source sentence ("spaghetti, tree market, singer's son"). The meaning of the source sentence cannot be guessed from the translation alone.

[illegible]

“Durchgeknallter als alle deutschen Elektroniker zusammen, abgespaceter als das Weltall. Und sinister bis zum letzten Schluß.”

“As the German electronic team, it is more isolated than the world's space.”

The translation content is **strongly detached** from the source sentence, as there is a significant proportion of the translated content that conveys different meaning to the source sentence.



5. The translation is **fully** detached from the source sentence content (the translation is **fully** separated from the source sentence content). Possible examples include translations that convey totally different meanings to that of the source sentence. In these translations, there is no evidence in the translation of translated source content.

Source Sentence	Translation	Explanation
“Lucio nur aufwecken um 1 Uhr nachts.”	'The staff were very friendly and helpful.'	The translation content is fully unsupported by the source sentence.
“Das war beim Frühstück, beim Essen, während der Sitzungen nicht nur im außenpolitischen Ausschuß die zentrale Frage, wo in Freund oder Feind, um es übertrieben zu sagen, eingeteilt wurde.”	'The room was very small and the bed was very comfortable and the bed was very comfortable.'	Not only is there evidence of repetitions (“bed was very comfortable”, the translation content is fully unsupported by the source sentence.

Rubric for Annotation



To summarize, you will be given a (source, translation) pair and you will need to fill the following rubric below:

Is this a correct translation for the given source sentence (see Sec. 1)?



- ☐ Yes 
☐ No 

If you answered No in 1., answer the questions below:



Does the translation contain erroneous repetition of words/spans of words/punctuation (see 2.1)?

- ☐ Yes 
☐ No 



Does the translation contain mistranslation of named entities (person names, organizations, locations, etc.) – (see 2.2)?

- ☐ Yes 
☐ No 


Is the translation an instance of omission (see 2.3)?

- ☐ Yes 
☐ No 

Is the translation content **strongly** unsupported by the source (see 2.4)?

- ☐ Yes 
☐ No 

Is the translation content **completely** unsupported by the source (see 2.4)?

- ☐ Yes 
☐ No 