

MTM 2014 LAB - ERROR ANNOTATION GUIDELINES

Task Description

In this task you are presented with an automatically translated sentence (Hypothesis) and its post-edition carried out by a professional translator (Reference). For each automatic translation, you are asked to compare it with its post-edition in order to identify the type of errors present and mark their position in the text.

The error typology is composed of five main classes:

- 1) Reordering errors
- 2) Missing words
- 3) Extra words
- 4) Lexical errors
- 5) Morphology errors

1) Reordering errors

A reordering error occurs when one or more words in the translation are in a wrong position. Reordering errors can affect single words or many words, and the reordering range can be local (e.g. three or less words around the error) or long (e.g. more than three words around the error).

When it is possible to identify individual words or blocks of words to be moved independently from the others, only the word(s) in the wrong position must be marked as errors.

REF: to target the desire centers of your brain

HYP: to target the desire of your brain *centers*

On the contrary, in those cases where the order of some words depends on other words and it is not possible to identify a precise word movement, the whole text span involved in the reordering error is to be marked.

REF: I want to talk about one of the biggest modern day choosing problems ...

HYP: I want to talk about one of the *day modern biggest problems choosing...*

2) Missing words

A missing word error occurs when a word appears in the reference, but not in the hypothesis.

REF: to target the desire *centers* of your brain

HYP: to target the desire of your brain

3) Superfluous words

A superfluous (extra) word error occurs when a word appears in the hypothesis, but not in the reference.

REF: to target the desire centers of your brain

HYP: to target the desire *hearth* centers of your brain

4) Lexical errors

Lexical errors include different phenomena, such as incorrect choice of words (including not recognized idioms), misspelled words, foreign source words passed through (including proper names).

REF: to target the desire centers of your brain

HYP: to target the desire *squares* of your brain

5) Morphology errors

A morphology error occurs when the base form of the translated word is correct but the full (surface) form is wrong due to errors with conjugation, tense, modality, declension, concordance in number, gender, case.

Example:

REF: If someone asked you to describe

HYP: If someone *asks* you to *described*

Annotation instructions

You will be presented with one of the following annotation settings:

1. annotation from scratch
2. revision of existing annotations. In this setting your task is to keep the tags you agree with, removing those you don't agree with, and adding new tags until the result is correct.

In both annotation settings, remember that:

- In the annotation interface, OUTPUT 1 corresponds to the reference text (post-edition), while OUTPUT 2 corresponds to the hypothesis text (MT output).
- All errors must be annotated in the hypothesis text (OUTPUT 2). Only in the case of missing words, the word(s) that is(are) missing in the hypothesis must be tagged in the reference text (OUTPUT 1).
- The reference sentence is to be considered correct: possible errors in the reference must not be annotated. The reference must be tagged only to mark words that are missing in the hypothesis.
- You can annotate both single words and text spans.
- You can annotate multiple errors in the same word/span.
- If the hypothesis requires too much annotation effort, you can mark it as containing "too many errors" (please try not do it! :-)