

Evaluation Guide Provided to Human Annotators

Evaluation Metrics

1. Idiom Translation: Does the translation of the idiom preserve its original meaning within the sentence context?
 - 0: Yes
 - 1: No

2. Fluency: Assuming the idiom is correctly translated, how semantically and syntactically accurate is the rest of the translation?
 - 5: The translation has no issues in terms of grammar or meaning.
 - 4: The translation has one issue (or two minor issues) in grammar or meaning.
 - 3: The translation has more than two grammatical issues, but the meaning is fairly understandable.
 - 2: The translation has numerous grammatical issues, or part of the sentence's meaning is missing or incomprehensible.
 - 1: The translation is completely incomprehensible or entirely different from the original sentence.

Evaluation Steps

To evaluate idiom translation:

1. Using the original sentence and the idiom, identify the intended idiom in the English sentence.
2. Without looking at the human translation, answer this question: What does the idiom mean?
3. Look at the human translation. Does the meaning you identified match the human translation?
 - Yes: idiom-translation = 1
 - No: idiom-translation = 0

To evaluate fluency:

Assume that the idiom has been translated correctly in terms of meaning. Score the rest of the sentence based on the criteria explained in the previous section.

Note: If you believe the machine translation conveys the intended meaning of the idiom but contains errors, consider those errors when evaluating fluency.

Examples

1. The idiom and the rest of the sentence are correctly translated.
 - Original Sentence: She passed the test with flying colors.
 - Gold Translation: عملکرد او در تست عالی بود. (She performed exceptionally well on the test.)
 - Machine Translation: عملکرد او در تست خیلی خوب بود. (Her performance on the test was exceptional.)

Scores: Idiom translation → 1, Fluency → 5

2. Idiom translation preserves meaning but is grammatically incorrect.
 - Original Sentence: She passed the test with flying colors.
 - Gold Translation: عملکرد او در تست عالی بود. (She performed exceptionally well on the test.)
 - Machine Translation: عملکرد او در تست خوب خیلی بود. (She performed well very on the test.)

Scores: Idiom translation → 1 (Even though the idiom translation has grammatical mistakes, it still conveys its original meaning.), Fluency → 4 (Grammatical errors in idiom translation impact fluency.)

3. Idiom translation does not preserve its original meaning, but the rest of the sentence is correctly translated.
 - Original Sentence: She passed the test with flying colors.
 - Gold Translation: عملکرد او در تست عالی بود. (She performed exceptionally well on the test.)
 - Machine Translation: عملکرد او در تست خنده دار بود. (She performed funnily on the test.)

Scores: Idiom translation: 0, Fluency: 5 (Fluency score is independent of the semantic correctness of the idiom translation.)

4. The idiom is correctly translated but has incorrect additions.
 - Original Sentence: She passed the test with flying colors.
 - Gold Translation: عملکرد او در تست عالی بود. (She performed exceptionally well on the test.)
 - Machine Translation: عملکرد او در تست عالی و خنده دار بود. (Her performance was exceptional and very funny on the test.)

Scores: Idiom translation: 0 (This translation does not convey the correct meaning of the idiom.), Fluency: 5