Criteria

1. Is the correction accurate?

When evaluating correctness, only target error is considered, other errors are not evaluated. The correction may not be identical to the one written in standard but still be tagged as a correct one.

'corrected' ("исправлена") — the target error is corrected the same way as in standard
'corrected but not as in standard' ("исправлена, но не как в эталоне") — the target error is fixed correctly but it is different from the expected correction
'is not corrected or corrected incorrectly' ("не исправлена или исправлена некорректно") — the target error is still present or is was corrected incorrectly (wrong word choice, making new error, etc.)
'is not present in the correction' ("отсутствует в исправлении") — after the correction the piece that contains the target error was removed; if removing the word was necessary for the correction (e.g. removing a preposition that is not needed) then the correction is marked as 'corrected'

2. Change of structure (stage 1)

During the first stage (evaluating 8 models) it was briefly rated whether the sentence has changed much after the correction. Changes in syntax, words' and phrases' removal, and addition of them were considered as major changes. Some changes were meaningful and helped with the error correction, and some changed the meaning of the sentence.

'has not changed' ("не изменилась") — there were no significant changes
'has changed but it was needed for the correction' ("изменилась, но этого требовало исправление") — there were some changes but they help to correct the error (there may be also a variation of correction that has no structure changes)
'has changed and it was not needed for the correction' ("изменилась, и для исправления этого не требовалось") — some changes are extra and do not affect the error correction

3. Change of structure (stage 2)

During the second stage (evaluating 3 models) structure changes were viewed more detailed. Multiple choices were available, which means it was possible to mark all the types of changes separately. Only the changes beyond the standard correction of the target error were evaluated.

'has not changed' ("не изменилась") — nothing apart from words related to the target error has changed, syntax has not changed either

'words have disappeared' ("исчезли слова") — some words that are not related to the target error are missing (these words must be independent part of speech)

'words have appeared' ("появились слова") — new words that were not present in the original sentence have appeared (these words must be independent part of speech)

'words have changed' ("слова поменялись") — some words were replaced with synonyms without visible reason (these words are not related to the target error)

'syntax has changed' ("поменялся синтаксис") — there were some changes in constructions, syntactic roles, word order, etc.

4. Evaluation of changes (stage 2)

Apart from indicating specific types of changes assessors also gave their own evaluation to the quality of the correction with a change in structure: whether the changes increased the quality of the text or they were unnecessary. To avoid subjectivity, the assessors' assessment was tested on a small sample using responses from independent experts.

'the quality has increased' ("качество текста улучшилось") — the changes has made text more coherent, logical or understandable, without misrepresenting author's idea

'the quality stayed the same' ("качество осталось таким же") — the changes had no significant effect on text's quality

'the quality has decreased' ("качество текста ухудшилось") — the changes were unnecessary, have misrepresented author's idea or made it difficult to understand

'no structure changes' ("изменений структуры не было") — in case it was marked with 'has not changed' previously

5. Pairwise comparison

In order to compare several corrections of the same error with each other, a pairwise comparison was made. First of all, the target error correction was evaluated and if the corrections were the same, assessors evaluated other changes and quality of the final sentence.

'model 1 is much better than model 2' ("модель 1 сильно лучше модели 2") — model 1 has fixed the target error much better

'model 1 is slightly better than model 2' ("модель 1 немного лучше модели 2") — model 1 has fixed the target error slightly better or the same, in general, the preference is given to the model 1's variation

'no difference' ("одинаково") — the sentence is corrected identical by both models or it is impossible to choose between them (e.g. both corrections are incorrect)