

Критерии

1. Корректно ли исправление

При оценке корректности рассматривается только целевая ошибка, другие ошибки в данном предложении не оцениваются. Исправление может не совпадать с эталоном, но являться корректным

“исправлена” — целевая ошибка исправлена в соответствии с эталонным исправлением

“исправлена, но не как в эталоне” — целевая ошибка исправлена корректно, но отличается от ожидаемого исправления

“не исправлена или исправлена некорректно” — ошибка осталась на месте или была исправлена некорректно (неправильный выбор слова, появление новой ошибки и др.)
“отсутствует в исправлении” — в результате исправления фрагмент с целевой ошибкой был убран; если убиение слова было не дополнительным изменением структуры, а одним из вариантов корректного исправления (например, убрать предлог, не сочетающийся с существительным), это помечается как “исправлена”

2. Изменение структуры (этап 1)

На первом этапе (оценка 8 моделей) кратко оценивалось, сильно ли изменилось предложение в процессе исправления. Сильными изменениями структуры считались изменения в синтаксисе, исчезновение слов и фраз или добавление новых.

Некоторые изменения были осмысленными и помогали исправить ошибку, а некоторые изменяли смысл, заложенный автором предложения

“не изменилась” — значимых исправлений не было

“изменилась, но этого требовало исправление” — есть изменения, но все они помогают в исправлении ошибки (при этом может существовать и вариант исправления без изменения структуры)

“изменилась, и для исправления этого не требовалось” — некоторые изменения излишни и не влияют на исправление ошибки

3. Изменение структуры (этап 2)

На втором этапе (оценка 3 моделей) изменения структуры были рассмотрены более подробно. Был доступен множественный выбор, т.е. можно было отметить все

встретившиеся типы изменений отдельно. Всё ещё оценивались только изменения, выходящие за рамки стандартного исправления целевой ошибки.

“не изменилась” — ничего кроме отдельных слов, связанных с целевой ошибкой, не поменялось, синтаксис не изменился.

“исчезли слова” — в исправлении не хватает слов, не связанных с целевой ошибкой, это самостоятельные части речи (исчезновение предлогов не засчитывается)

“появились слова” — в исправлении появились самостоятельные слова, которых не было в оригинале (ср. “слова поменялись”)

“слова поменялись” — некоторые слова были заменены синонимами без видимой причины (не связанные с целевой ошибкой)

“поменялся синтаксис” — поменялись используемые конструкции, синтаксические роли, слова в предложении поменялись местами и пр.

4. Оценка изменений (этап 2)

Помимо указания на конкретные типы изменений асессоры давали собственную оценку качеству исправления с изменением структуры: повысили ли изменения структуры качество текста или были излишни. Чтобы избежать субъективности, оценка асессоров была проверена на небольшой выборке с помощью ответов независимых экспертов.

“качество текста улучшилось” — изменения сделали текст более связным, логичным или понятным, не исказив задумку автора

“качество осталось таким же” — изменения не оказали значимого влияния на качество текста

“качество текста ухудшилось” — изменения были излишними, исказили изначальный смысл или затруднили понимание

“изменений структуры не было” — в случае, если в пункте 3 не было отмечено изменений структуры

5. Попарное сравнение

Для того, чтобы сравнить несколько исправлений одной и той же ошибки между собой, было произведено попарное сравнение. Прежде всего оценивалось исправление целевой ошибки и лишь в случае, когда целевые исправления совпадали, асессоры прибегали к оценке остальных изменений и качества итогового предложения.

“модель 1 сильно лучше модели 2” — целевая ошибка исправлена моделью 1
значительно лучше

“модель 1 немного лучше модели 2” — целевая ошибка исправлена моделью 1
незначительно лучше или так же, в целом предпочтение отдаётся варианту от модели
1

“одинаково” — предложения исправлены идентично, либо невозможно выбрать
(например, оба исправления некорректны)