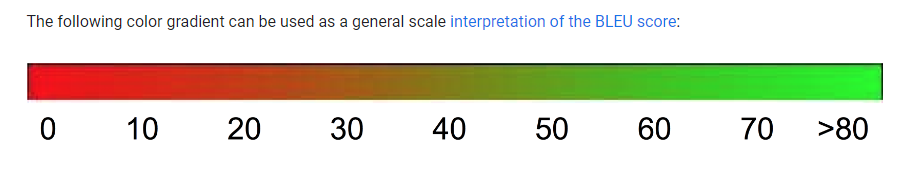


BLEU (BiLingual Evaluation Understudy) – метрика от 0 до 1, оценивающая, насколько машинный перевод близок к высококачественному (например, ручному) переводу. 0 означает, что нет пересечений между кандидатом (версией, полученной с помощью машинного перевода) и референсом (текстом-образцом), 1 означает, что все идеально совпадает, но на самом деле единицу даже людям сложно получить.

В целом BLEU предназначен для оценки качества работы модели, поэтому лучше его считать для целого корпуса, но для одного и того же.

Google Cloud приводит такую интерпретацию результата:



Иными словами, от 40% точности начинаются переводы, которые в целом понятны, хоть и не точны, а от 60 – переводы, которые в среднем будут лучше, чем сделанные вручную.

Что касается примеров переводов предложений, то мне кажется, что «предсказания» (кандидаты) были недооценены.

В первом случае один синоним заменен менее точным и с другим оттенком (conventional/usual) и потерян факт погружения под лед. Тем не менее, это достаточно точный перевод, структура In addition \_\_, there is \_\_\_ осталась на месте. Могу предположить, что проблема со словом подледный заключалась в том, что это достаточно редкое слово, не встречавшееся в тренировочных данных, поэтому при переводе акцент был сделан на семантику корня, и она была передана правильно. ~~А что мы еще хотим.~~

Со вторым предложением дела обстоят куда интереснее: машина передала значение одного образного выражения другим, гораздо более близким по тематике, чем было в референсе. Уместно ли это – зависит от контекста, но в вакууме это точно нельзя назвать плохим переводом.

В третьем предложении синонимы опять оказываются заменены на более «простые» (fairness/justice, scarce/small). Более того, scarce и small в этом контексте вообще не очень синонимичны – small это скорее про размер, а scarce про наличие, а поскольку «малые» может обозначать и то, и то, было выбрано более топорное значение. К тому же, структура предложения оказывается передана менее «творчески» - если в русском мы начинали с «Также,», то вот и в английском начнем с «Also,», никаких переставлений для более естественного звучания.

Итак, можно сделать (ожидаемый) вывод, что наибольшими препятствиями для высокого BLEU score в данном случае оказались редкие слова и фразеологические обороты.