Вера Латышкина

* + Что значит кандидат и референс?

Кандидат – выходные данные машинного перевода (то, что сравнивается по качеству с эталонным переводом).  
Референс – эталонный, точно правильный перевод, на который ориентируется метрика.

* + Для чего подойдет BLEU-Score: подсчета отдельных переводов или результатов работы системы на целом корпусе?

Для целых корпусов (потому что для правдоподобной оценки нужен большой обучающий датасет, чем больше данных, тем адекватнее оценка).

* + Какие "скоры" считаются высокими, а какие - низкими?

Шкала «скоров» – от 0 (макс. низкий показатель) до 1 (макс. высокий), чем ближе число к единице, тем точнее совпадение с референсом.

* Проанализируйте результаты в ячейках ниже. Порассуждайте:
  + На ваш взгляд, какие слова, контексты вызвали трудности у автоматической системы?  
    1. Фразеологизмы  
       Для «# Что посеешь - то и пожнешь» оба варианта адекватные по смыслу и естественные для живого языка (в русском тоже можно придумать вариации типа «как аукнется, так и откликнется»), но формально не совпадают вообще;
    2. «underwater and even under-ice» – слово «under(-)ice» конструируется в контексте как своего рода неологизм, машинный вариант опускает приставку, чтобы получилось нормальное словарное слово, система видит «under-ice» и «ice» как разные слова;
    3. Лексическая/контекстная сочетаемость: в случаях «conventional/usual hockey», «variety/version of this game», «scarce/small resources» – более специальный (устойчивый)/более нейтральный вариант сочетаемости – по смыслу более-менее совпадают, но н-граммы разные;
    4. Контекстно синонимичные слова/конструкции: «fairness/justice», «is how/ has to do with how» - разные слова и разные длины равнозначных по смыслу н-граммов (=надо сравнивать 2- и 4-граммы, а система сравнивает только равные друг другу);
    5. Лишние слова, которых нет в референсе – машинный вариант ставит лишние «the» (тоже увеличивает длину н-грамма);
    6. Порядок слов: разные позиции «also» в 3 предложении (попадает в разные н-граммы);
  + Руководствуясь знаниями о том, как работают нейросети, предположите, что вызвало ошибки.  
    1. Последовательная обработка токенов;
    2. Нет выравнивания для идиом.