

# МАШИННЫЙ ПЕРЕВОД СВЯЗНЫХ ТЕКСТОВ: ОЦЕНКА КАЧЕСТВА И ПРОБЛЕМЫ

**Н.В. Лукашевич**

МГУ имени М.В. Ломоносова

[louk\\_nat@mail.ru](mailto:louk_nat@mail.ru)

## Введение

Современный машинный перевод достиг значительного прогресса, в ряде работ указывается, что оценка качества перевода автоматическими системами уже достигла качества профессионального перевода (Hassan et al., 2018). Однако такие высокие оценки касаются только переводов в рамках отдельных предложений. При экспертной оценке качества перевода текста в целом, человеческий перевод получает значительно более высокие оценки (Laubli et al., 2018).

Это связано с тем, что подавляющее число современных систем машинного перевода переводят текст по отдельным предложениям. В результате не учитываются различные механизмы связности текста, и переводы в различных предложениях могут не соответствовать друг другу.

В данной статье мы представим краткие описания недавних экспериментов, в которых качество машинного перевода сравнивалось с человеческим, а также наиболее частые проблемы перевода на уровне текста.

## Эксперименты, оценивающие близость машинного перевода к профессиональному

В работе Hassan et al. 2018 была поставлена экспериментальная задача оценить качество китайско-английского машинного перевода новостных текстов и сравнить с профессиональными переводами этих же данных. Для оценки качества перевода оценщикам, владеющим и китайским, и английским языками, выдавались тексты на исходном и целевом языке. Оценщики не являлись профессиональными переводчиками, а были собраны через систему краудсорсинга (краудсорсеры), при этом обеспечивались проверки компетентности оценщиков в обоих языках.

Оценщиками предъявлялись отдельные предложения, и они должны были ответить на вопрос: «Как точно переводной текст передает смысл исходного текста» по шкале от 0 (совсем не передает) до 100 (полностью передает).

Каждое предложение оценивалось тремя оценщиками. В результате экспериментов авторы сделали вывод, что переводы, сделанные предложенной системой машинного перевода, получили оценки, сравнимые с оценками профессиональных переводов, качество переводов человека и компьютера статистически не отличимо.

В исследовании Laubli et al. 2018 сравнение результатов машинного и автоматического перевода было продолжено: в частности, проводилась оценка качества перевода не только для отдельных предложений, но и для текста в целом.

Оценщикам предлагалось два перевода (профессиональный и автоматический), оценщик должен был выбрать, какой перевод ему нравится больше. В эксперименте оценщики должны были оценить переводы по двум параметрам: *adequacy* (адекватность) и *fluency* (качество текста). При оценке адекватности оценщики видели исходный текст и его переводы и должны были ответить на вопрос: какой перевод передает смысл текста более адекватно. При оценке качества текста оценщики видели только переводы и оценивали, в каком тексте лучше качество английского языка. В качестве оценщиков выступали профессиональные переводчики с опытом по крайней мере трехлетней переводческой деятельности. Оценщики оценивали целые документы и отдельные предложения в случайном порядке. Для оценки использовались данные, опубликованные в вышеупомянутой работе Hassan et al. 2018, что позволяет сопоставлять результаты двух исследований.

В результате оценки были сделаны следующие выводы. Адекватность перевода отдельных предложений профессионального и машинного перевода оказалась сравнимой, как и в предыдущем эксперименте Hassan et al. 2018. Однако при переводе текстов в целом оценщики отдавали явный приоритет человеческому переводу. При оценке качества текста (*fluency*) оценщики ставили более высокие оценки человеческому переводу как на уровне предложения, так и на уровне текста. Эти выводы были подтверждены статистическими тестами.

Сравнительные исследования были продолжены в работе Freitag M. et al. 2021. Были рассмотрены две пары языков: англо-немецкий перевод и китайско-английский перевод. В качестве участников тестирования были взяты лучшие системы машинного перевода с тестирования WMT (World Machine Translation), а также две онлайн-системы перевода. По результатам тестирования было показано, что все-таки утверждения работы Hassan et al. 2018 не подтверждаются: для обеих пар перевода и всех применяемых оценок человеческий перевод получил более высокие оценки. В частности, это было объяснено тем, что в данном исследовании оценщиками служили профессиональные переводчики, а не краудсорсеры, что позволило точнее оценить качество перевода.

### Проблемы машинного перевода в переводе полных текстов

Как известно, неотъемлемой характеристикой связного текста является когезия – связь между языковыми выражениями в разных частях текста, которая соединяет их в единое целое. Когезия может быть подразделена на несколько подвидов: когезия на основе дискурсивных маркеров (союзов, союзных слов), когезия на основе референции (Kibrik 2011), эллипсис, лексическая связность (Halliday, Hasan 1976). Именно последние три явления и представляют наиболее частые проблемы при переходе машинного перевода с уровня предложения на уровень текста.

В работе Voita et al. 2019 авторы собрали специальный корпус и разметили ошибки машинного перевода на уровне текста. Было выявлено, что такие проблемы связаны в основном с тремя явлениями связности текста: дейксисом, эллипсисом и лексической связностью.

**Дейксис** — это референциальные выражения, интерпретация которых зависит от контекста. Наиболее частой проблемой стал перевод личных местоимений, например, в переводе на русский язык может получиться непоследовательный перевод «ты-вы»:

(1) EN: *We haven't really spoken much since your return. Tell me, what's on your mind these days?*

RU: *Мы не разговаривали с тех пор, как вы вернулись. Скажи мне, что у тебя на уме в последнее время?*

**Эллипсис** представляет собой пропуск в предложении одного или нескольких слов, которые могут быть восстановлены из контекста. Проблема с переводом эллиптических конструкций связана, во-первых, с необходимостью воспроизвести пропущенные слова из контекста, если в целевом языке не разрешается эллипсис такого рода. Во-вторых, пропущенная конструкция может влиять на синтаксис остального предложения, например, на окончания существительного в русском языке.

(2) EN: *You call her your friend but have you been to her home? Her work?*

RU: *Ты называешь её своей подругой, но ты был у неё дома? Её работа?*

**Лексическая связность**, в частности, подразумевает, что если для некоторого слова или имени возможны несколько переводов, то нужно переводить одинаково, а при машинном переводе часто возникает непоследовательность с переводом имен. Например, перевод на русский язык может получиться таким:

(3) EN: *Not for Julia. Julia has a taste for taunting her victims.*

RU: *Не для Джулии. Юлия умеет дразнить своих жертв.*

В работе Voita et al. 2019 также указывается, что современные автоматические метрики качества машинного перевода не чувствительны к таким проблемам, поскольку их доля в текстах относительно низкая. Однако, понятно, что появление хотя бы одной такой неточности в тексте сразу снижает оценку качества перевода для человека.

### Заключение

В последние годы качество машинного перевода значительно улучшилось, что позволило ставить вопрос, насколько качество перевода деловой прозы приблизилось к качеству профессионального перевода. В ряде работ были сделаны утверждения о сравнимости качества машинного и профессионального перевода. Если при некоторых условиях такая сравнимость возможна при переводе отдельных предложений, то при переводе целых текстов есть еще много проблем, которые разработчики систем машинного перевода должны научиться преодолевать. В подавляющем числе случаев эти проблемы связаны с таким свойством текстов как когезия.

### Литература

- Freitag M. et al. 2021. Experts, errors, and context: A large-scale study of human evaluation for machine translation. *Transactions of the Association for Computational Linguistics*, v. 9, 1460–1474.
- Halliday M.A.K., & Hasan R. 1976. *Cohesion in English*. English Language Series. London: Longman.
- Hassan H. et al. 2018. *Achieving human parity on automatic chinese to english news translation*. arXiv preprint arXiv:1803.05567.
- Kibrik A.A. 2011. *Reference in discourse*. Oxford: Oxford University Press.
- Läubli S., Sennrich R., Volk M. 2018. Has Machine Translation Achieved Human Parity? A Case for Document-level Evaluation. In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Stroudsburg, PA: Association for Computational Linguistics, 4791–4796.
- Voita E., Sennrich R., Titov I. 2019. When a Good Translation is Wrong in Context: Context-Aware Machine Translation Improves on Deixis, Ellipsis, and Lexical Cohesion. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Proceedings of the Conference*. Stroudsburg, PA: Association for Computational Linguistics, 1198–1212.