



Чем полезны нейросети при обучении студентов в программах автоматизированного перевода

А.В. Ковальчук

В свое время от выпускников переводческого факультета мне доводилось слышать, что дисциплина «Информационные технологии в переводе» у них читалась как теоретический курс, без аудиторной работы за компьютером. И, конечно, никаких кат-программ.

Хочется верить, что подготовка современных переводчиков в академической среде за последние годы стала значительно ближе к реалиям отрасли. В том числе в непрофильных вузах, таких как Сеченовский университет и МАИ. Там студентов обучают не только классическому переводу, но и работе в программах автоматизированного перевода: дома и в аудитории — «Смарткат», на производственной практике — «Традос».

Кроме того, есть еще и машинный перевод, и постредактирование, и общение с нейросетями... Этими технологиями — в отличие от кат-программ, которым надо прицельно обучаться, — студенты активно пользуются сами и по совершенно разным поводам, хотя и не всегда осознают их оборотную сторону. Запретить? Не получится. Значит, надо научить студентов применять технологии по делу и с умом, показать их ограничения: слабых мест у всей этой цифры более чем достаточно.

Будем исходить из того, что коллеги-преподаватели, которым адресована эта статья, работают или планируют работать со студентами в той или иной кат-программе. Здесь необходимый минимум — текст оригинала. На начальном этапе, вероятно, не стоит все усложнять. Познакомиться с оболочкой программы, узнать, что там в принципе есть, какие разделы и функции, разобраться в основных возможностях и принципах работы — все это обилие информации нужно усвоить, отработать и довести до автоматизма. После того, как студенты освоились, можно имитировать работу переводчика в бюро переводов. Думается, у многих коллег есть соответствующий профессиональный опыт, хотя бы кратковременный; следовательно, общее представление о специфике задач тоже имеется. В остальном помогут нейросети, и использовать их можно следующим образом.

Создаем кат-проект... А что у нас есть?

Кроме текста оригинала, скорее всего, ничего. Мало кто располагает доступом к корпоративным ресурсам переводческого бюро или обзавелся собственным архивом переводческой памяти и глоссариев по соответствующим тематикам. Но такой архив можно сделать.

1. Начнем с легенды

OK, DeepSeek, придумай жизненную ситуацию. Мы работаем в бюро переводов, наш постоянный заказчик из фармкомпании разработал новый препарат. От проектного менеджера мы получили четыре файла: оригинал, память переводов, глоссарий и стайлгайд. Перевод делал врач, по медицине все хорошо, но клиент остался недоволен качеством: проблемы стиля, унификации, грамотности. Перевод в файле памяти tmx отдали нам, лингвисту, на доработку.

В ответ на этот промпт нейросеть выдает несколько вариантов вполне реалистичных историй. Выбрав один, на мой взгляд, самый правдоподобный и наименее разухабистый, я внес в него минимум правок и получил коммуникативную ситуацию, типичную для насыщенных событиями будней в переводческом бюро:

Вы работаете медицинским редактором в бюро переводов MedTrans Global. Ваш постоянный клиент — фармацевтическая компания Ganssen.Oncology — готовится к запуску нового препарата Простатин, для лечения локализованного рака предстательной железы.

Менеджер проекта прислал вам пакет документов, а именно: презентацию «ТО ЛРПЖ. rptx», список требований от заказчика, стайлгайд, память переводов и глоссарий. Работа была начата в SmartCat другим переводчиком — внештатным врачом-онкологом. Он уже выполнил перевод сегментов и выгрузил память переводов в формате TMX.

Однако клиент, взглянув на фрагменты перевода, выразил недовольство качеством. Перевод точен по сути (врач, так сказать, «не ошибся в диагнозах»), но стилистически «сырой», содержит кальки и неунифицированную терминологию. Клиент требует редактуры.

Ваша задача как лингвиста — загрузить презентацию в SC, подключить к проекту этот TMX и отредактировать перевод.

Цель редактирования: подготовить финальную, «чистовую» версию перевода, которая будет использоваться в информационном пакете для медицинских представителей Ganssen.Oncology.

По мере необходимости ситуацию можно усложнять, внося в нее детали, обязательные к учету. Для переводчика важно быть внимательным к мелочам, вчитываться и вдумываться в техническое задание, оценивать параметры ситуации, а не делать что-то свое, не переводить текст в пустоту, но соотносить его с обстоятельствами коммуникации. Для студентов, по их же отзывам, это становится каким-то откровением.

Постоянный клиент нашего переводческого бюро — это фармацевтическая компания Ganssen.Oncology. Мы для нее переводим тексты про онкологические препараты, в основном материалы клинических исследований, но также ОХЛП. Давай представим, что в этой фармкомпании есть сотрудник, с которым мы все это время взаимодействуем. От него мы получаем тексты на перевод, ему их сдаем, он оценивает качество перевода, иногда что-то возвращает. За годы совместной работы у него появился список требований, рекомендаций и пожеланий о том, как нужно и как не нужно переводить для Ganssen.Oncology. Придумайте такой перечень требований от медицинского специалиста, контактирующего на постоянной основе с переводческим бюро.

2. Добавим список требований от заказчика

Нейросеть способна генерировать длинные, хорошо структурированные списки подобного рода требований. Разумеется, что любые сведения, предоставленные нейросетью, необходимо проверять! Преподаватель перевода, поскольку сам практикующий переводчик, навыком поиска и проверки информации владеет в достаточной мере, чтобы с минимумом усилий редактировать нейросеть и на выходе получать грамотный достоверный текст. Прочитую самый первый ответ¹, точнее его преамбулу и первый из десяти пунктов:

Памятка для переводческого бюро (от Алексея, Ganssen.Oncology)

Коллеги, мы работаем давно, и я вам доверяю. Но онкология — это зона нулевой терпимости к смысловым искажениям. Пациенты, врачи и регуляторы (FDA, EMA, Минздрав) получают именно то, что написано у вас. Поэтому ниже — не капризы, а жесткие требования, выработанные годами правок. Спасибо, что приняли это как руководство к действию.

1. Терминология — абсолютная стабильность

- Требование: никакой синонимии ради красоты слога. Один английский термин → один русский эквивалент в пределах одного документа и всей линейки препаратов.
- Пример строгого соответствия:
 - *Adverse event (AE)* — только «нежелательное явление» (НЯ), не «побочное», не «неблагоприятное».
 - *Serious adverse event (SAE)* — «серьезное нежелательное явление» (СНЯ).
 - *Outcome* — «исход», а не «результат», не «развязка».
 - *Overall survival (OS)* — «общая выживаемость» (не «совокупная»).
 - *Progression-free survival (PFS)* — «выживаемость без прогрессирования» (ВБП). Только так!
- Рекомендация: каждый новый препарат Ganssen.Oncology — запрашивайте у нас актуальный глоссарий по данной молекуле. <...>

В этом фрагменте я внес лишь несколько правок по стилю и пунктуации. Кстати, о стиле. Нейросеть в целом удачно имитирует корпоративную среду, но иногда увлекается. Из пункта 10, посвященного общим ожиданиям заказчика от исполнителя, цитирую: «Если я прислал текст в пятницу вечером к понедельнику — предупредите в пятницу же, успеваете или нет. Тишина хуже, чем отказ». Очень натуральная реплика, но, наверное, в академической среде не стоит слишком увлекаться чересчур бойким, сварливым, напористым, а иногда и агрессивным корпоративным регистром.

¹ У дипсика есть опция «Обновить»: на один и тот же промпт нейросеть отвечает повторно, неограниченное число раз, причем всегда по-разному.

3. Найдем стайлгайд

Стайлгайд, или Руководство по стилю и правописанию, — это внутренний документ, доступный для ознакомления всем сотрудникам переводческой компании, в том числе внештатным фрилансерам.

Руководство по стилю и правописанию специализированное агентство медицинского и фармацевтического перевода  Москва, 2024 г.	Общие принципы	3
	Требования и рекомендации	4
	Имена собственные	5
	1. Имена и фамилии людей	5
	2. Названия организаций	5
	3. Наименования лекарственных препаратов, медицинских изделий и фармацевтических субстанций	5
	PhD,MD	6
	Адрес	6
	Даты	8
	Аббревиатуры	8
	Числительные	8
	Прописные и строчные буквы	9
	Пунктуация	10
	1. Кавычки	10
	2. Косая черта	10
	3. Пробел	10
	Что такое обратный перевод и как с ним работать?	11

Как правило, стайлгайд рассылается по электронной почте в самом начале сотрудничества с бюро, а то и вовсе на этапе выполнения тестового задания. Мол, ознакомьтесь с нашими правилами и следуйте им. Коллеги-преподаватели, скорее всего, у вас, ну или у ваших знакомых, такой документ имеется, поищите в почте. Как вариант, можно объяснить задачу нейросети и сгенерировать аналогичный документ, разумеется, с последующей тщательной проверкой его содержания. Не обязательно большой, страниц на 10–20. За основу хорошо взять материалы, выложенные в Интернете в открытый доступ, особенно на сайте «Грамота.ру».

4. Сделаем файл памяти переводов

По заданной легенде, перевод у нас уже есть, но качество русского текста оставляет желать лучшего. Сырой машинный перевод будет как нельзя кстати. В принципе, нейросеть умеет генерировать текст для файла tmx — распространенного формата переводческой памяти. Однако этот метод я не рекомендую потому, что нейросеть слишком часто выполняет задачи не в полном объеме или с самыми неожиданными, непредсказуемыми искажениями; результат требует хотя бы беглого просмотра. Текст в формате tmx содержит в себе элементы кода, что затрудняет ознакомление с билингвалом.

```
<?xml version="1.0" encoding="UTF-8"?>
<tmx version="1.4">
  <header creationtool="Manual" creationtoolversion="1.0" segtype="sentence" o-tmf="UTF-8" admin
lang="en" srclang="en" datatype="plaintext"/>
  <body>
    <tu>
      <tuv xml:lang="en"><seg>Background and unmet needs</seg></tuv>
      <tuv xml:lang="ru"><seg>Предпосылки и неудовлетворённые потребности</seg></tuv>
    </tu>
    <tu>
      <tuv xml:lang="en"><seg>Introduction to LPC</seg></tuv>
      <tuv xml:lang="ru"><seg>Введение в локализованный рак предстательной железы (ЛРПЖ)</seg></
tuv>
    </tu>
  </body>
</tmx>
```

Если не tmx, то что? Программа «Смарткат», например, поддерживает файлы памяти в формате xlsx. Экселевская таблица с оригиналом и переводом — это максимально удобно. Отсюда и соответствующий промпт, и ожидаемый результат:

Извлеки текст из файла, расположи все предложения в одну колонку и переведи на русский язык. На выходе нужна таблица из двух колонок с английским оригиналом слева и его русским переводом справа.

< 1 / 2 >

Вот таблица с извлечёнными предложениями из файла (оригинал на английском слева, перевод на русский справа). Все предложения расположены по одному в строке.

English original	Русский перевод
What does localised prostate cancer refer to?	Что подразумевается под локализованным раком предстательной железы?
Localised prostate cancer refers to cancer that is restricted to the prostate gland.	Локализованный рак предстательной железы означает рак, ограниченный предстательной железой.

Посмотрим результат. Иногда вдруг видишь, что машина бросила текст — приходится написать: «А дальше?». Доделывает. Или, допустим, потеряны подписи к изображениям и текстовые выноски, слова и словосочетания в таблицах. Ну правильно, в задаче сказано «расположи все предложения», поэтому нейросеть только с ними и работает, убирая из текста все прочее.

Добившись от программы желаемого результата, обе колонки копируем в эксель (рекомендую гугл-таблицы), даем им правильные названия (в «Смарткате» для английских сегментов — En, для русских — Ru) и сохраняем в формате xlsx.

Файл памяти	
Файл Правка Вид Вставка Формат Данные Инструменты Расширения Справка	
100% 123 По ум... 10 В I A	
A1	en
A	B
1	en
2	What does localised prostate cancer refer to?
3	Localised prostate cancer refers to cancer that is restricted to the prostate gland.
4	Locally advanced prostate cancer refers to cancer that may have spread into the seminal vesicles and nearby tissues.
5	Typical outcomes for patients with localised prostate cancer in the US include a 5-year survival rate of approximately 97%.
6	~74% of men diagnosed with prostate cancer in the US have clinically localised disease.
7	Despite the use of radiation therapy or RP with curative intent, 20–40% of patients experience biochemical recurrence.
8	After biochemical recurrence, patients are typically given ADT.
9	CRPC develops if biochemical recurrence occurs after ADT, despite castration.
10	M1 metastatic disease ensues in many cases; the most common site of metastasis is the bone.
11	Unmet needs in localised prostate cancer include the need for more effective treatments with fewer side effects.
12	Based on the results from a global landscape analysis conducted by the men's health foundation, the most common unmet need is the need for more effective treatments.
13	Develop tests with higher specificity and sensitivity to improve disease screening.
14	Determine the most effective way of treating biochemical recurrence to enhance patient outcomes.
15	Improve the use of ADT to minimise TRAEs.
16	Conduct genomic profiling earlier to optimise treatment and identify likely responders.
17	Develop and validate novel biomarkers for treatment selection.

Загрузить этот файл памяти можно через раздел «Лингвистические ресурсы» своего смарткат-аккаунта, а затем подключить к проекту и провести автоматическую подстановку сегментов перевода из памяти².

Таким образом, мы получим кат-проект, где уже дан перевод, но качество его оценено заказчиком как неприемлемо низкое (по легенде, сгенерированный мною в самом начале через дипсик). Студентам предлагается выступить в роли редактора. Причем, помимо текста оригинала и файла памяти *xlsx*, необходимо изучить и обдумать коммуникативную ситуацию, ознакомиться с требованиями заказчика и корпоративным стилем. Можно ли добавить что-то еще? Можно!

5. Вытащим из сети глоссарии

Схожим образом, что и файл переводческой памяти, создается терминологическая база. За основу можно взять словари, размещенные в Интернете также в открытом доступе. Если глоссарий уже организован в две колонки (и не просто выглядит таковым, а именно копируется как две колонки), достаточно перенести его в эксель, озаглавить колонки *En Term1* и *Ru Term1* соответственно и сохранить в формате *xlsx*. Так, на сайте study-english.info³ находим англо-русский словарь по теме «Медицина». Буквально за пару минут у нас есть терминологическая база на 1400 единиц, которую можно подключить к кат-проекту также через раздел «Лингвистические ресурсы».

² В «Смарткате» есть своя опция машинного перевода. Ее платный вариант для студенческих проектов не подходит. Вариант без оплаты дает слишком низкое качество подстрочника. Если использовать для тех или иных целей машинный перевод как основу будущего текста, то целесообразно выбирать что-то заведомо более высокого качества, например, вариант дипсика или гугла.

³ Дата обращения: 28.05.2026.

	A	B	C
1	en Term 1	ru Term 1	
2	abscess	абсцесс, гнойник	
3	absolute risk reduction	снижение абсолютного риска	
4	abstract	реферат, краткий обзор, резюме; мн.ч. - тезисы	
5	acellular	ацеллюлярный; бесклеточный	
6	acquired immunity	приобретенный иммунитет	
7	acquired immunodeficiency syndrome	синдром приобретённого иммунодефицита (СПИД)	
8	active immunity	активный иммунитет	
9	acute	острый	
10	acute flaccid paralysis (AFP)	острый вялый паралич (ОВП)	
11	acute gastrointestinal infection	острая кишечная инфекция (ОКИ)	
12	acute respiratory disease (ARD)	острое респираторное заболевание (ОРЗ)	
13	acute respiratory infection (ARI)	острая респираторная инфекция (ОРИ)	
14	acute respiratory viral infection (ARVI)	острая респираторная вирусная инфекция (ОРВИ)	

Если словарь, найденный в сети, оформлен не в виде таблицы, то задачу по размещению английских терминов и их русских соответствий в две колонки можно поручить нейросети с последующим копированием результата в эксель.

6. Обзаведемся корпусом параллельных текстов

В кат-программах важнейший инструмент — поиск по конкордансу. С помощью него переводчик/редактор получает точечный доступ ко всему корпусу двуязычных текстов, содержащихся в подключенных к проекту файлах переводческой памяти. Это ценный корпоративный ресурс, в нем лингвистический опыт, накопленный бюро за годы своей деятельности. Очевидно, что у преподавателя, а тем более у студентов, доступа к таким ресурсам быть не может. Однако некоторый аналог можно создать самому. В идеале — найти по тематике проекта параллельные тексты, то есть тексты-аналоги на английском, качественно переведенные на русский. Выровнять их не проблема, с этим справится нейросеть. Если параллельных текстов нет, можно сделать их самим.

По легенде проекта у нас, напомним, «завернутый» заказчиком перевод с английского на русский, по теме «рак предстательной железы». В качестве ориентира нам прежде всего интересны русские тексты по этой же теме. Находим их в Интернете: научные статьи про ЛРПЖ, справочно-энциклопедические статьи, документы, например, клинические рекомендации, в частности, по стадированию. Промпт можно составить так:

Из русского текста извлеки все предложения. Расположи их в одну колонку. В другой колонке, расположенной слева, размести английский перевод этих предложений.

Поясню нюанс. Конкордансный поиск в кат-программах имеет два режима: поиск по языку оригинала и по языку перевода. В данном случае «оригинал»⁴ будет на английском — сгенерированный машиной текст. По нему можно вести поиск с целью посмотреть, как это будет по-русски, то есть в образцовом аутентичном тексте. Но в целом английский машинный перевод, конечно, обладает сравнительно низкой ценностью. Другое дело — русский оригинал. В нем и предлагается вести поиск в первую очередь. А когда таких билингвальных файлов к проекту подключается несколько, мы получаем корпус образцовых русских текстов, с которым студенты могут работать.

Низке приведены все предложения из предоставленного русского текста (начиная с аннотации и ключевых слов, заканчивая последним предложением раздела 3.3.4) в одной колонке, а слева — их английский перевод. Пустые строки между предложениями сохранены для удобства чтения.

English Translation	Russian Original
DOI: 10.18027/2224-5057-2023-13-3s2-1-640-660	DOI: 10.18027/2224-5057-2023-13-3s2-1-640-660
Citation: Nosov D.A., Volkova M.I., Gladkov O.A., Karabina E.V., Krylov V.V., Matveev V.B. et al. Practical recommendations for drug treatment of prostate cancer. RUSSCO Practical Recommendations, part 1. Malignant tumors, 2023 (vol. 13), #3s2, pp. 640–660.	Цитирование: Носов Д.А., Волкова М.И., Гладков О.А., Карабина Е.В., Крылов В.В., Матвеев В.Б. и соавт. Практические рекомендации по лекарственному лечению рака предстательной железы. Практические рекомендации RUSSCO, часть 1. Злокачественные опухоли, 2023 (том 13), #3s2, стр. 640–660.
PROSTATE CANCER	РАК ПРЕДСТАТЕЛЬНОЙ ЖЕЛЕЗЫ
Authors: Nosov D.A., Volkova M.I., Gladkov O.A., Karabina E.V., Krylov V.V., Matveev V.B., Usychkin S.V., T. Mitin	Коллектив авторов: Носов Д.А., Волкова М.И., Гладков О.А., Карабина Е.В., Крылов В.В., Матвеев В.Б., Усичкин С.В., Т. Митин
Key words: prostate cancer, androgen deprivation, docetaxel, cabazitaxel, abiraterone acetate,	Ключевые слова: рак предстательной железы, андрогенная депривация, доцетаксел, ка-

Напомню, что здесь мы действуем по той же схеме, что описали в пункте 4, когда делали файл памяти переводов: копируем в эксель, называем колонки En и Ru соответственно, загружаем в «Лингвистические ресурсы» кат-аккаунта и подключаем к своему проекту.

7. Нужны ли все эти сложности?

Безусловно да. Все эти технические манипуляции — представленная фрагментарно, хотя и в концентрированном виде, рутинная современная цифровая переводчика (и уж тем более преподавателя, который таких переводчиков обучает). Умение работать в кат-программах не сводится к тому, чтобы перейти по ссылке, полученной по почте

⁴ В файле памяти английский машинный перевод будет в колонке слева, то есть займет место оригинала, тогда как русский оригинал — справа, в позиции перевода. Так получается потому, что файл мы подключим к проекту «с английского на русский». По техническим причинам в языковой паре направление перевода должно совпадать обязательно!

от проектного менеджера, и там, найдя сегментированный оригинал и частично выполненный перевод, методично нажимать на клавиши Ctrl+Enter, записывая сегмент за сегментом в память.

Необходимо, во-первых, знать инструменты (а их много!) и уметь с ними работать. Во-вторых, понимать хотя бы в общих чертах принцип действия кат-программ, как в них функционируют глоссарии и память переводов, уметь все это настроить под себя. В-третьих, быть ближе к будням переводческой отрасли, которая не обходится без терминологических баз и переводческой памяти — а эти ресурсы надо откуда-то брать.

Думается, что мастерство в работе с технологиями проявляется не только в том, чтобы использовать их по прямому назначению, но также в умении преподавателя приспособить любой высокотехнологичный продукт под свои академические нужды, извлечь из него максимум пользы для студентов.