Машинный перевод

*Не следует путать с Автоматизированный перевод. Эта статья о переводчиках — компьютерных программах; о переводчиках-людях.*

Ельчанинова А.Р.

Оглавление

1. Машинный перевод………………………………………………………………………….2

### 2. Автоматизированный перевод…………………………………………………………………………..3

## 3. История машинного перевода………………………………………………………….…4-5

## 4. Виды систем машинного перевода……………………………6

## 5. Качество машинного перевода…………………………………7-8

## 6. Предварительная подготовка текста………………………….9

## 7. Что можно и что нельзя переводить машинным

## способом ..………………………………………………………………………10

8. Зачем нужен машинный перевод?.................................11

**1. Машинный перевод.**

**Маши́нный перево́д** — процесс перевода текстов (письменных, а в идеале и устных) с одного естественного языка на другой с помощью специальной компьютерной программы. Так же называется направление научных исследований, связанных с построением подобных систем.

За более чем полвека своего существования машинный перевод из полуфантастического проекта стал реальностью. Разработчики систем МП уже не боятся, что бьются над заведомо нерешаемой задачей, а переводчики — что восстание машин лишит их хлеба насущного. Компьютер вряд ли когда-нибудь заменит переводчика во всех областях, однако машинный перевод уже активно применяется как полезный инструмент, пусть и подходящий для решения строго определённых задач. Для того чтобы понять, в каких случаях стоит и в каких не стоит прибегать к машинному переводу, и чтобы полученный результат не разочаровал, важно знать принципы работы МП, его ограничения и область его применимости.

### 2. Автоматизированный перевод

Вместо *«машинный»* иногда употребляется слово **автоматический**, что не влияет на смысл. Однако термин [**автоматизированный перевод**](https://ru.wikipedia.org/wiki/%D0%90%D0%B2%D1%82%D0%BE%D0%BC%D0%B0%D1%82%D0%B8%D0%B7%D0%B8%D1%80%D0%BE%D0%B2%D0%B0%D0%BD%D0%BD%D1%8B%D0%B9_%D0%BF%D0%B5%D1%80%D0%B5%D0%B2%D0%BE%D0%B4) имеет совсем другое значение — при нём программа просто *помогает человеку* переводить тексты.

Автоматизированный перевод предполагает такие формы взаимодействия:

* Частично автоматизированный перевод: например, использование переводчиком-человеком компьютерных [словарей](https://ru.wikipedia.org/wiki/%D0%A1%D0%BB%D0%BE%D0%B2%D0%B0%D1%80%D1%8C).
* Системы с разделением труда: компьютер обучен переводить только фразы жёстко заданной структуры (но делает это так, чтобы исправлять за ним не требовалось), а всё, не уложившееся в схему, отдаёт человеку.

В англоязычной терминологии также различаются термины [англ.](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) *machine translation, MT* (полностью автоматический перевод) и [англ.](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) *machine-aided* или [англ.](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) *machine-assisted translation (MAT)* (автоматизированный); если же надо обозначить и то, и другое, пишут M(A)T.

Существуют два принципиально разных подхода к построению алгоритмов машинного перевода: [основанный на правилах](https://ru.wikipedia.org/wiki/%D0%9C%D0%B0%D1%88%D0%B8%D0%BD%D0%BD%D1%8B%D0%B9_%D0%BF%D0%B5%D1%80%D0%B5%D0%B2%D0%BE%D0%B4_%D0%BD%D0%B0_%D0%BE%D1%81%D0%BD%D0%BE%D0%B2%D0%B5_%D0%BF%D1%80%D0%B0%D0%B2%D0%B8%D0%BB) (rule-based) и статистический, или [основанный на статистике](https://ru.wikipedia.org/wiki/%D0%A1%D1%82%D0%B0%D1%82%D0%B8%D1%81%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B8%D0%B9_%D0%BC%D0%B0%D1%88%D0%B8%D0%BD%D0%BD%D1%8B%D0%B9_%D0%BF%D0%B5%D1%80%D0%B5%D0%B2%D0%BE%D0%B4) (statistical-based). Первый подход является традиционным и используется большинством разработчиков систем машинного перевода (ПРОМТ в России, SYSTRAN во Франции, Linguatec в Германии и др.). Ко второму типу относится популярный сервис [Яндекс.Переводчик](https://ru.wikipedia.org/wiki/%D0%AF%D0%BD%D0%B4%D0%B5%D0%BA%D1%81.%D0%9F%D0%B5%D1%80%D0%B5%D0%B2%D0%BE%D0%B4%D1%87%D0%B8%D0%BA" \o "Яндекс.Переводчик), Переводчик Google, а также новый сервис от ABBYY.

**3. История машинного перевода**

Мысль использовать ЭВМ для перевода была высказана в [1947 году](https://ru.wikipedia.org/wiki/1947_%D0%B3%D0%BE%D0%B4) в [США](https://ru.wikipedia.org/wiki/%D0%A1%D0%A8%D0%90), сразу после появления первых ЭВМ. Первая публичная [демонстрация](https://ru.wikipedia.org/wiki/%D0%94%D0%B5%D0%BC%D0%BE%D0%BD%D1%81%D1%82%D1%80%D0%B0%D1%86%D0%B8%D1%8F) машинного перевода (так называемый [Джорджтаунский эксперимент](https://ru.wikipedia.org/wiki/%D0%94%D0%B6%D0%BE%D1%80%D0%B4%D0%B6%D1%82%D0%B0%D1%83%D0%BD%D1%81%D0%BA%D0%B8%D0%B9_%D1%8D%D0%BA%D1%81%D0%BF%D0%B5%D1%80%D0%B8%D0%BC%D0%B5%D0%BD%D1%82" \o "Джорджтаунский эксперимент)) состоялась в [1954 году](https://ru.wikipedia.org/wiki/1954_%D0%B3%D0%BE%D0%B4). Несмотря на примитивность той системы (словарь в 250 слов, [грамматика](https://ru.wikipedia.org/wiki/%D0%93%D1%80%D0%B0%D0%BC%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0_(%D0%BE%D0%BF%D0%B8%D1%81%D0%B0%D0%BD%D0%B8%D0%B5_%D1%8F%D0%B7%D1%8B%D0%BA%D0%B0)) из 6 правил, перевод нескольких простых фраз), этот эксперимент получил широкий резонанс: начались исследования в [Англии](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D1%8F), [Болгарии](https://ru.wikipedia.org/wiki/%D0%91%D0%BE%D0%BB%D0%B3%D0%B0%D1%80%D0%B8%D1%8F), [ГДР](https://ru.wikipedia.org/wiki/%D0%93%D0%94%D0%A0), [Италии](https://ru.wikipedia.org/wiki/%D0%98%D1%82%D0%B0%D0%BB%D0%B8%D1%8F), [Китае](https://ru.wikipedia.org/wiki/%D0%9A%D0%B8%D1%82%D0%B0%D0%B9), [Франции](https://ru.wikipedia.org/wiki/%D0%A4%D1%80%D0%B0%D0%BD%D1%86%D0%B8%D1%8F), [ФРГ](https://ru.wikipedia.org/wiki/%D0%A4%D0%A0%D0%93), [Японии](https://ru.wikipedia.org/wiki/%D0%AF%D0%BF%D0%BE%D0%BD%D0%B8%D1%8F) и других странах; в том же 1954 году и в [СССР](https://ru.wikipedia.org/wiki/%D0%A1%D0%A1%D0%A1%D0%A0).

К середине [1960-х](https://ru.wikipedia.org/wiki/1960-%D0%B5) в США для практического использования были предоставлены две системы русско-английского перевода:

* MARK (в Департаменте иностранной техники ВВС США);
* GAT (разработка Джорджтаунского университета, использовалась в Национальной лаборатории атомной энергии в Окридже и в центре Евратома в г. Испра, Италия).

Однако созданная для оценки подобных систем комиссия [ALPAC](https://ru.wikipedia.org/w/index.php?title=ALPAC&action=edit&redlink=1)[[en]](https://en.wikipedia.org/wiki/ALPAC) пришла к выводу, что в силу низкого качества машинно переведённых текстов эта деятельность в условиях США нерентабельна. Хотя комиссия рекомендовала продолжать и углублять теоретические разработки, в целом её выводы привели к росту [пессимизма](https://ru.wikipedia.org/wiki/%D0%9F%D0%B5%D1%81%D1%81%D0%B8%D0%BC%D0%B8%D0%B7%D0%BC), снижению финансирования, часто к полному прекращению работ по этой тематике.

Тем не менее, в ряде стран исследования продолжались, чему способствовал постоянный прогресс вычислительной техники. Особенно существенным фактором стало появление мини- и [персональных компьютеров](https://ru.wikipedia.org/wiki/%D0%9F%D0%B5%D1%80%D1%81%D0%BE%D0%BD%D0%B0%D0%BB%D1%8C%D0%BD%D1%8B%D0%B9_%D0%BA%D0%BE%D0%BC%D0%BF%D1%8C%D1%8E%D1%82%D0%B5%D1%80), а с ними всё более сложных словарных, [поисковых](https://ru.wikipedia.org/wiki/%D0%9F%D0%BE%D0%B8%D1%81%D0%BA%D0%BE%D0%B2%D0%B0%D1%8F_%D1%81%D0%B8%D1%81%D1%82%D0%B5%D0%BC%D0%B0) и т. п. систем, ориентированных на работу с естественноязыковыми данными. Росла и необходимость в переводе как таковом ввиду роста международных связей. Все это привело к новому подъёму этой области, наступившему примерно с середины [1970-х](https://ru.wikipedia.org/wiki/1970-%D0%B5). В [1980-е](https://ru.wikipedia.org/wiki/1980-%D0%B5) наступило время широкого практического использования переводческих систем, сложился рынок коммерческих разработок по этой теме.

Впрочем, мечты, с которыми человечество полвека назад взялось за задачу машинного перевода, в значительной мере остаются мечтами: высококачественный перевод текстов широкой тематики по-прежнему недостижим. Однако несомненным является ускорение работы переводчика при использовании систем машинного перевода: по оценкам конца 1980-х, до пяти раз.

В настоящее время существует множество коммерческих проектов машинного перевода. Одним из пионеров в области машинного перевода была компания [SYSTRAN](https://ru.wikipedia.org/w/index.php?title=SYSTRAN&action=edit&redlink=1). В [России](https://ru.wikipedia.org/wiki/%D0%A0%D0%BE%D1%81%D1%81%D0%B8%D1%8F) большой вклад в развитие машинного перевода внесла группа под руководством проф. [Р. Г. Пиотровского](https://ru.wikipedia.org/wiki/%D0%9F%D0%B8%D0%BE%D1%82%D1%80%D0%BE%D0%B2%D1%81%D0%BA%D0%B8%D0%B9,_%D0%A0%D0%B0%D0%B9%D0%BC%D1%83%D0%BD%D0%B4_%D0%93%D0%B5%D0%BD%D1%80%D0%B8%D1%85%D0%BE%D0%B2%D0%B8%D1%87) ([Российский государственный педагогический университет имени А. И. Герцена](https://ru.wikipedia.org/wiki/%D0%A0%D0%BE%D1%81%D1%81%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D0%B3%D0%BE%D1%81%D1%83%D0%B4%D0%B0%D1%80%D1%81%D1%82%D0%B2%D0%B5%D0%BD%D0%BD%D1%8B%D0%B9_%D0%BF%D0%B5%D0%B4%D0%B0%D0%B3%D0%BE%D0%B3%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B8%D0%B9_%D1%83%D0%BD%D0%B8%D0%B2%D0%B5%D1%80%D1%81%D0%B8%D1%82%D0%B5%D1%82_%D0%B8%D0%BC%D0%B5%D0%BD%D0%B8_%D0%90._%D0%98._%D0%93%D0%B5%D1%80%D1%86%D0%B5%D0%BD%D0%B0), [Санкт-Петербург](https://ru.wikipedia.org/wiki/%D0%A1%D0%B0%D0%BD%D0%BA%D1%82-%D0%9F%D0%B5%D1%82%D0%B5%D1%80%D0%B1%D1%83%D1%80%D0%B3)).

**4. Виды систем машинного перевода**

**Системы машинного перевода делятся на три категории**: *системы на* *основе грамматических правил* (Rule-Based Machine Translation, RBMT*), статистические системы* (Statistical Machine Translation, SMT) и наиболее перспективные *гибридные системы*, сочетающие преимущества тех и других. Системы RBMT анализируют текст и строят его перевод на базе встроенных словарей и набора правил для данной языковой пары. В системах SMT применяется принцип статистического анализа: в программу загружаются большие (в миллионы слов) объёмы текстов на исходном языке и их переводы, выполненные человеком. Программа анализирует статистику межъязыковых соответствий, словоупотребления, синтаксических конструкций и т. д., и позже опирается на неё при выборе вариантов перевода — этот процесс называется самообучением. Систему может обучать и человек, корректируя выдаваемые переводы. Именно так работает широко известный сервис «Переводчик Google». Благодаря способности статистических и гибридных систем МП обучаться, накапливая языковые данные, качество перевода у них повышается с каждым переведённым текстом.

**5. Качество машинного перевода**

Качество перевода зависит от тематики и [стиля](https://ru.wikipedia.org/wiki/%D0%A1%D1%82%D0%B8%D0%BB%D0%B8_%D1%80%D0%B5%D1%87%D0%B8) исходного текста, а также грамматической, синтаксической и лексической родственности языков, между которыми производится перевод. Машинный перевод [художественных текстов](https://ru.wikipedia.org/wiki/%D0%A4%D1%83%D0%BD%D0%BA%D1%86%D0%B8%D0%BE%D0%BD%D0%B0%D0%BB%D1%8C%D0%BD%D1%8B%D0%B5_%D1%81%D1%82%D0%B8%D0%BB%D0%B8_%D1%80%D0%B5%D1%87%D0%B8#%D0%A5%D1%83%D0%B4%D0%BE%D0%B6%D0%B5%D1%81%D1%82%D0%B2%D0%B5%D0%BD%D0%BD%D1%8B%D0%B9_%D1%81%D1%82%D0%B8%D0%BB%D1%8C) практически всегда оказывается неудовлетворительного качества. Тем не менее для технических документов при наличии специализированных машинных словарей и некоторой настройке системы на особенности того или иного типа текстов возможно получение перевода приемлемого качества, который нуждается лишь в небольшой редакторской корректировке. Чем более [формализован](https://ru.wikipedia.org/wiki/%D0%A4%D0%BE%D1%80%D0%BC%D0%B0%D0%BB%D0%B8%D0%B7%D0%B0%D1%86%D0%B8%D1%8F) стиль исходного документа, тем большего качества перевода можно ожидать. Самых лучших результатов при использовании машинного перевода можно достичь для текстов, написанных в техническом (различные описания и руководства) и [официально-деловом стиле](https://ru.wikipedia.org/wiki/%D0%A4%D1%83%D0%BD%D0%BA%D1%86%D0%B8%D0%BE%D0%BD%D0%B0%D0%BB%D1%8C%D0%BD%D1%8B%D0%B5_%D1%81%D1%82%D0%B8%D0%BB%D0%B8_%D1%80%D0%B5%D1%87%D0%B8#%D0%9E%D1%84%D0%B8%D1%86%D0%B8%D0%B0%D0%BB%D1%8C%D0%BD%D0%BE-%D0%B4%D0%B5%D0%BB%D0%BE%D0%B2%D0%BE%D0%B9_%D1%81%D1%82%D0%B8%D0%BB%D1%8C).

Применение машинного перевода без настройки на тематику (или с намеренно неверной настройкой) служит предметом многочисленных бытующих в [Интернете](https://ru.wikipedia.org/wiki/%D0%98%D0%BD%D1%82%D0%B5%D1%80%D0%BD%D0%B5%D1%82) шуток. Из старых и наиболее популярных примеров таких шуток наиболее известен текст перевода документации к [драйверу](https://ru.wikipedia.org/wiki/%D0%94%D1%80%D0%B0%D0%B9%D0%B2%D0%B5%D1%80) [мыши](https://ru.wikipedia.org/wiki/%D0%9A%D0%BE%D0%BC%D0%BF%D1%8C%D1%8E%D1%82%D0%B5%D1%80%D0%BD%D0%B0%D1%8F_%D0%BC%D1%8B%D1%88%D1%8C), известный как [«Гуртовщики Мыши»](http://www.lib.ru/ANEKDOTY/mouse_driver.txt), заявленный как «перевод компьютерной документации системой машинного перевода Poliglossum на основе медицинского, коммерческого и юридического словарей». Из кратких — фраза «*Our cat gave birth to three kittens — two whites and one black*», которую онлайн-переводчик «[ПРОМТ](https://ru.wikipedia.org/wiki/%D0%9F%D0%A0%D0%9E%D0%9C%D0%A2)» (версия 7.0, [2007](https://ru.wikipedia.org/wiki/2007)) превращал в «Наш кот родил трёх котят — двух белых и одного [афроамериканца](https://ru.wikipedia.org/wiki/%D0%90%D1%84%D1%80%D0%BE%D0%B0%D0%BC%D0%B5%D1%80%D0%B8%D0%BA%D0%B0%D0%BD%D1%86%D1%8B)». Если «афроамериканца» ещё можно было сделать «чёрным», написав «*black kitten*», то «коту» так и не получалось сменить пол: например, *female cat* переводился как «самка кот».

Чаще всего подобные шутки связаны с тем, что программа не распознаёт контекст фразы и переводит термины дословно, к тому же не отличая собственных имён от обычных слов. Тот же переводчик ПРОМТ превращает «Лев Толстой» в «Lion Thick» («толстый лев»), «[bra-ket notation](https://ru.wikipedia.org/wiki/%D0%91%D1%80%D0%B0_%D0%B8_%D0%BA%D0%B5%D1%82" \o "Бра и кет)» в «примечание Кети лифчика», «[Lie algebra](https://ru.wikipedia.org/wiki/%D0%90%D0%BB%D0%B3%D0%B5%D0%B1%D1%80%D0%B0_%D0%9B%D0%B8" \o "Алгебра Ли)» — в «алгебру Лжи», «[eccentricity vector](https://ru.wikipedia.org/wiki/%D0%92%D0%B5%D0%BA%D1%82%D0%BE%D1%80_%D1%8D%D0%BA%D1%81%D1%86%D0%B5%D0%BD%D1%82%D1%80%D0%B8%D1%81%D0%B8%D1%82%D0%B5%D1%82%D0%B0" \o "Вектор эксцентриситета)» — в «вектор оригинальности», «[Shawnee Smith](https://ru.wikipedia.org/wiki/%D0%A8%D0%BE%D1%83%D0%BD%D0%B8_%D0%A1%D0%BC%D0%B8%D1%82" \o "Шоуни Смит)» в «индеец племени шони Смит» и т. п. [Переводчик Google](https://ru.wikipedia.org/wiki/Google_Translate), наоборот, слово «[*rice*](https://ru.wikipedia.org/wiki/%D0%A0%D0%B8%D1%81)» часто принимал за фамилию [госсекретаря США](https://ru.wikipedia.org/wiki/%D0%A0%D0%B0%D0%B9%D1%81,_%D0%9A%D0%BE%D0%BD%D0%B4%D0%BE%D0%BB%D0%B8%D0%B7%D0%B0).

## 6. Предварительная подготовка текста

Предварительная подготовка материалов может значительно упростить задачу системе МП и редакторам, которые будут дорабатывать сырой машинный перевод. Такая подготовка начинается ещё на этапе написания исходного текста — с этой целью для технических писателей и авторов разрабатываются стандарты, соблюдение которых позволяет сделать текст более простым для понимания и перевода, как машинного, так и человеческого.

**Существуют три правила, выполнение которых наиболее существенно повышает качество машинного перевода с английского языка.**

1. Использование глаголов в действительном залоге вместо герундия;
2. Использование активного залога вместо пассивного;
3. Отказ от использования составных предложений и однородных членов.

В идеале каждое предложение должно содержать одну законченную мысль. Именно это правило, в равной степени применимое для всех языков, является самым эффективным из трёх.

**Постредактирование** — машинный перевод, редактируемый человеком.

**Предредактирование** – человек выполняет предварительную подготовку текста для машинного перевода.

## Интерредактирование – человек вмешивается в процесс машинного перевода, устраняя ошибки и конфликты.

## Смешанные способы комбинации вышеперечисленных способов.

## 7. Что можно и что нельзя переводить машинным способом

Главное преимущество машинного перевода в том, что он позволяет быстро справиться с очень большими объёмами текста и поэтому иногда оказывается экономически выгоднее перевода вручную. При этом следует помнить, что качество машинного перевода всегда будет уступать человеческому. Поэтому использовать его целесообразно лишь в определённых случаях

Во-первых, машинным способом могут переводиться материалы длявнутреннего пользования, например, когда требуется в общих чертах понять содержание сайта, статей или писем на иностранном языке или найти сообщения на ту или иную тему в прессе на нескольких языках мира. Во-вторых,технические и узкоспециальные тексты, которые затем пойдут на редактирование специалистам по данной тематике — в этом случае машинный перевод используется в качестве подстрочника, на основе которого технический специалист будет создавать финальный текст, опираясь на свои знания в предметной области.

Многие типы материалов в принципе не предназначены для машинного перевода. Так, нельзя доверять машине тексты, где неточность перевода может поставить под угрозу здоровье человека, работоспособность сложного прибора или крупный контракт — сэкономленное время здесь не оправдывает риска. Любые документы, подразумевающие юридическую ответственность (договоры, гарантийные обязательства), требуют контроля человека. Машинный перевод непригоден для маркетинговых материалов, где текст фактически переосмысливается в новом культурном контексте и создается заново.

В целом, приемлемого качества можно ожидать при переводе строго формализованных технических текстов, в то время как художественные и рекламные тексты машинному переводу не поддаются.

**8. Зачем нужен машинный перевод?**

**Программы-переводчики оказываются просто незаменимыми, когда возникает необходимость:**

* быстро понять общий смысл текста,
* быстро проанализировать многоязычную информацию из интернета,
* оптимизировать перевод большого объема текстов по одной тематике (процесс разбивается на два этапа: машинный перевод и постредактирование человеком),
* сделать информацию на сайте понятной иностранным пользователям, встроив функцию перевода на сайт,
* быстро вести переписку с иностранными партнерами, а также свободно общаться в социальных сетях и на форумах,
* извлечь информацию из большого объема текстов, опираясь на лингвистические характеристики слов. Например, выяснить, сколько раз в тексте встречаются имена собственные, какие это имена и какие действия связаны с этими именами.

Спасибо за внимание!