

# Проектная работа по информатике по теме “Машинный перевод”

Подготовил студент первого курса ТиМп нем.яз.

Малашенкова Елизавета

## Машинный и автоматизированный перевод.

Размытые формулировки и неточные термины нередко служат причиной для непрекращающихся споров, особенно среди непрофессионалов. Ярким примером такой ситуации стали «автоматизированный перевод» и «машинный перевод».

Путаница между ними возникла из-за того, что и первое, и второе чаще всего ассоциируется с выполнением работы программами вместо человека. И в эту ловушку попадают многие – заказчики, студенты специализированных факультетов или начинающие переводчики.

Что такое машинный перевод и автоматизированный перевод

**Автоматизированный перевод** – перевод текстов с одного языка на другой человеком с использованием специализированных программ, приложений.

**Машинный перевод** – процесс перевода текстов с одного языка на другой машиной посредством специальной компьютерной программы.

Между ними есть три главных различия:

- • Трудозатраты переводчика.
- • Специализированное программное обеспечение.
- • Качество.

Так, при **автоматизированном переводе** основную работу выполняет переводчик, а специальные программы выступают только в качестве вспомогательного инструмента. Его главной целью является сокращение времени процесса, обеспечение единообразия терминов и общего соответствия (качества).

**Машинный же перевод** производит сама программа, а человек только редактирует полученный результат. И здесь уже цель – заменить труд человека, получая при этом быстрый перевод низкого качества.

### Специальные программы

Системы **автоматизированного перевода** – это собирательное определение для специализированных программ и интернет-сервисов, которые используют переводчики в процессе работы. Они хорошо подходят для работы над художественными, юридическими и техническими переводами. К ним относятся:

- • Отдельные либо встроенные редакторы, которые помогают автоматически проверить грамматику текста и правописание слов.
- • ПО, таблицы, текстовые редакторы, обеспечивающие управление терминологией (MultiTerm, Termex и т.д.).
- • ПО, с помощью которого осуществляется менеджмент переводческих проектов.
- • САТ-инструменты, использующие ТМ (Translation Memory – Память переводов), которые включают в себя образцы ранее переведённых текстов или предложений. Яркими её представителями стали программы Trados, Déjà Vu, MemoQ, MemSource, Wordfast.
- • Корпусы, представляющие собой большой ряд документов с использованием одного или нескольких языков. С их помощью составляется сжатое описание употребления слов и выражений в общих случаях или с учётом какой-либо определённой предметной темы.

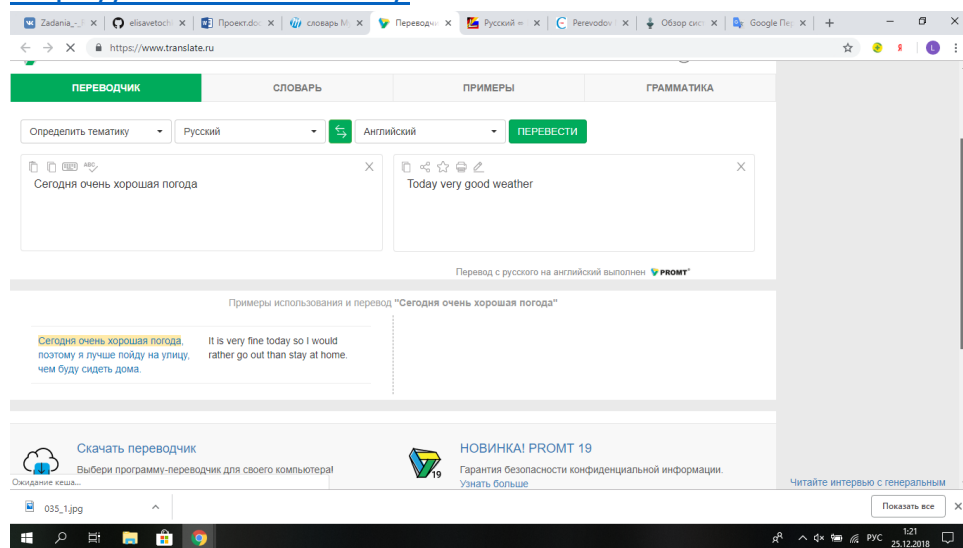
**Машинный перевод** также представлен набором программ и интернет-сервисов. Наиболее популярные из них: PROMT, Google translate, Яндекс.Переводчик.

### Системы машинного перевода

На сегодняшний день, одними из часто используемых систем являются такие системы машинного перевода, как:

-PROMT 2000/ХТ компании PROMT

<https://www.translate.ru/>

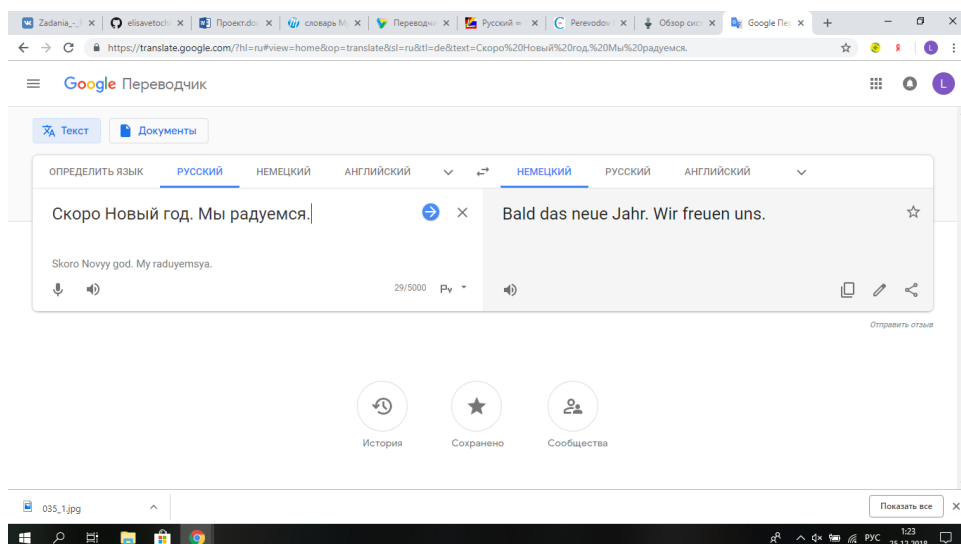


-Retrans Vista компаний Vista и Advantis



-Переводчик Google Translate

<https://translate.google.com/?hl=ru>



## **К преимуществам программ машинного перевода относятся:**

1. Быстрый доступ и высокая скорость.

2. Экономичность.

Система машинного перевода в онлайн -режиме не требует оплаты, необходим только доступ к Интернету.

3. Защита и безопасность информации

4. Гибкость и универсальность.

Понятие гибкость подразумевает возможность настройки либо на конкретную предметную область (специализированные словари), либо настройки на конкретную книгу или текст (словари, созданные пользователями).

## **Минусы:**

1. Электронные переводчики адекватно переводят простые части речи, но не всегда справляются с переводом падежей, устойчивых оборотов, фразеологизмов, построения предложения.

Некоторые слова электронный переводчик оставляет без перевода.

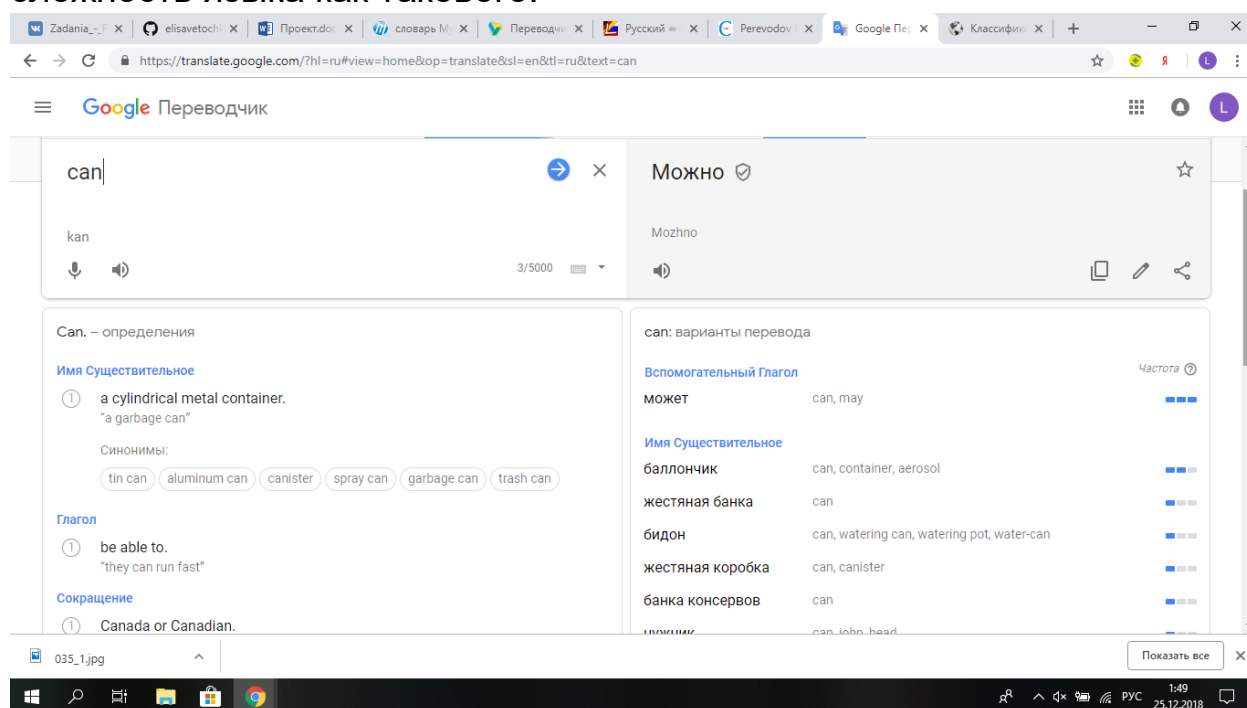
В таких случаях необходимо подбирать синонимы, перестраивать структуру предложения, т. е. зачастую перевод нуждается в корректировке, редактировании. А это очень кропотливая работа и она отнимает довольно таки много сил и времени.

Например: словосочетание

«Theater machen» PROMT переводит как «театры делают», но на самом деле это устойчивое выражение переводится «ломать комедию».

# Классификация систем машинного перевода

**Полностью автоматизированный машинный перевод.** Этот вид машинного перевода и подразумевается большинством людей, когда они говорят о машинном переводе. Смысл здесь прост: в компьютер вводится текст на одном языке, этот текст обрабатывается и компьютер выводит этот же текст на другом языке. К сожалению, реализация такого вида автоматического перевода сталкивается с определенными препятствиями, которые еще предстоит преодолеть. Основной проблемой является сложность языка как такового.



**Автоматизированный машинный перевод при участии человека.** Этот вид машинного перевода теперь вполне осуществим. Говоря о машинном переводе при участии человека, обычно подразумевают редактирование текстов как до, так и после их обработки компьютером. Люди-переводчики изменяют тексты так, чтобы они были понятны машинам. После того, как компьютер сделал перевод, люди опять-таки редактируют грубый машинный перевод, делая текст на выходном языке правильным. Помимо такого порядка работы, существуют системы МП, во время перевода требующие постоянного присутствия человека-переводчика, помогающего компьютеру делать перевод особенно сложных или неоднозначных конструкций.

**Машинный перевод с помощью человека применим в большей степени к текстам с ограниченным вокабуляром узко-ограниченной тематики.** Экономичность использования машинного перевода с помощью человека – вопрос все еще спорный. Сами программы обычно достаточно дорогостоящи, а для работы некоторых из них требуется специальное оборудование. Предварительному и последующему редактированию необходимо обучаться, да и работа эта не из приятных. Создание и поддержание в рабочем состоянии баз данных слов – процесс трудоемкий и зачастую требует специальных навыков. Однако для организации, переводящей большие объемы текстов в четко-определенной тематической сфере, машинный перевод с помощью человека может оказаться достаточно экономичной альтернативой традиционному человеческому переводу.

**Перевод, осуществляемый человеком с использованием компьютера.** При этом подходе человек-переводчик ставится в центр процесса перевода, в то время как программа компьютера расценивается в качестве инструмента, делающего процесс перевода более эффективным, а перевод – точным. Это обычные электронные словари, которые обеспечивают перевод требуемого слова, возлагая на человека ответственность за выбор нужного варианта и смысл переведенного текста. Такие словари значительно облегчают процесс перевода, но требуют от пользователя определенного знания языка и затрат времени на его осуществление. И все же сам процесс перевода значительно ускоряется и облегчается.

## **Причина создания машинного перевода**

История машинного перевода берет начало в XVII веке, когда такими философами, как Лейбниц и Декарт было выдвинуто предположение о существовании некоего кода, соединяющего между собой слова разных языков. Все предположения носили гипотетический характер, и никому не удавалось в действительности создать машину для перевода.

Первые заявления на получение патента на «переводческую машину» были поданы в середине 30-х годов XX века. Одно заявление было подано французским изобретателем Ж. Арцруни, просто создавшим автоматический двуязычный словарь на перфоленте. Другое было сделано советским ученым Петром Троянским, чье изобретение было более детальным. Оно включало в себя, как двуязычный словарь,

## Лабораторная работа

Yandex Переводчик

ТЕКСТ САЙТ КАРТИНКА

malaschenkova.liza2015

РУССКИЙ ↔ АНГЛИЙСКИЙ

В этом доме живет большая семья

31 / 10000

А big family lives in this house

Перевести в Google Bing

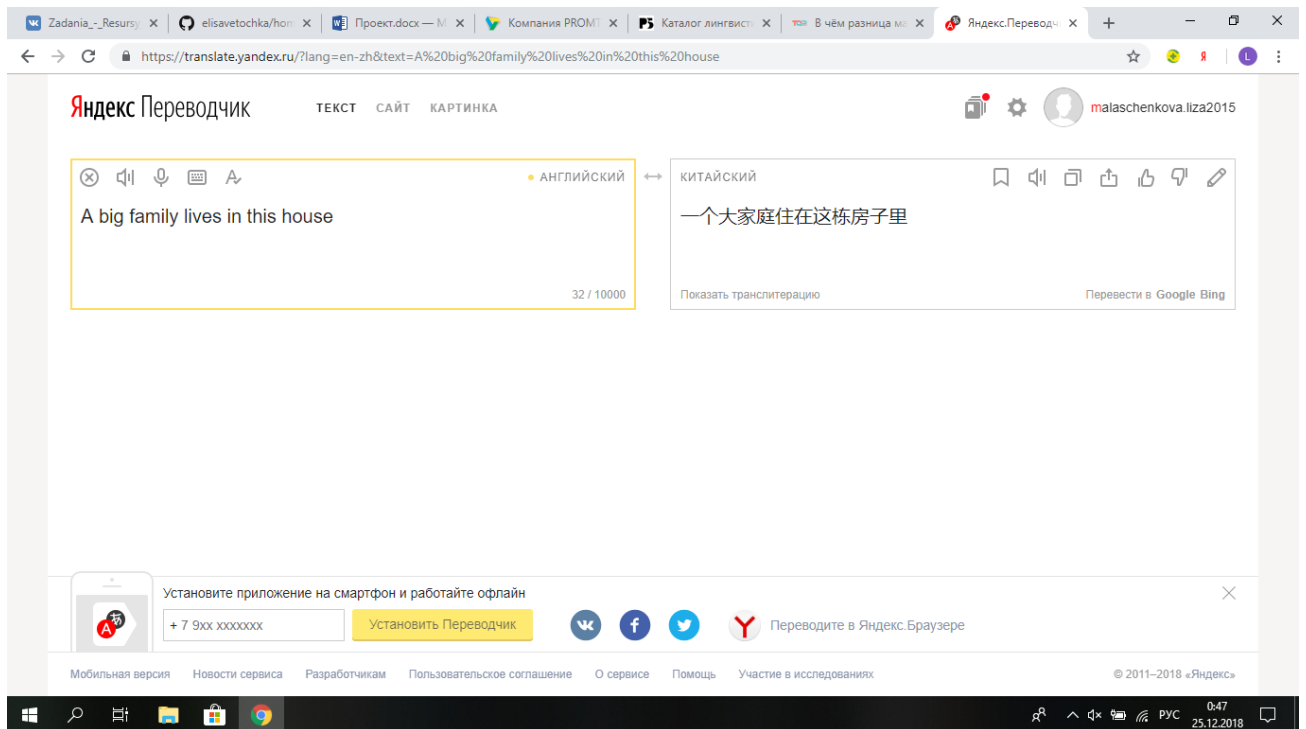
Установите приложение на смартфон и работайте офлайн

+7 9xx xxxxxxxx Установить Переводчик

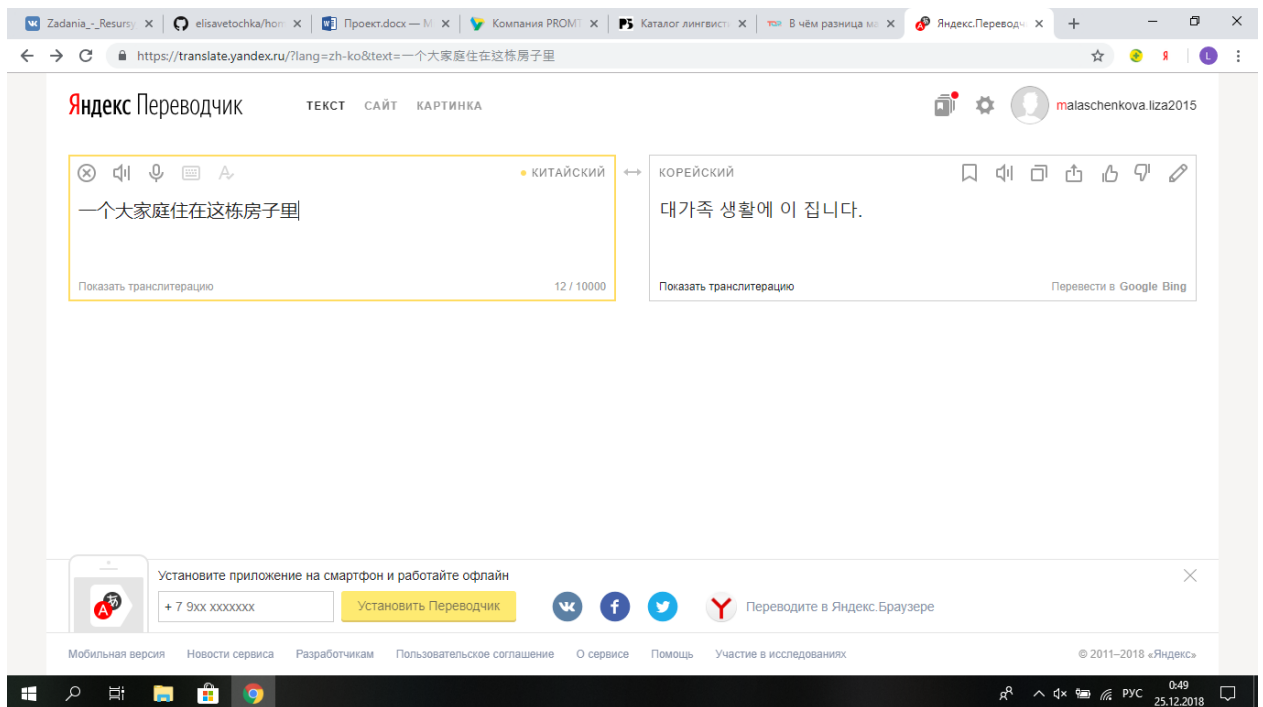
Переводите в Яндекс Браузере

© 2011–2018 «Яндекс»

Далее меняем местами русское и английское предложение. А вместо русского выбираем китайский.



Теперь меняем местами английский и китайский, а вместо английского выбираем корейский.





Yandex.Translate interface showing a translation from Korean to Xhosa. The URL is <https://translate.yandex.ru/?lang=ko-xh&text=대가족생활에이집니다>. The text being translated is "대가족 생활에 이 집니다." (Daegajok saenghwal-e i jipnida). The translation result is "Enkulu usapho abahlala kule ndlu." (Enkulu usapho abahlala kule ndlu.). The interface includes a search bar, a list of recent translations, and a footer with links to the mobile version, news, developer tools, and user agreement.

Yandex.Translate

ТЕКСТ САЙТ КАРТИНКА

malaschenkova.liza2015

КОРЕЙСКИЙ ↔ КОСА

대가족 생활에 이 집니다.

Enkulu usapho abahlala kule ndlu.

Показать транслитерацию 14 / 10000

Перевести в Google Bing

Установите приложение на смартфон и работайте офлайн

+7 9xx xxxxxxxx Установить Переводчик

Переводите в Яндекс.Браузере

Мобильная версия Новости сервиса Разработчикам Пользовательское соглашение О сервисе Помощь Участие в исследованиях © 2011–2018 «Яндекс»

Yandex.Translate interface showing a translation from Xhosa to Japanese. The URL is <https://translate.yandex.ru/?lang=xh-ja&text=Enkulu%20usapho%20abahlala%20kule%20ndlu>. The text being translated is "Enkulu usapho abahlala kule ndlu." (Enkulu usapho abahlala kule ndlu.). The translation result is "家に住んでいる偉大な家族。" (Ie ni sunde iru taishinaka no kaizoku.). The interface includes a search bar, a list of recent translations, and a footer with links to the mobile version, news, developer tools, and user agreement.

Yandex.Translate

ТЕКСТ САЙТ КАРТИНКА

malaschenkova.liza2015

КОСА ↔ ЯПОНСКИЙ

Enkulu usapho abahlala kule ndlu.

家に住んでいる偉大な家族。

Показать транслитерацию 33 / 10000

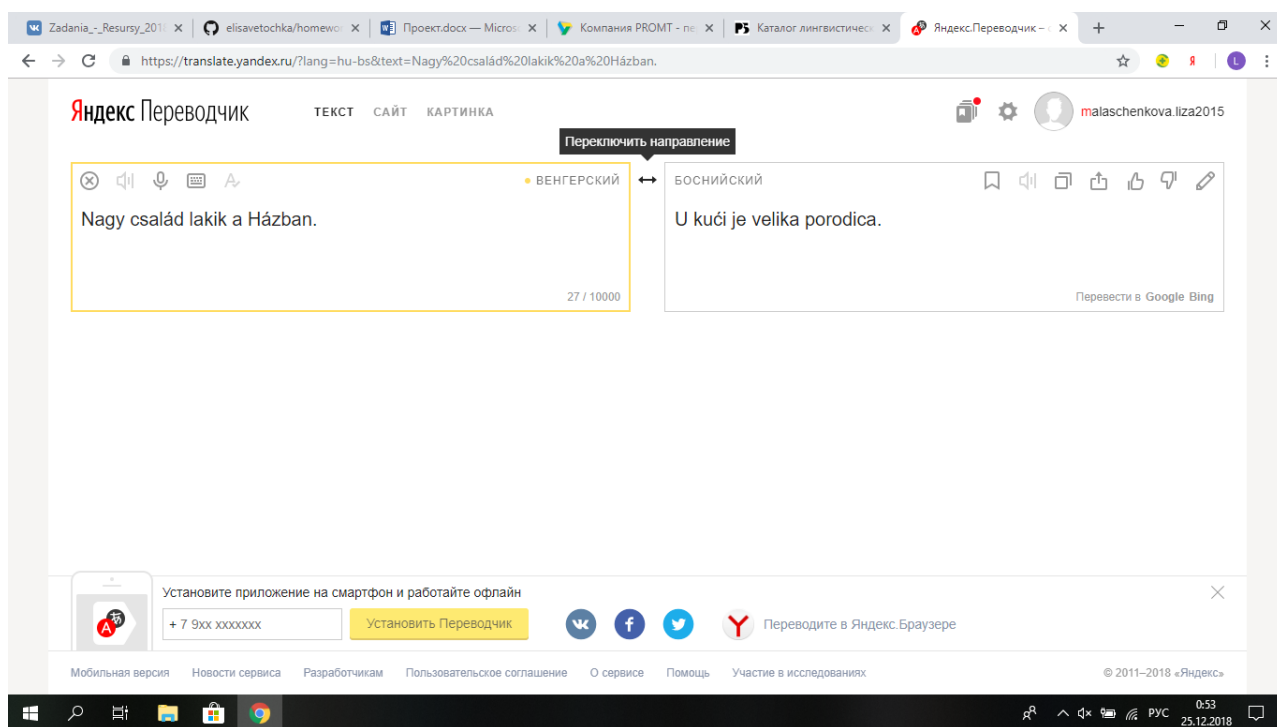
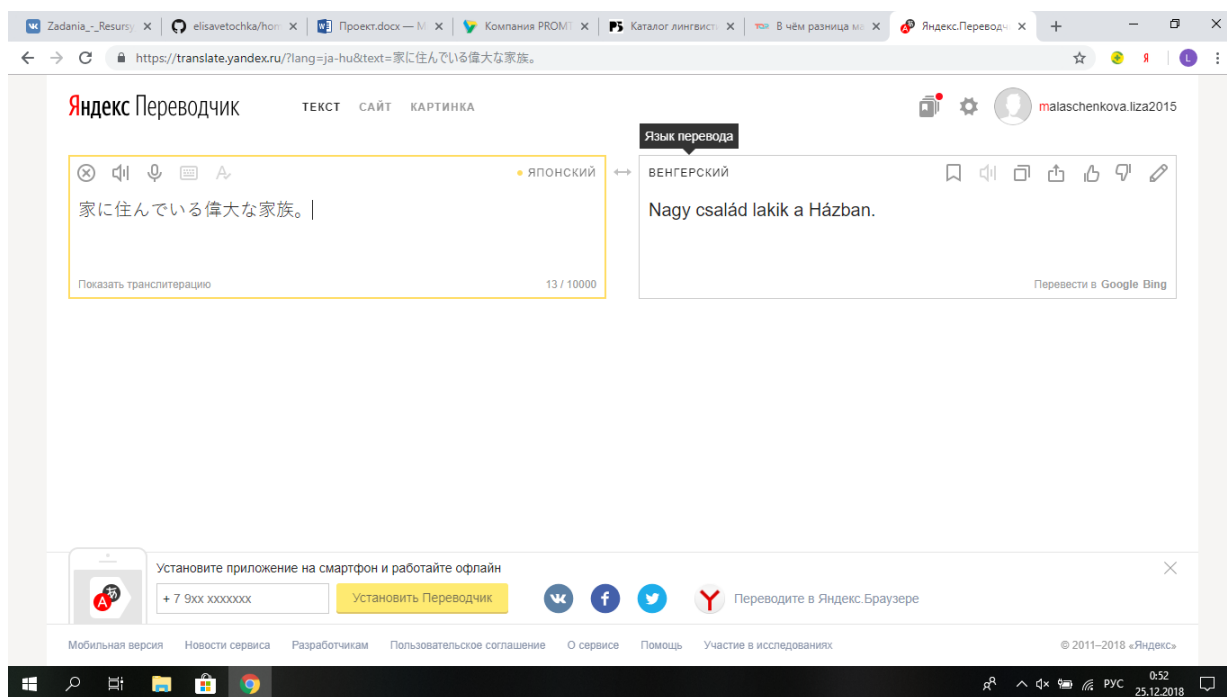
Перевести в Google Bing

Установите приложение на смартфон и работайте офлайн

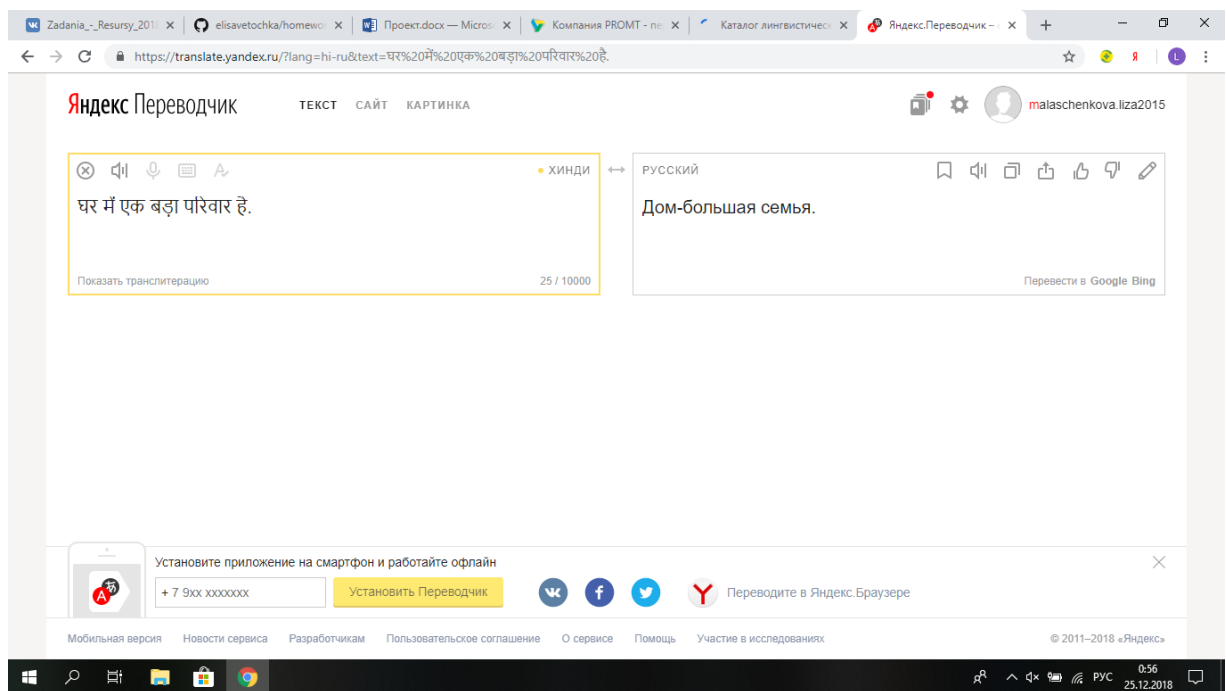
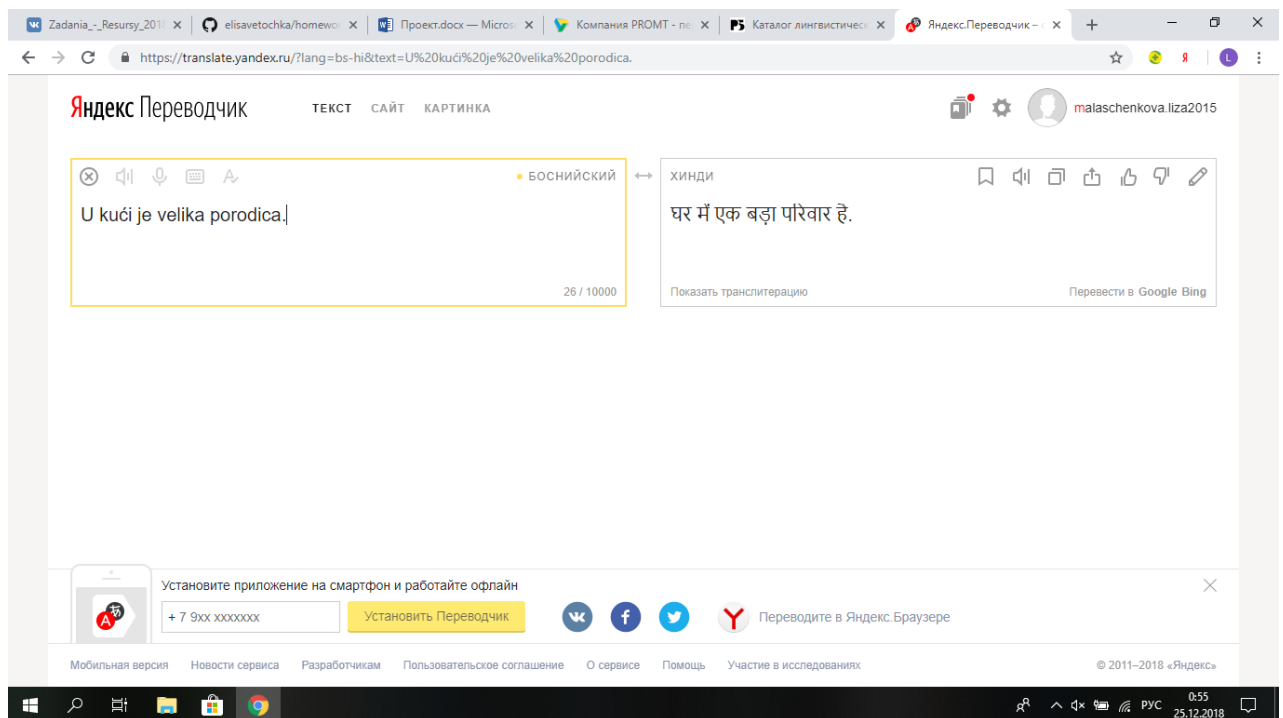
+7 9xx xxxxxxxx Установить Переводчик

Переводите в Яндекс.Браузере

Мобильная версия Новости сервиса Разработчикам Пользовательское соглашение О сервисе Помощь Участие в исследованиях © 2011–2018 «Яндекс»



Все эти манипуляции мы проводим для того, чтобы понять с какой точностью переводит яндекс слова с одного языка на другие, если в итоге мы вернемся к исходному языку.



Исходя из проделанной работы ,можно сделать вывод ,что не всегда машинный перевод является правильным и он не может запомнить с самого начала смысл вводимого предложения.

Тоже самое проделайте с предложением “На огороде крестьяне посадили много лука”  
Скрины проделанной работы скиньте в GitHub

## Система переводческой памяти

**Память переводов** (ПП, [англ. translation memory](#), ТМ, иногда также называемая «накопитель переводов») — [база данных](#), содержащая набор ранее [переведённых сегментов текста](#).

Одна запись в такой базе данных соответствует сегменту или «[единице перевода](#)», за которую обычно принимается одно [предложение](#) (реже — часть [сложносочинённого предложения](#) либо [абзац](#)). Если единица перевода исходного текста в точности совпадает с единицей перевода, хранящейся в базе, она может быть автоматически подставлена в перевод. Новый сегмент может также слегка отличаться от хранящегося в базе. Такой сегмент может быть также подставлен в перевод, но переводчик будет должен внести необходимые изменения.

Помимо ускорения процесса перевода повторяющихся фрагментов и изменений, внесенных в уже переведённые тексты, системы ПП также обеспечивают единообразие перевода терминологии в одинаковых фрагментах, что особенно важно при техническом переводе. С другой стороны, если переводчик регулярно подставляет в свой перевод точные соответствия, извлеченные из баз переводов, без контроля их использования в новом контексте, качество переведённого текста может ухудшиться.

В каждой конкретной системе ПП данные хранятся в своем собственном формате (текстовый формат в Wordfast, база данных Access в Deja Vu), но существует международный стандарт [TMX](#) ([англ. Translation Memory eXchange format](#)), который основан на XML и может генерироваться практически всеми системами ПП. Благодаря этому сделанные переводы можно использовать в разных приложениях, то есть переводчик, работающий с [OmegaT](#), может использовать ПП, созданную в [Trados](#), и наоборот.

Большинство систем ПП как минимум поддерживают создание и использование словарей пользователя, создание новых баз данных на основе [параллельных текстов](#) ([англ. alignment](#)), а также полуавтоматическое извлечение терминологии из оригинальных и параллельных текстов.

**В соответствии с обзорами использования систем ПП к наиболее популярным системам относятся:**

- memoQ
- XTM Cloud
- Memsource
- [Deja Vu](#)
- [OmegaT](#) (бесплатная система, распространяемая по лицензии [GNU GPL](#))
- [SDLX](#)<sup>[3]</sup>
- [Trados](#)
- [SmartCAT](#) (бесплатная программа со встроенной функцией распознавания PDF)
- [STAR Transit](#)
- [Wordfast](#) (реализована как набор макросов для [MS Word](#), новая версия 2009 года является отдельным приложением)

**Преимущества и недостатки**

### Преимущества

- Сокращение времени и объёма работы переводчика.
- Улучшение последовательности перевода, особенно при работе группы переводчиков над одним проектом.
- Увеличение прибыли за счёт увеличения производительности труда переводчика, группы переводчиков.
- Повышение качества услуг за счёт увеличения точности и единообразия перевода терминов, особенно в специализированных текстах.

### Недостатки

Может делать перевод более «сухим»; утрачивается сама суть текста, если перевод с использованием накопителя переводов выполняется переводчиком низкой квалификации.

- Часто отсутствует связь предлагаемого программой предложения/текста с соседними предложениями и с текстом в целом.
- Оригинал должен быть в электронном виде.
- Одна незамеченная ошибка может распространиться на весь проект.
- Необходимо обучение работе в самой программе, а при смене работы — возможно, и не один раз (если работодатели пользуются разными программами ТМ).
- Подходит не ко всем видам текстов.
- Высокая стоимость лицензионного ПО.

## **ТАБЛИЦА 2. РЕЗУЛЬТАТЫ АВТОМАТИЧЕСКОЙ ОЦЕНКИ МАШИННОГО ПЕРЕВОДА ВЫБОРОЧНОЙ СОВОКУПНОСТИ ИЗ КОРПУСА RAURAL ДЛЯ РАЗЛИЧНЫХ СИСТЕМ С ПОМОЩЬЮ МЕТРИКИ BLEU**

Система перевода	Значение BLEU
Гибридная система PROMT	29,2
RBMT-система PROMT	16,7
Статистическая система PROMT	27,3
Google Translate	15,2

### Задания

1. Для закрепления всего пройденного пройдите по ссылке

[https://docs.google.com/forms/d/e/1FAIpQLSdpUAUQyUnyOwr7b1b4cwhB\\_fpoVWRV-Hs3cf0Bzfbtq0ygDQ/viewform](https://docs.google.com/forms/d/e/1FAIpQLSdpUAUQyUnyOwr7b1b4cwhB_fpoVWRV-Hs3cf0Bzfbtq0ygDQ/viewform)

2. Также предлагаю вам провести небольшую исследовательскую работу. Вам нужно перевести стихотворение на русский язык на разных сайтах и

сравнить

перевод:

*Der Brief, den du geschrieben,  
er macht mich gar nicht bang;  
du willst mich nicht mehr lieben,  
aber dein Brief ist lang.  
Zwölf Seiten, eng und zierlich!  
Ein kleines Manuskript!  
Man schreibt nicht so ausführlich,  
wenn man den Abschied gibt.*

**3. Найдите фразеологизмы на разных языках мира и переведите их на русский язык, а также найдите их аналоги в русском языке.**

**Пример:**

be out of sorts — in bad humor	не в духе
--------------------------------	-----------

**4. Откройте видео на иностранном языке(конечно, который вы хорошо знаете) на YouTube и нажмите, чтобы отображались субтитры. Сделайте вывод о том , правильные или нет были субтитры.**

**5. Сравните перевод сериала “Ривердейл” на разных сайтах . Изменения проходят в некоторых словах или меняются даже имена?**

## **Полезные ссылки**

<https://ru.smartcat.ai/corp/cattools/>

<https://translate.google.com/?hl=ru>

<https://translate.yandex.ru/>

[https://www.isi.edu/research\\_groups/nlg/home](https://www.isi.edu/research_groups/nlg/home)

<https://www.multitran.ru/>

<https://www.translate.ru/>

<https://dict.leo.org/%D0%BD%D0%B5%D0%BC%D0%B5%D1%86%D0%BA%D0%B8%D0%B9-%D1%80%D1%83%D1%81%D1%81%D0%BA%D0%B8%D0%B9/>

<http://www.perevodov.net/>

## **Полезные ресурсы**

<https://www.tra-service.ru/shop>

<https://www.wordfast.net/>

<https://context.reverso.net/%D0%BF%D0%B5%D1%80%D0%B5%D0%B2%D0%BE%D0%B4/%D1%80%D1%83%D1%81%D1%81%D0%BA%D0%B8%D0%B9-%D0%B0%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9/%D1%84%D1%80%D0%B0%D0%B7%D0%B5%D0%BE%D0%BB%D0%BE%D0%B3%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B8%D0%B9+%D1%81%D0%BB%D0%BE%D0%B2%D0%B0%D1%80%D1%8C>

