


# Машинный перевод

## План:

Введение

- История машинного перевода (МП)
- Классификация систем МП
- Системы МП
- Электронные словари
- Онлайнновый перевод
- Как улучшить качество перевода

# **Машинный перевод**



**Переводом называется  
процесс и результат  
создания на основе**

**исходного текста на одном  
языке равноценного ему в  
коммуникативном  
отношении текста на другом  
языке.**

# **Коммуникативная эквивалентность**

**- качество текста перевода,  
которое позволяет ему  
выступать в процессе  
общения носителей разных  
языков в качестве  
полноправной замены  
исходного текста  
(оригинала) в сфере  
действия языка перевода**

# Три основных требования коммуникативной эквивалентности

- текст перевода должен в возможно более полном объеме передавать содержание оригинала;
- текст перевода должен соответствовать нормам языка перевода;
- текст перевода должен быть примерно сопоставим с оригиналом по своему объему.

# Перевод может осуществляться:

- ✓ с одного языка на другой – неродной, родственный, близкородственный;
- ✓ с литературного языка на его диалект и наоборот, или с диалекта одного языка на другой литературный язык;
- ✓ с языка древнего периода на данный язык в его современном состоянии (например, с древнерусского языка на современный русский, со староанглийского на современный английский и т.д.).

# **Машинный перевод**

**т.е. перевод, строящийся на  
использовании машинной  
определенных и постоянных  
для данного вида материала  
соответствий между  
словами и  
грамматическими  
явлениями разных языков**

# Две основные группы словарей

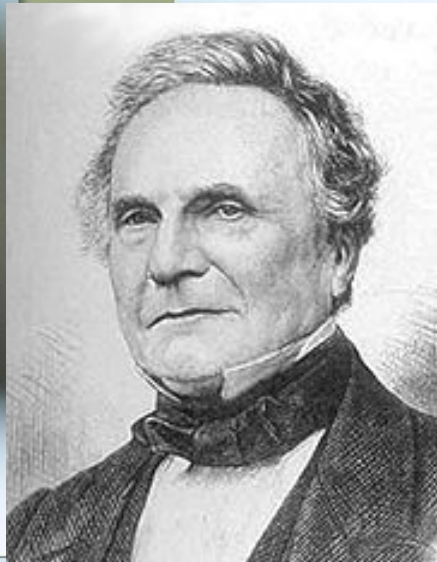
- ❖ **системы машинного перевода (*machine translation system*)**
- ❖ **электронные словари (*electronic dictionary*)**



# I. История машинного перевода



# Чарльз Бэббидж



Чарльз Бэббидж родился 26 декабря 1791 г. на юго-западе Англии в маленьком городке Тотнес, в графстве Девоншир. Скончался около полуночи 18 октября 1871 г. на руках у сына, не дожив до своего 80-летия двух месяцев. Бэббидж похоронен на кладбище Кензел Грин 24 октября.

Выдвинув концепцию универсальной цифровой вычислительной машины с программным управлением, Бэббидж на много лет опередил свое время не только с точки зрения идеи, но и с позиций возможности ее технического осуществления.

# Этапы развития МП


**1. 40-е: первые  
шаги**



# Уоррен Уивер



1947 - дата рождения машинного перевода как научного направления. Уоррен Уивер (Warren Weaver), руководитель отделения естественных наук Рокфеллеровского фонда, написал письмо Норберту Винеру, в котором рассматривал задачу перевода текстов с одних языков на другие как еще одну область применения техники дешифрования.



**"I have a text in front of me which is written in Russian but I am going to pretend that it is really written in English and that it has been coded in some strange symbols. All I need to do is strip off the code in order to retrieve the information contained in the text"**

**Концепция *interlingva*: стадия передачи информации разделена на два этапа; на первом этапе исходное предложение переводится на язык-посредник (созданный на базе упрощенного английского языка), а затем результат этого перевода представляется средствами выходного языка.**

## 2. 50-е: первое разочарование

“John was looking for his toy box. Finally he found it. The box was in the *pen*. John was very happy”.

(“Джон искал свою игрушечную коробку. Наконец он ее нашел. Коробка была в *манеже*. Джон был очень счастлив”).

## 3. 60-е: низкий старт

- США университет Brigham Young University в Прово, штат Юта (ранние коммерческие системы WEIDNER и ALPS);
- в Канаде - группы исследователей, в числе которых TAUM в Монреале с ее системой МЕТЕО;
- в Европе — группы GENA (Гренобль) и SUSY (Саарбрюкен).
- работа в этой области отечественных лингвистов, таких, как И.А.Мельчук и Ю.Д. Апресян (Москва)

## 4. 70-80-е: новый импульс

- Комиссия Европейских общин (СЕС) покупает англо-французскую версию Systran, а также систему перевода с русского на английский.
- СЕС заказывает разработку французско-английской и итальянско-английской версий.
- Проекты EUROTRA, основанного на разработках групп SUSY и GETA.
- Расширение деятельности по созданию систем МП в Японии (главным образом основанных на технологии interlingva, разработанной Уивером в конце 40-х гг.).
- В США Панамериканская организация здравоохранения (РАНО) заказывает разработку испанско-английского направления (система SPANAM).
- ВВС США финансируют разработку системы МП в Лингвистическом исследовательском центре при Техасском университете в Остине.
- Группа TAUM в Канаде достигает заметных успехов в разработке своей системы МЕТЕО.

## 5. От 90-х к XXI веку

- Systran, IBM, L&H (Lernout & Hauspie), Language Engineering Corporation, Transparent Language, Nova Incorporated, Trident Software, Atril, LingoWare, Ata Software, Lingvistica b.v.
- PROMT (PROgrammer's Machine Translation)



## II. Классификация систем машинного перевода по Лари Чайлдсу

- - *полностью автоматический перевод;*
- - *автоматизированный машинный перевод при участии человека;*
- - *перевод, осуществляемый человеком с использованием компьютера.*

# Translation Memory (TM)

## Список наиболее известных систем TM:

- - Transit швейцарской фирмы Star,
- - Trados (США),
- - Translation Manager от IBM,
- - Eurolang Optimizer французской фирмы LANT,
- - DejaVu от ATRIL (США),
- - WordFisher (Венгрия).

# Общие функции систем ТМ :

- ❖ - Функция сопоставления (Alignment). Одно из преимуществ систем ТМ – это возможность использования уже переведенных материалов по данной тематике.
- ❖ - Наличие фильтров импорта – экспорта. Это свойство обеспечивает совместимость систем ТМ с множеством текстовых процессоров и издательских систем и дает переводчику относительную независимость от заказчика.
- ❖ - Механизм поиска нечетких или полных совпадений.
- ❖ - Поддержка тематических словарей.
- ❖ - Средства поиска фрагментов текста.

# Системы машинного перевода (МП)

## Система машинного перевода

включает в себя **двухязычные словари**, снабженные необходимой грамматической информацией (морфологической, синтаксической и семантической) для обеспечения передачи эквивалентных, вариантных и трансформационных переводных соответствий, а также алгоритмические средства грамматического анализа, реализующие какую-либо из принятых для автоматической переработки текста формальных грамматик.

# Последовательность формальных операций в системе МП:

- 1. На первом этапе** осуществляется ввод текста и поиск входных словоформ во входном словаре (словаре языка, с которого производится перевод) с сопутствующим морфологическим анализом, в ходе которого устанавливается принадлежность данной словоформы к определенной лексеме (слову как единице словаря).

# Последовательность формальных операций в системе МП:

**2. Следующий этап** включает в себя перевод идиоматических словосочетаний, фразеологических единств или штампов данной предметной области (например, при англо-русском переводе обороты типа *in case of*, *in accordance with* получают единый цифровой эквивалент и исключаются из дальнейшего грамматического анализа); определение основных грамматических (морфологических, синтаксических, семантических и лексических) характеристик элементов входного текста, производимое в рамках входного языка; разрешение омографии (конверсионной омонимии словоформ – скажем, англ. *round* может быть существительным, прилагательным, наречием, глаголом или же предлогом); лексический анализ и перевод лексем.

# Последовательность формальных операций в системе МП:

3. **Окончательный грамматический анализ**, в ходе которого доопределяется необходимая грамматическая информация с учетом данных выходного языка (например, при русских существительных типа *сани*, *ножницы* глагол должен стоять в форме множественного числа, несмотря на то, что в оригинале может быть и единственное число).

# Последовательность формальных операций в системе МП:

**4. Синтез выходных словоформ и предложения в целом на ВЫХОДНОМ ЯЗЫКЕ.**



## Две концепции развития систем МП:

1. **Модель «большого словаря со сложной структурой»**, которая заложена в большинство современных программ-переводчиков;
2. **Модель «смысл-текст»**, впервые сформулированная А.А. Ляпуновым, но пока что не реализована ни в одном коммерческом продукте.

# Системы машинного перевода:

- ❖ PROMT 2000/XT компании PROMT;
- ❖ Retrans Vista компаний Vista и Advantis;
- ❖ Сократ – набор программ компании Арсеналь.

# **Применение систем МП не оправдано, поскольку:**

- Системы МП не дают приемлемого качества выходного текста.**
- Системы МП не гарантируют соблюдения единства терминологии, особенно при работе коллектива переводчиков над большим проектом.**

# Система МП PROMT XT.

**Первая проблема** - проблема создания больших словарей для систем.

**Вторая проблема** - научить систему распознавать устойчивые обороты.

**Третья проблема** - записать все правила написания предложений в виде программы.

# Система МП PROMT XT.

*"You are given 12 identical-looking coins, one of which is counterfeit and weighs slightly more or less (you don't know which) than the others. You are given a beam balance which lets you put the same number of coins on each side and observe which side (if either) is heavier. How can you identify the counterfeit and tell whether it is heavy or light, in 3 weighings?"*

# Система МП PROMT ХТ.

*"У вас есть 12 одинаковых по виду монет, одна из которых - фальшивая и весит немного больше или меньше, чем остальные (вы не знаете, какая именно). Имеются рычажные весы, на чаши которых вы можете класть равное число монет и смотреть, какая из чаш перевесила (или весы остались в равновесии). Как за 3 взвешивания определить фальшивую монету и узнать, легче она или тяжелее остальных?"*

# Система МП PROMT XT.

*"Вам дают 12 идентично-выглядящих монет, одна из которых - подделка и весит немного более или менее (Вы не знаете который) чем другие. Вам дают баланс луча, который позволяет Вам помещать то же самое число {номер} монет на каждой стороне и наблюдать {соблюдать}, какая сторона (если любой) более тяжел. Как Вы можете идентифицировать подделку и сказать, тяжело ли это или легко, в 3 взвешиваниях?"*

# Система МП PROMT ХТ.

**"*if any* - если это имеет место**

At the start of every month I have to send him an account of my earnings, if any. — В начале каждого месяца я должен посылать ему отчет о моих заработках, если таковые имелись."



# ПЕРЕВОДЧИК ДЛЯ ОФИСА

- **PROMT 98 или Stylus**  
(фирма ПРОМТ)
- **ПАРС (фирма Лингвистика**  
**93»»)**

# КТО В ДОМЕ ПЕРЕВОДЧИК?

- ❖ **Magic Gooddy (от компании ПРОМТ)**
- ❖ **"Сократ Персональный 3.0.3«**
- ❖ **"Коперник"**

# **ЭЛЕКТРОННЫЕ СЛОВАРИ**

**DIC digital integrated circuit**

**LINGVO**

# ЭС

- 1. По используемой операционной системе.**
- 2. По способу загрузки.**
- 3. По количеству подключаемых словарных баз.**
- 4. По возможностям расширения словарной базы.**
- 5. По режиму перевода.**

# ЭЛЕКТРОННЫЕ СЛОВАРИ

- ❖ **Lingvo** компании Abbyu
- ❖ **МультиЛекс**, разработанный фирмой МедиаЛингва.

# Преимущества электронных словарей

- ❑ Специфика словарного ответа в том, что он дает весьма разнообразную информацию о слове или словосочетании, а не просто переводное соответствие.
- ❑ Содержат транскрипцию, но и могут произносить слова.
- ❑ Одновременный поиск не только по названию словарной статьи, но и по всему огромному объему словарей

# V. Онлайн-перевод информации в Интернете

Две технологии машинного перевода: **традиционная, на основе правил (rule-based machine translation)**, и


**статистическая (statisticalbased machine translation)**

# GOOGLE TRANSLATE

1. Принцип работы основан на статистическом вычислении вероятности совпадений фраз из исходного текста с фразами, которые хранятся в базе системы перевода.
2. Для функционирования нужна база текстов на исходном языке и их переводов (база параллельных текстов), которую программа анализирует на основе определенных алгоритмов.
3. Плюсом статистической технологии является отсутствие необходимости разработки лингвистических алгоритмов в принципе.
4. Минусом статистических систем является отсутствие учета грамматических правил входного и выходного языков.



# Перевод по правилам




- технологии перевода на основе лингвистических алгоритмов. По такому принципу работают системы PROMT, Systran, Linguatrec

# Исходный текст ВВС




Flexible paper batteries could meet the energy demands of the next generation of gadgets, says a team of researchers.

# Перевод Google




Гибкие батареи документ может удовлетворить энергетические потребности следующего поколения гаджеты, утверждает группа исследователей.

# Перевод PROMT 8.0



Гибкие бумажные батареи могли удовлетворить требованиям энергии следующего поколения устройств, говорит команда исследователей.

# Перевод PROMT 8.0 после настройки



Гибкие бумажные батареи могли удовлетворить требованиям энергии следующего поколения устройств, говорит команда исследователей.

# VI. КАК УЛУЧШИТЬ КАЧЕСТВО ПЕРЕВОДА?

- 1) Исход работы в значительной мере решается еще до ее начала.
- 2) Бойтесь опечаток!
- 3) Хороший словарь – половина успеха.
- 4) Строим "пирамиду".
- 5) Не надо спешить!
- 6) Различайте общее и частное!

# Параметры МШ



- **оперативность;**



- **гибкость;**



- **скорость;**

- **точность.**

# ЛИТЕРАТУРА

- ✓ Ванников Ю.В. Языковая сложность текста как фактор трудности перевода (Методическое пособие). – М.: Всесоюзный центр переводов, 1988.
- ✓ Вейзе А. А., Мирончиков И. К. Перевод технического текста с английского языка на русский. – Мн.: МГЛУ, 1995.
- ✓ Васильев А. (Компьютер на месте переводчика). // Подводная лодка. – 1998, № 6
- ✓ Винокуров А.А., Чуканов В.О. Новый метод оценки машинного перевода.// Информационные технологии и системы. Hardware Software Security. Тенденции и перспективы – Сборник статей: М., Международная академия информатизации, 1997.
- ✓ Гвишиани Н.Б. Язык научного общения (вопросы методологии). –М.: Высш. шк.; 1986.
- ✓ Марчук Ю.Н. основы компьютерной лингвистики. Учебное пособие. – М.:Изд-во МГОУ, 2002. – 236 с.



# ЛИТЕРАТУРА

<http://www.promt.ru>

<http://www.socrat.ru>

<http://www.translate.ru>

<http://www.a-z.ru/person/belonogov/index.htm#I0>

<http://www.computerra.ru/> ("Что могут словари?" Денис Зельцер);

[http://www.lingvoda.ru/transforum/articles/pdf/selegey\\_a1.pdf](http://www.lingvoda.ru/transforum/articles/pdf/selegey_a1.pdf)  
("Электронные словари и компьютерная лексикография", Владимир Селегей, компания АБВУ);

[http://www.promt.ru/mtw/articles/article\\_Sokolova.phtml](http://www.promt.ru/mtw/articles/article_Sokolova.phtml)  
("Как переводит компьютер", Автор: Соколова Светлана);

<http://www.promt.ru:8000/mtw/class.phtml>;

<http://www.promt.ru:8000/mtw/developer.phtml>;

[www.multilex.ru](http://www.multilex.ru);



**СПАСИБО ЗА  
ВНИМАНИЕ!**