

Математические модели в лингвистике

8. Машинный перевод: обзор методов и оценка качества

Мати Пентус, Александр Пиперски,
Алексей Сорокин

МГУ, межфакультетский курс,
осенний семестр 2017–2018 учебного года

BRANCH OF PHYSICAL CHEMISTRY BIOLOGY

Squirrel institute

phone: (495) 632-78-71

fax: (495) 632-78-71 , (4967) 73-24-56

e-mail address: protres@vega.protres.ru

www-page: <http://www.protres.ru>

Deputy director

dcs **Efimov** Aleksandr Vasil'evich

phone: (4967)73-28-06

e-mail address: efimov@protres.ru

dbS **Fedorov** Oleg Vasil'evich

phone: (4967)73-09-15 , (27)73-09-15

e-mail address: fedorov@vega.protres.ru

Deputy director

Shaklunov Vitalii Nikolaevich

phone: (4967)73-07-22 , (27)73-07-22

e-mail address: vitali@vega.protres.ru

RUS | ENG



Attribute search

- Organizations
- Persons

RAS organizations structure

Last modification: 11.03.2010

Resources search

119991 Москва, Ленинский просп., 14
Телефон: (495) 938-0309 (Справ. бюро); Факс: (495) 938-1844 (Лен.пр.32а), (495)
954-3320 (Лен.пр.14)

[На главную страницу](#)
[В начало страницы](#)

Аннотация МФК (2015)

Russian English Albanian Detect language ▾



Greek English Albanian ▾

Translate

Лингвистика — это не только наука о том, как грамотно писать и говорить на родном и иностранных языках. Лингвистов гораздо больше интересует, как устроена языковая способность человека, как развиваются и в каких отношениях между собой находятся языки, а также какими свойствами обладают тексты на естественных языках. ✕

Ä Py ▾

Linguistics - it is not only the science of how to correctly write and speak in the native and foreign languages. Linguists are more interested in how to construct the linguistic ability of the person how to develop and what are the relations between a language and what properties have a natural language.



Wrong?

Аннотация МФК (2017)

русский английский китайский Определить язык ▾



английский сербский русский ▾

Перевести

Лингвистика — это не только наука о том, как грамотно писать и говорить на родном и иностранном языках. Лингвистов гораздо больше интересует, как устроена языковая способность человека, как развиваются и в каких отношениях между собой находятся языки, а также какими свойствами обладают тексты на естественных языках. ✕



317/5000

Linguistics is not only a science of how to write and write literally in native and foreign languages. Linguists are much more interested in how the language ability of a person is arranged, how languages are developed and in what relations are languages, and also what properties are available in natural language texts.



Предыстория машинного перевода

Предыстория машинного перевода

- ▶ Машинный перевод (МП) обсуждается с XVII века

Предыстория машинного перевода

- ▶ Машинный перевод (МП) обсуждается с XVII века
- ▶ Готтфрид Вильгельм Лейбниц (1646–1716)

Предыстория машинного перевода

- ▶ Машинный перевод (МП) обсуждается с XVII века
- ▶ Готтфрид Вильгельм Лейбниц (1646–1716)
- ▶ Рене Декарт (1596–1650)

Лейбниц: универсальный язык

Лейбниц: универсальный язык

- ▶ *characteristica universalis* — универсальный пиктографический язык для выражения мыслей (в первую очередь научных)

Лейбниц: универсальный язык

- ▶ *characteristica universalis* — универсальный пиктографический язык для выражения мыслей (в первую очередь научных)
- ▶ Даёт возможность перевода:
язык 1 \Leftrightarrow *characteristica universalis* \Leftrightarrow язык 2

Лейбниц: универсальный язык

- ▶ *characteristica universalis* — универсальный пиктографический язык для выражения мыслей (в первую очередь научных)
- ▶ Даёт возможность перевода:
язык 1 \Leftrightarrow *characteristica universalis* \Leftrightarrow язык 2
- ▶ К концу жизни сам Лейбниц разочаровался в идее универсального языка

История МП

История МП

- ▶ XX век: развитие компьютеров \Rightarrow
практическая возможность машинного
перевода

История МП

- ▶ XX век: развитие компьютеров \Rightarrow практическая возможность машинного перевода
- ▶ Первые реальные попытки машинного перевода — 1950-е годы

История МП

- ▶ XX век: развитие компьютеров \Rightarrow практическая возможность машинного перевода
- ▶ Первые реальные попытки машинного перевода — 1950-е годы
- ▶ У истоков машинного перевода стояли военные нужды

История МП

История МП

- ▶ Развитие МП началось почти одновременно с началом Холодной войны

История МП

- ▶ Развитие МП началось почти одновременно с началом Холодной войны
- ▶ Первая языковая пара: русский \Leftrightarrow английский

Джорджтаунский эксперимент

Джорджтаунский эксперимент

- ▶ “Russian is turned into English by a fast electronic translator” (New York Times, 8 января 1954 года, **стр. 1**)

Джорджтаунский эксперимент

- ▶ “Russian is turned into English by a fast electronic translator” (New York Times, 8 января 1954 года, **стр. 1**)
- ▶ Компания IBM и Институт языков и лингвистики Джорджтаунского университета (Вашингтон)

Джорджтаунский эксперимент

Джорджтаунский эксперимент

- ▶ Более 60 предложений (в основном по органической химии)

Джорджтаунский эксперимент

- ▶ Более 60 предложений (в основном по органической химии)
- ▶ 250 морфем (корней и окончаний)

Джорджтаунский эксперимент

- ▶ Более 60 предложений (в основном по органической химии)
- ▶ 250 морфем (корней и окончаний)
- ▶ 6 синтаксических правил

Джорджтаунский эксперимент

Джорджтаунский эксперимент

- ▶ Starch is produced by mechanical method from potatoes

Джорджтаунский эксперимент

- ▶ Starch is produced by mechanical method from potatoes
- ▶ Мезhdunarodnoye ponyimaniye yavlyayetsya vazhnim faktorom v ryeshyeniyyi polyityichyeskix voprosov = International understanding constitutes an important factor in decision of political questions.

Джорджтаунский эксперимент

- ▶ Starch is produced by mechanical method from potatoes
- ▶ Мyezhdunarodnoye ponyimaniye yavlyayetsya vazhnim faktorom v ryeshyeniyyi polyityichyeskix voprosov = International understanding constitutes an important factor in decision of political questions.
- ▶ Качество перевода?

Джорджтаунский эксперимент

Джорджтаунский эксперимент

- ▶ Эксперимент вселил в широкие круги веру в быстрый успех МП

Джорджтаунский эксперимент

- ▶ Эксперимент вселил в широкие круги веру в быстрый успех МП
- ▶ Леон Достерт (один из создателей системы): реальный МП — через 3–5 лет

Джорджтаунский эксперимент

- ▶ Эксперимент вселил в широкие круги веру в быстрый успех МП
- ▶ Леон Достерт (один из создателей системы): реальный МП — через 3–5 лет
- ▶ Вслед за США, работа над МП началась и в СССР

Статистика МП (1959)

Статистика МП (1959)

[Bar-Hillel 1960]:

- ▶ 11 групп в США, ок. 150 человек

Статистика МП (1959)

[Bar-Hillel 1960]:

- ▶ 11 групп в США, ок. 150 человек
- ▶ 7 групп в СССР, ок. 300 человек

Статистика МП (1959)

[Bar-Hillel 1960]:

- ▶ 11 групп в США, ок. 150 человек
- ▶ 7 групп в СССР, ок. 300 человек
- ▶ 2 группы в Великобритании, ок. 25 человек

Статистика МП (1959)

[Bar-Hillel 1960]:

- ▶ 11 групп в США, ок. 150 человек
- ▶ 7 групп в СССР, ок. 300 человек
- ▶ 2 группы в Великобритании, ок. 25 человек
- ▶ 1 группа в Италии, ? человек

Статистика МП (1959)

[Bar-Hillel 1960]:

- ▶ 11 групп в США, ок. 150 человек
- ▶ 7 групп в СССР, ок. 300 человек
- ▶ 2 группы в Великобритании, ок. 25 человек
- ▶ 1 группа в Италии, ? человек
- ▶ 1 группа в Израиле, 4 человека

WHEN A USER TAKES A PHOTO,
THE APP SHOULD CHECK WHETHER
THEY'RE IN A NATIONAL PARK...

SURE, EASY GIS LOOKUP.
GIMME A FEW HOURS.

... AND CHECK WHETHER
THE PHOTO IS OF A BIRD.

I'LL NEED A RESEARCH
TEAM AND FIVE YEARS.



IN CS, IT CAN BE HARD TO EXPLAIN
THE DIFFERENCE BETWEEN THE EASY
AND THE VIRTUALLY IMPOSSIBLE.

Правильный МП

- ▶ Компьютерная лингвистика–1:
моделирование человеческого перевода

Правильный МП

- ▶ Компьютерная лингвистика–1:
моделирование человеческого перевода
- ▶ Лингвистический анализ текста +
синтез текста с учётом
лингвистической информации

Правильный МП

- ▶ Компьютерная лингвистика–1:
моделирование человеческого перевода
- ▶ Лингвистический анализ текста +
синтез текста с учётом
лингвистической информации
- ▶ Проблемы: см. лекцию № 2

Правильный МП и теория языка

- ▶ Некоторые проблемы, ставшие явными при разработке МП, стали обсуждаться в теории языка

Правильный МП и теория языка

- ▶ Некоторые проблемы, ставшие явными при разработке МП, стали обсуждаться в теории языка
- ▶ Пример: **лексические функции** (А. К. Жолковский, И. А. Мельчук)

Лексические функции

Лексические функции

[Фролова (без года)]:

- ▶ *People who have pneumonia*

Лексические функции

[Фролова (без года)]:

- ▶ *People who have pneumonia*
- ▶ ? *Люди, которые имеют пневмонию*

Лексические функции

[Фролова (без года)]:

- ▶ *People who have pneumonia*
- ▶ ? *Люди, которые имеют пневмонию*
- ▶ ? *Люди, у которых есть пневмония*

Лексические функции

[Фролова (без года)]:

- ▶ *People who have pneumonia*
- ▶ ? *Люди, которые имеют пневмонию*
- ▶ ? *Люди, у которых есть пневмония*
- ▶ *Люди, которые болеют пневмонией*

Лексические функции

Лексические функции

[Фролова (без года)]:

- ▶ *Schizophrenia is a severe illness*

Лексические функции

[Фролова (без года)]:

- ▶ *Schizophrenia is a severe illness*
- ▶ ? *Шизофрения — суровая болезнь*

Лексические функции

[Фролова (без года)]:

- ▶ *Schizophrenia is a severe illness*
- ▶ ? *Шизофрения — суровая болезнь*
- ▶ *Шизофрения — тяжёлая болезнь*

Лексические функции

- ▶ Лексические функции (ЛФ) моделируют выражение частотных типовых значений

Лексические функции

- ▶ Лексические функции (ЛФ) моделируют выражение частотных типовых значений
- ▶ ЛФ ставит в соответствие слову W_0 способ выражения соответствующего значения в данном языке

Лексические функции: примеры

- ▶ $\text{Magn}(\textit{illness}) = \textit{severe}$

Лексические функции: примеры

- ▶ $\text{Magn}(\textit{illness}) = \textit{severe}$
- ▶ $\text{Magn}(\textit{болезнь}) = \textit{тяжёлая}$

Лексические функции: примеры

- ▶ $\text{Magn}(\textit{illness}) = \textit{severe}$
- ▶ $\text{Magn}(\textit{болезнь}) = \textit{тяжёлая}$
- ▶ $\text{Oper}_1(\textit{pneumonia}) = \textit{have}$

Лексические функции: примеры

- ▶ $\text{Magn}(\textit{illness}) = \textit{severe}$
- ▶ $\text{Magn}(\text{болезнь}) = \text{тяжёлая}$
- ▶ $\text{Oper}_1(\textit{pneumonia}) = \textit{have}$
- ▶ $\text{Oper}_1(\text{пневмония}) = \text{болеть (+ твор.)}$

Лексические функции в переводе

- ▶ *Schizophrenia is a severe illness*

Лексические функции в переводе

- ▶ *Schizophrenia is a severe illness*
- ▶ *Schizophrenia is a Magn(illness) illness*

Лексические функции в переводе

- ▶ *Schizophrenia is a severe illness*
- ▶ *Schizophrenia is a Magn(illness) illness*
- ▶ *Шизофрения — Magn(болезнь) болезнь*

Лексические функции в переводе

- ▶ *Schizophrenia is a severe illness*
- ▶ *Schizophrenia is a Magn(illness) illness*
- ▶ *Шизофрения — Magn(болезнь) болезнь*
- ▶ *Шизофрения — тяжёлая болезнь*

Лексическая функция Magn

- ▶ $\text{Magn}(\text{rain}) = \text{heavy, cats and dogs}$

Лексическая функция Magn

- ▶ $\text{Magn}(\text{rain}) = \textit{heavy, cats and dogs}$
- ▶ $\text{Magn}(\text{дождь}) = \textit{сильный, проливной}$

Лексическая функция Magn

- ▶ $\text{Magn}(\text{rain}) = \textit{heavy, cats and dogs}$
- ▶ $\text{Magn}(\text{дождь}) = \textit{сильный, проливной}$
- ▶ NB: лексическая функция ставит в соответствие своему аргументу не конкретное слово, а функцию вероятности

Источники знаний о лексических функциях

- ▶ В правилочном МП значения лексических функций отражаются в словаре (напр., [Мельчук, Жолковский 1984])

Источники знаний о лексических функциях

- ▶ В правилочном МП значения лексических функций отражаются в словаре (напр., [Мельчук, Жолковский 1984])
- ▶ Можно пытаться автоматически извлечь эту информацию из больших собраний текстов

Лексические функции: *аплодисменты* */ applause I*

Лексические функции: аплодисменты / *applause* II

- ▶ Magn: бурный / продолжительный / восторженный / громкий || *heavy / lengthy / prolonged / loud / thunderous / ringing / terrific*
- ▶ AntiMagn: сдержанный / скупой / редкий / жидкий || *light / weak / thin*
- ▶ Ver: заслуженный || *well-merited*

Лексические функции: аплодисменты / *applause* III

- ▶ Labor₁₂: *встречать / награждать* || *meet / greet / hail*
- ▶ Oper₁: Ø || *give*
- ▶ Oper₂: *вызывать* || *draw / win / get*
- ▶ IncerFunc₀: *раздаваться / грянуть / вспыхивать* || *break out*

Лексические функции: аплодисменты / *applause* IV

- ▶ FinFunc₀: *с молчать / стихать / прекращаться* || *subside / die out*

Правильный МП

- ▶ Трудозатраты не соответствуют качеству

Правильный МП

- ▶ Трудозатраты не соответствуют качеству
- ▶ PROMT: Linguistics — it not only science how competently to write and speak on native and foreign languages. Interests linguists much more as language ability of the person is arranged as develop and in what relations among themselves there are languages, and also what properties texts in natural languages possess.

МП на основе примеров

МП на основе примеров

- ▶ 1980-е годы: если есть готовые образцы переводов, можно заменять в них варьирующиеся части

МП на основе примеров

- ▶ 1980-е годы: если есть готовые образцы переводов, можно заменять в них варьирующиеся части
- ▶ *Сколько стоит этот красный зонтик? =*
= How much is that red umbrella?

МП на основе примеров

- ▶ 1980-е годы: если есть готовые образцы переводов, можно заменять в них варьирующиеся части
- ▶ *Сколько стоит этот красный зонтик? =
= How much is that red umbrella?*
- ▶ *Сколько стоит этот зелёный чемодан? =
= How much is that green suitcase?*

Статистический МП: истоки

Статистический МП: истоки

- ▶ Уоррен Уивер (1949): “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.”

Статистический МП

Статистический МП

- ▶ 1990-е годы: развитие статистического МП

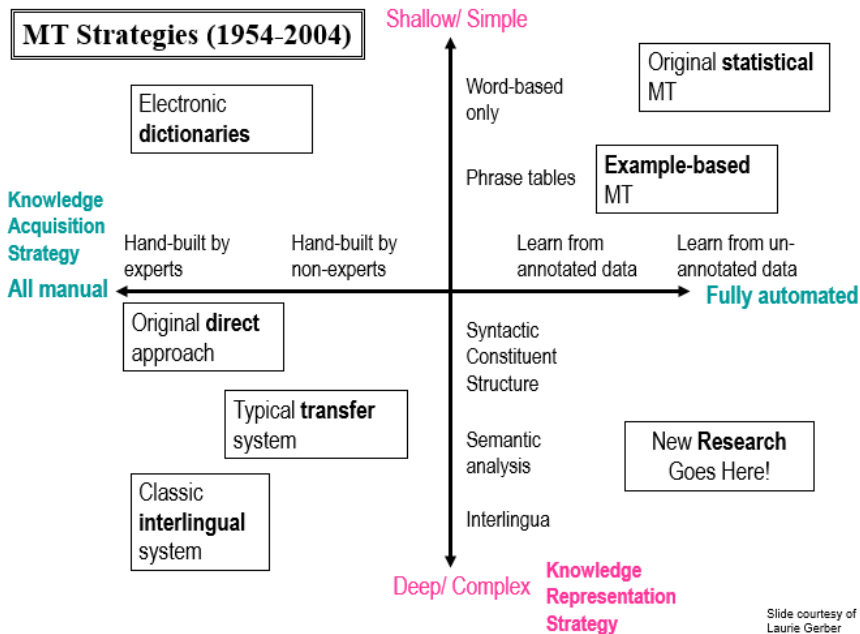
Статистический МП

- ▶ 1990-е годы: развитие статистического МП
- ▶ Компьютерная лингвистика–2

Статистический МП

- ▶ 1990-е годы: развитие статистического МП
- ▶ Компьютерная лингвистика–2
- ▶ Если есть корпус текстов на языке А и их переводных эквивалентов на языке В, можно искать соответствующие друг другу фрагменты без знания языков

MT Strategies (1954-2004)



Статистический МП

Статистический МП

- ▶ Подробнее о том, как это делается, — на следующей лекции

Оценка МП

Оценка МП

- ▶ Как оценить качество перевода?

Оценка МП

- ▶ Как оценить качество перевода?
- ▶ [Koehn 2010, 217–246]

Расхождения у переводчиков I

Расхождения у переводчиков II

- ▶ Israel is responsible for the airport's security.
- ▶ Israel is responsible for safety work at this airport.
- ▶ Israel presides over the security of the airport.
- ▶ Israel took charge of the airport security.
- ▶ The safety of this airport is taken charge of by Israel.
- ▶ This airport's security is the responsibility of the Israeli security officials.

Оценка МП

Оценка МП

- ▶ Не существует идеально правильного перевода

Оценка МП

- ▶ Не существует идеально правильного перевода
- ▶ МП А: Israeli officials responsibility of airport safety

Оценка МП

- ▶ Не существует идеально правильного перевода
- ▶ МП А: Israeli officials responsibility of airport safety
- ▶ МП В: airport security Israeli officials are responsible

Оценка МП

- ▶ Не существует идеально правильного перевода
- ▶ МП А: Israeli officials responsibility of airport safety
- ▶ МП В: airport security Israeli officials are responsible
- ▶ Что из этого лучше и почему?

Ручная оценка

Ручная оценка

Кто оценивает?

- ▶ В идеале: ассессоры, знающие и исходный язык, и язык перевода

Ручная оценка

Кто оценивает?

- ▶ В идеале: ассессоры, знающие и исходный язык, и язык перевода
- ▶ В реальности: ассессоры, знающие язык перевода и имеющие образцовые переводы, сделанные вручну

Ручная оценка

Кто оценивает?

- ▶ В идеале: ассессоры, знающие и исходный язык, и язык перевода
- ▶ В реальности: ассессоры, знающие язык перевода и имеющие образцовые переводы, сделанные вручну

Ручная оценка

Кто оценивает?

- ▶ В идеале: ассессоры, знающие и исходный язык, и язык перевода
- ▶ В реальности: ассессоры, знающие язык перевода и имеющие образцовые переводы, сделанные вручную

Что оценивается?

Ручная оценка

Кто оценивает?

- ▶ В идеале: ассессоры, знающие и исходный язык, и язык перевода
- ▶ В реальности: ассессоры, знающие язык перевода и имеющие образцовые переводы, сделанные вручну

Что оценивается?

- ▶ Качество языка перевода (fluency)

Ручная оценка

Кто оценивает?

- ▶ В идеале: ассессоры, знающие и исходный язык, и язык перевода
- ▶ В реальности: ассессоры, знающие язык перевода и имеющие образцовые переводы, сделанные вручну

Что оценивается?

- ▶ Качество языка перевода (fluency)
- ▶ Адекватность оригиналу (adequacy)

Качество языка и адекватность

Качество языка и адекватность

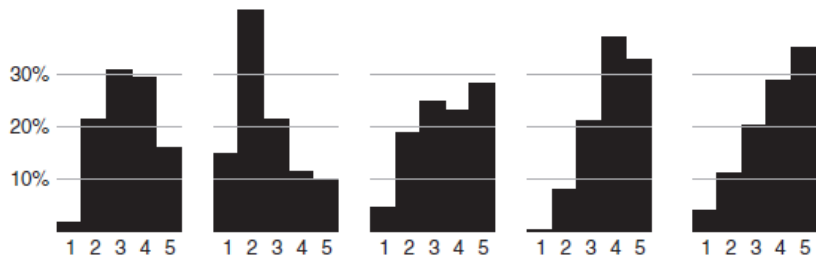
- ▶ МП А: Israeli officials responsibility of airport safety

Качество языка и адекватность

- ▶ МП А: Israeli officials responsibility of airport safety
- ▶ МП В: airport security Israeli officials are responsible

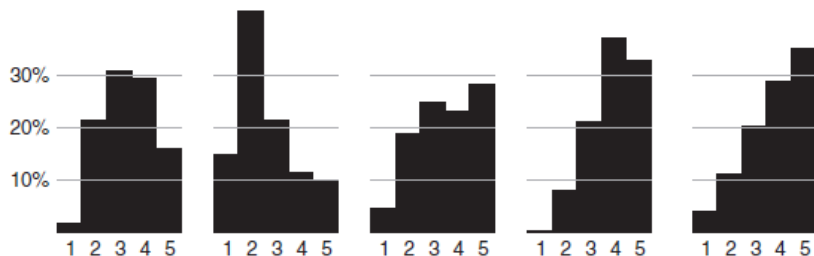
Расхождения в оценке адекватности

Расхождения в оценке адекватности



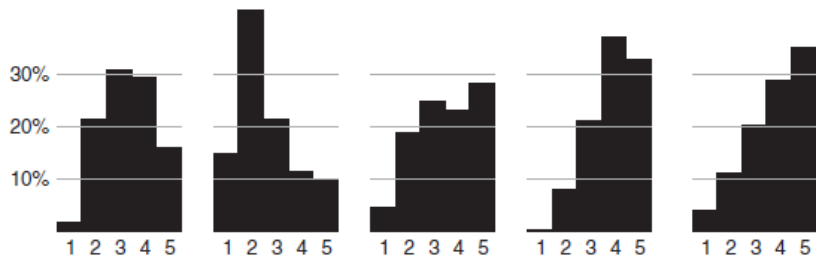
- ▶ Разные ассессоры по-разному понимают шкалу оценки

Расхождения в оценке адекватности



- ▶ Разные ассессоры по-разному понимают шкалу оценки
- ▶ То же касается и оценки качества языка

Расхождения в оценке адекватности



- ▶ Разные ассессоры по-разному понимают шкалу оценки
- ▶ То же касается и оценки качества языка
- ▶ Необходимо нормирование

Ручная оценка

Ручная оценка

- ▶ Разное восприятие адекватности до и после прочтения образцового перевода

Ручная оценка

- ▶ Разное восприятие адекватности до и после прочтения образцового перевода
- ▶ Более надёжны ответы на вопрос «какой перевод лучше — А или В?»

Автоматическая оценка

Автоматическая оценка

Мера должна быть:

Автоматическая оценка

Мера должна быть:

- ▶ недорогой

Автоматическая оценка

Мера должна быть:

- ▶ недорогой
- ▶ измеримой в процессе разработки систем

Автоматическая оценка

Мера должна быть:

- ▶ недорогой
- ▶ измеримой в процессе разработки систем
- ▶ содержательной

Автоматическая оценка

Мера должна быть:

- ▶ недорогой
- ▶ измеримой в процессе разработки систем
- ▶ содержательной
- ▶ устойчивой

Автоматическая оценка

Мера должна быть:

- ▶ недорогой
- ▶ измеримой в процессе разработки систем
- ▶ содержательной
- ▶ устойчивой
- ▶ правильной :)

Автоматическая оценка

Автоматическая оценка

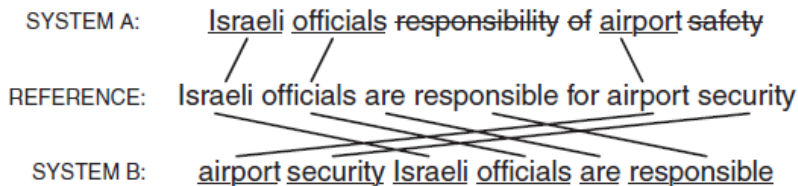
- ▶ Автоматическая оценка МП ориентируется на образцовые переводы, сделанные вручную

Автоматическая оценка

- ▶ Автоматическая оценка МП ориентируется на образцовые переводы, сделанные вручную
- ▶ Правильная автоматическая оценка МП должна хорошо коррелировать с ручной оценкой по качеству языка и адекватности

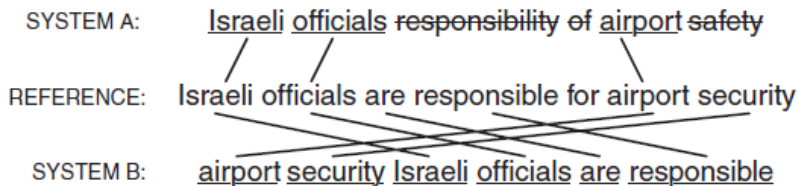
Точность и полнота

Точность и полнота



- ▶ **Точность (precision)** — какая доля слов машинного перевода правильна

Точность и полнота



- ▶ **Точность (precision)** — какая доля слов машинного перевода правильна
- ▶ **Полнота (recall)** — какая доля слов образцового перевода передана

Точность и полнота

Точность и полнота

- ▶ s — число слов, совпадающих в машинном и в образцовом переводе

Точность и полнота

- ▶ s — число слов, совпадающих в машинном и в образцовом переводе
- ▶ o — длина машинного перевода (output)

Точность и полнота

- ▶ s — число слов, совпадающих в машинном и в образцовом переводе
- ▶ o — длина машинного перевода (output)
- ▶ r — длина образцового перевода (reference)

Точность и полнота

- ▶ c — число слов, совпадающих в машинном и в образцовом переводе
- ▶ o — длина машинного перевода (output)
- ▶ r — длина образцового перевода (reference)
- ▶ Точность $P = \frac{c}{o}$

Точность и полнота

- ▶ c — число слов, совпадающих в машинном и в образцовом переводе
- ▶ o — длина машинного перевода (output)
- ▶ r — длина образцового перевода (reference)
- ▶ Точность $P = \frac{c}{o}$
- ▶ Полнота $R = \frac{c}{r}$

Точность и полнота

- ▶ c — число слов, совпадающих в машинном и в образцовом переводе
- ▶ o — длина машинного перевода (output)
- ▶ r — длина образцового перевода (reference)
- ▶ Точность $P = \frac{c}{o}$
- ▶ Полнота $R = \frac{c}{r}$
- ▶ Любую одну из этих мер легко завязать

F-мера

F-мера

- ▶ F -мера отражает баланс точности и полноты (среднее гармоническое)

F-мера

- ▶ F-мера отражает баланс точности и полноты (среднее гармоническое)
- ▶
$$F = \frac{P \times R}{(P + R)/2}$$

F-мера

- ▶ F-мера отражает баланс точности и полноты (среднее гармоническое)
- ▶
$$F = \frac{P \times R}{(P + R)/2}$$
- ▶ При $P = \frac{c}{o}$ и $R = \frac{c}{r}$:

F-мера

- ▶ F-мера отражает баланс точности и полноты (среднее гармоническое)
- ▶
$$F = \frac{P \times R}{(P + R)/2}$$
- ▶ При $P = \frac{c}{o}$ и $R = \frac{c}{r}$:

F-мера

- ▶ F-мера отражает баланс точности и полноты (среднее гармоническое)
- ▶
$$F = \frac{P \times R}{(P + R)/2}$$
- ▶ При $P = \frac{c}{o}$ и $R = \frac{c}{r}$:
$$F = \frac{2c}{r + o}$$

Точность, полнота, F-мера

Точность, полнота, F-мера

SYSTEM A: Israeli officials responsibility of airport safety

REFERENCE: Israeli officials are responsible for airport security

SYSTEM B: airport security Israeli officials are responsible

Точность, полнота, F-мера

SYSTEM A: Israeli officials responsibility of airport safety

REFERENCE: Israeli officials are responsible for airport security

SYSTEM B: airport security Israeli officials are responsible

Мера	A	B
точность	50%	100%
полнота	43%	86%
F-мера	46%	92%

Word Error Rate (WER)

Word Error Rate (WER)

- ▶ Можно измерять расстояние между машинным и образцовым переводом так же, как мы измеряем расстояние между словарным словом и словом с опечаткой (считая символами не буквы, а целые слова)

Word Error Rate (WER)

- ▶ Можно измерять расстояние между машинным и образцовым переводом так же, как мы измеряем расстояние между словарным словом и словом с опечаткой (считая символами не буквы, а целые слова)
- ▶ **Word Error Rate (WER)** — взвешенное расстояние Левенштейна между машинным и образцовым переводом

Word Error Rate (WER)

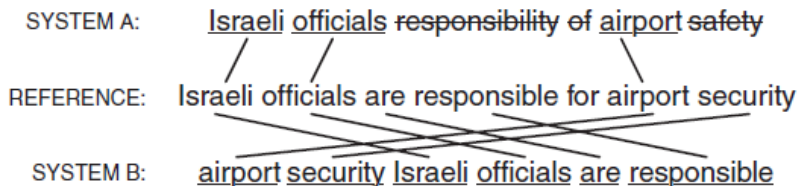
Word Error Rate (WER)

		Israeli	officials	responsibility	of	airport	safety
	0	1	2	3	4	5	6
Israeli	1	0	1	2	3	4	5
officials	2	1	0	1	2	3	4
are	3	2	1	1	2	3	4
responsible	4	3	2	2	2	3	4
for	5	4	3	3	3	3	4
airport	6	5	4	4	4	3	4
security	7	6	5	5	5	4	4

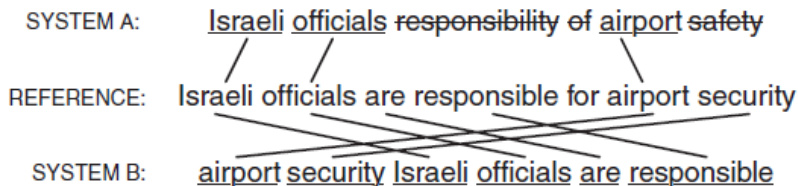
		airport	security	Israeli	officials	are	responsible
	0	1	2	3	4	5	6
Israeli	1	1	2	2	3	4	5
officials	2	2	2	3	2	3	4
are	3	3	3	3	3	2	3
responsible	4	4	4	4	4	3	2
for	5	5	5	5	5	4	3
airport	6	5	6	6	6	5	4
security	7	6	5	6	7	6	5

Word Error Rate (WER)

Word Error Rate (WER)



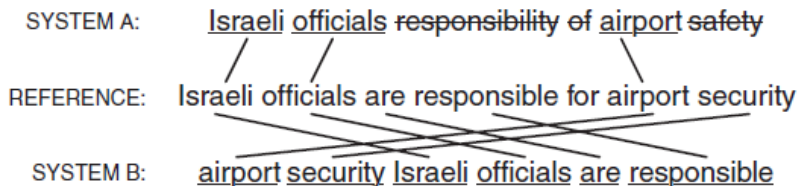
Word Error Rate (WER)



- ▶ $WER = \text{число редакторских операций} / r$

Мера	A	B
WER	57%	71%

Word Error Rate (WER)



- ▶ WER = число редакторских операций / r
- ▶ NB: чем выше WER, тем хуже

Мера	A	B
WER	57%	71%

BLEU

BLEU

SYSTEM A: Israeli officials responsibility of airport safety

2-GRAM MATCH

1-GRAM MATCH

REFERENCE: Israeli officials are responsible for airport security

SYSTEM B: airport security Israeli officials are responsible

2-GRAM MATCH

4-GRAM MATCH

- ▶ Можно учитывать не только отдельные слова, но и сочетания n слов (n -граммы)

BLEU

SYSTEM A: Israeli officials responsibility of airport safety
2-GRAM MATCH 1-GRAM MATCH

REFERENCE: Israeli officials are responsible for airport security

SYSTEM B: airport security Israeli officials are responsible
2-GRAM MATCH 4-GRAM MATCH

- ▶ Можно учитывать не только отдельные слова, но и сочетания n слов (n -граммы)
- ▶ **BLEU (Bilingual Evaluation Understudy)** — самая популярная мера в современной оценке МП

BLEU

BLEU

SYSTEM A: Israeli officials responsibility of airport safety
2-GRAM MATCH 1-GRAM MATCH

REFERENCE: Israeli officials are responsible for airport security

SYSTEM B: airport security Israeli officials are responsible
2-GRAM MATCH 4-GRAM MATCH

- ▶ P_n — точность на уровне n -грамм

BLEU

SYSTEM A: Israeli officials responsibility of airport safety
 2-GRAM MATCH 1-GRAM MATCH

REFERENCE: Israeli officials are responsible for airport security

SYSTEM B: airport security Israeli officials are responsible
 2-GRAM MATCH 4-GRAM MATCH

- ▶ P_n — точность на уровне n -грамм
- ▶ $BLEU_n = \min(1, \frac{o}{r}) \prod_{i=1}^n P_i$

BLEU

SYSTEM A: Israeli officials responsibility of airport safety
 2-GRAM MATCH 1-GRAM MATCH

REFERENCE: Israeli officials are responsible for airport security

SYSTEM B: airport security Israeli officials are responsible
 2-GRAM MATCH 4-GRAM MATCH

- ▶ P_n — точность на уровне n -грамм
- ▶ $BLEU_n = \min(1, \frac{o}{r}) \prod_{i=1}^n P_i$
- ▶ (первый множитель — штраф за краткость)

SYSTEM A: Israeli officials responsibility of airport safety
 2-GRAM MATCH 1-GRAM MATCH

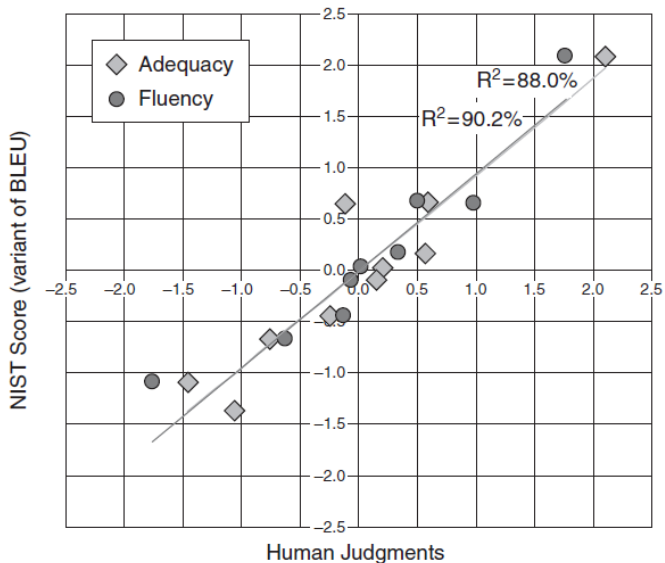
REFERENCE: Israeli officials are responsible for airport security

SYSTEM B: airport security Israeli officials are responsible
 2-GRAM MATCH 4-GRAM MATCH

Мера	A	B	Мера	A	B
штраф	6/7	6/7			
P_1	3/6	6/6	$BLEU_1$	43%	86%
P_2	1/5	4/5	$BLEU_2$	9%	69%
P_3	0/4	2/4	$BLEU_3$	0%	34%
P_4	0/3	1/3	$BLEU_4$	0%	11%

BLEU и ручная оценка

BLEU и ручная оценка



BLEU: за

BLEU: за

- ▶ +: можно сравнивать с несколькими образцовыми переводами

BLEU: за

- ▶ +: можно сравнивать с несколькими образцовыми переводами
- ▶ +: хорошо коррелирует с ручной оценкой качества языка перевода и адекватности

BLEU: против

VLEU: против

- ▶ -: не учитываются различия в значимости слов (имена, *не*, ...)

VLEU: против




- ▶ -: не учитываются различия в значимости слов (имена, *he*, ...)
- ▶ -: не учитывается грамматичность (дискриминация правильного МП)




BLEU: против

- ▶ -: не учитываются различия в значимости слов (имена, *не*, ...)
- ▶ -: не учитывается грамматичность (дискриминация правильного МП)
- ▶ -: значения плохо интерпретируемы и не сравнимы при разных тестах (что значит, напр., $BLEU_3 = 34\%$?)

BLEU: против

- ▶ -: не учитываются различия в значимости слов (имена, *he*, ...)
- ▶ -: не учитывается грамматичность (дискриминация правильного МП)
- ▶ -: значения плохо интерпретируемы и не сравнимы при разных тестах (что значит, напр., $BLEU_3 = 34\%$?)
- ▶ -: постредактирование перевода улучшает ручную оценку, но не BLEU

-  Иомдин Л. Л. и Л. Л. Цинман. Лексические функции и машинный перевод // *Труды международного семинара по компьютерной лингвистике и ее приложениям (Диалог'97)*. М., 1997. С. 291–297.
-  Мельчук, И. А. и Жолковский А. К. 1984. *Толково-комбинаторный словарь русского языка*. Wien: Wiener Slavistischer Almanach, Sonderband 14.
-  Фролова Т. Значение лексических функций для качества машинного перевода.
iitp.ru/upload/publications/2834/frolova.doc

-  Bar-Hillel, Yehoshua. 1960. “The present status of automatic translation of languages.” *Advances in computers* 1.1: 91-163.
-  Knight, Kevin. 1997. “Automating knowledge acquisition for machine translation.” *AI Magazine* 18(4): 81–96.
-  Koehn, Philipp. 2010. *Statistical Machine Translation*. Cambridge: CUP.