

ЛН-21: Анна Воеводина & Алина Сукманова

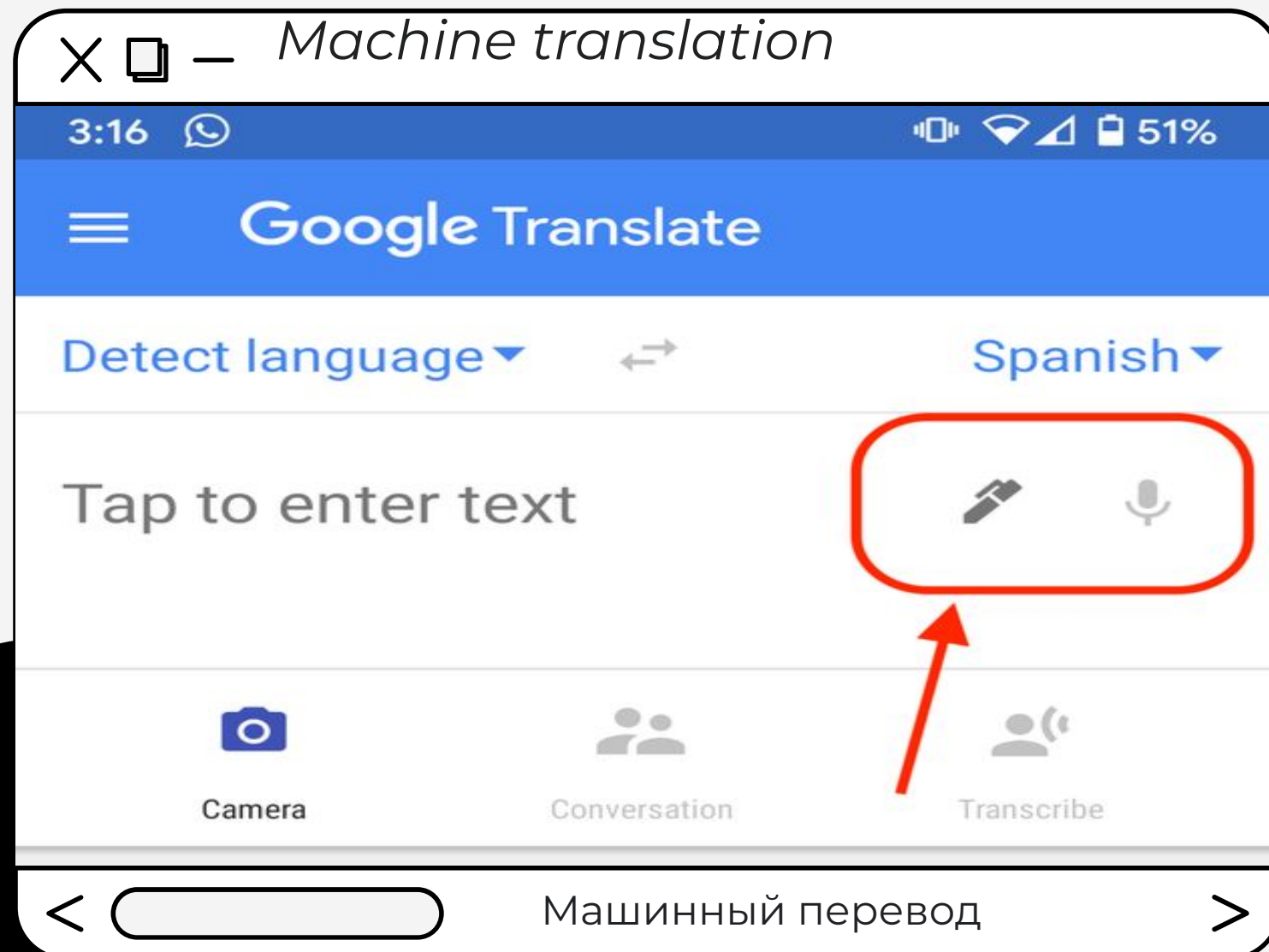
Search



Машинный перевод



MT vs CAT



Машинный перевод — процесс перевода текстов с одного естественного языка на другой с помощью специальной компьютерной программы.

История машинного перевода



Пётр Смирнов-Троянский



Уоррен Уивер

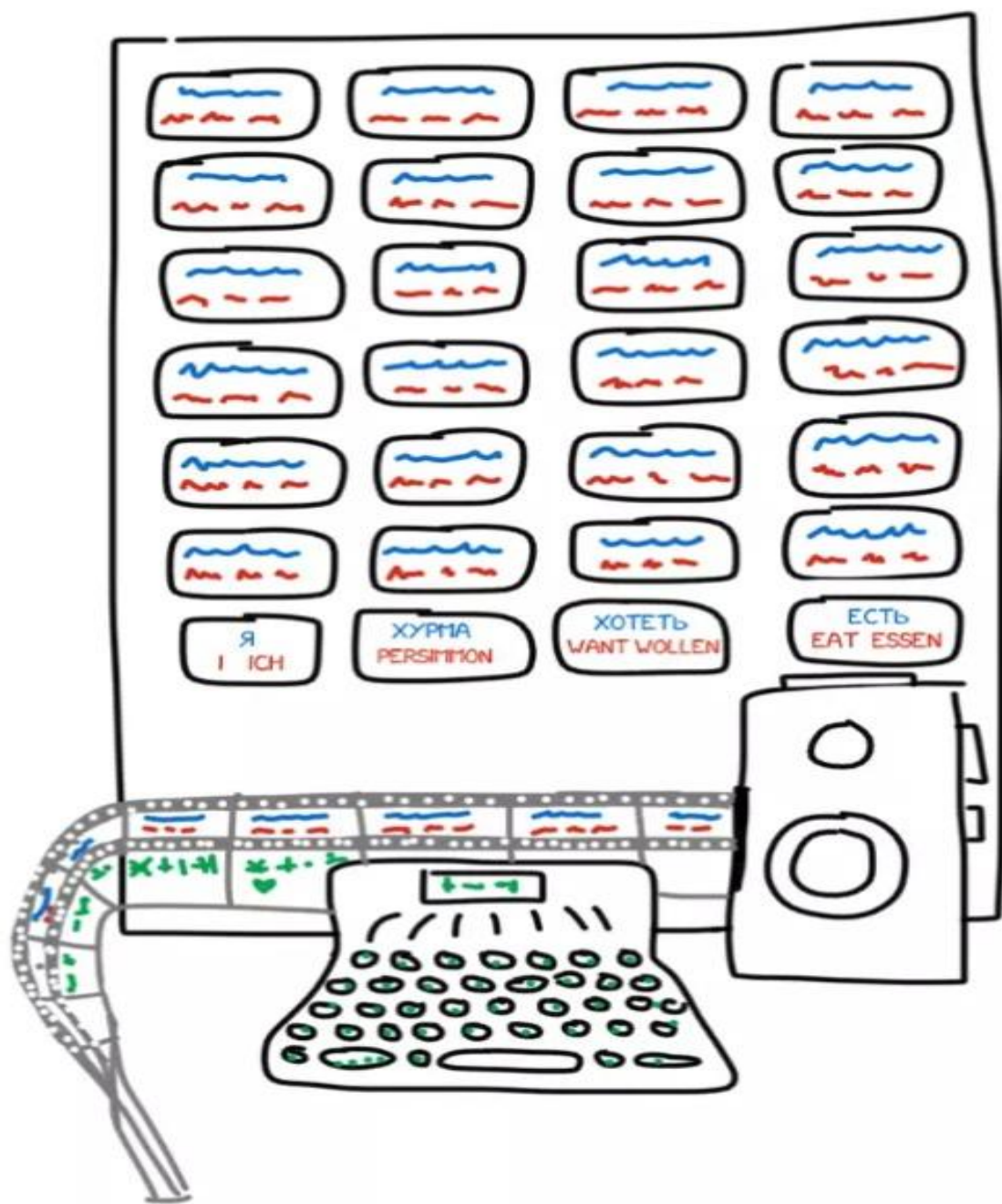


Джорджтаунский эксперимент



Доклад ALPAC





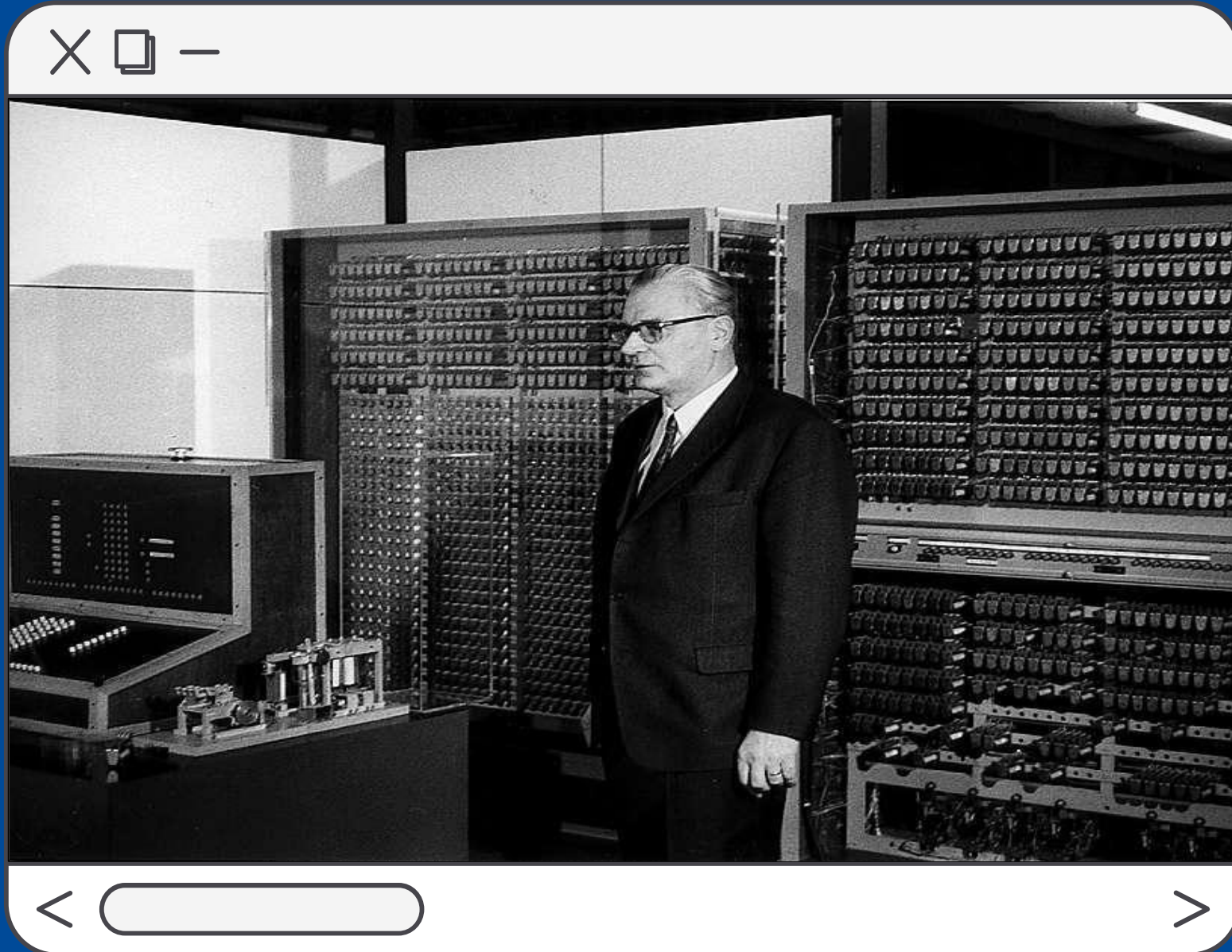
Машина Троянского

Пётр Петрович Смирнов-Троянский (январь 1894 года – 24 мая 1950 года) — советский педагог и учёный, один из пионеров автоматизированного перевода.

В 1933 году П.П. Смирнов-Троянский обратился в Академию наук СССР с разработанной им конструкцией «машины для подбора и печатания слов при переводе с одного языка на другой»

Изобретение Смирнова-Троянского было встречено в академических кругах скептически и на долгое время забыто. Вновь вспомнили о нём только в 1959 году, когда И.К. Бельской и Д. Ю. Пановым была опубликована статья «Переводная машина П.П. Смирнова-Троянского: сборник материалов о машине для перевода с одного языка на другие, предложенной П. П. Смирновым-Троянским в 1933 г.»

Я I ICH YO	ХОТЕТЬ WANT WOLLEN QUERER	МНОГО MANY VIEL MUCHO	ХУРМА PERSIMMON PERSIMONE CAQU
МЕСТ., ЕД. Ч., ИМ. П.	ГЛАГ., I. Л., ЕД. Ч., НЕСОВ., НАСТ. ВР., ДЕЙСТВ. ЗАЛОГ	ЧИСЛ., ИМ. П.	СУЩ., МН. Ч., РОД. П., НЕОДУШ.



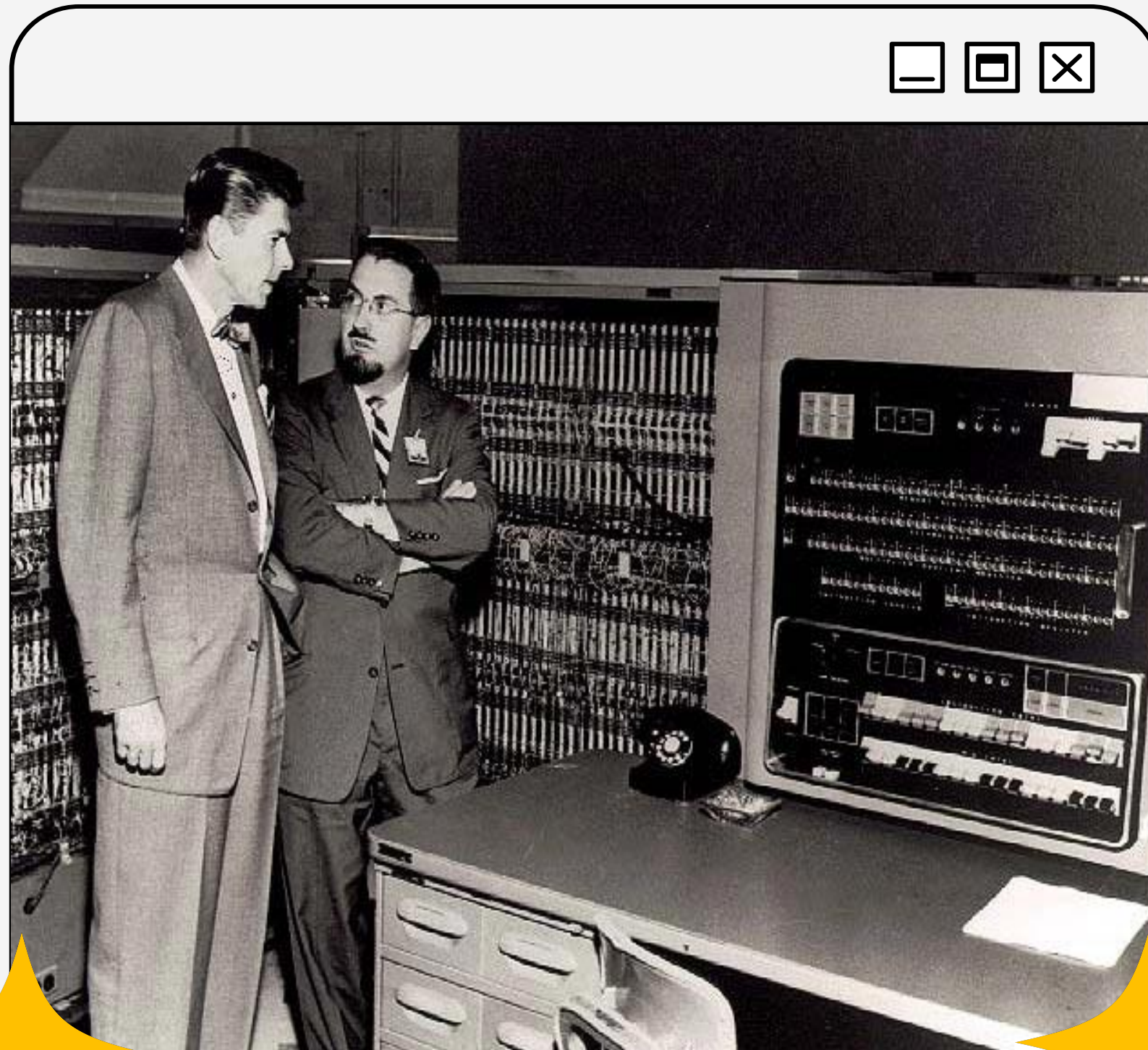
Доктор **Конрад Эрнст Отто Цузе** (*Dr. Konrad Ernst Otto Zuse*)

«Когда я вижу текст на русском языке,
я говорю себе, что на самом деле
он написан по-английски и зашифрован
при помощи странных знаков.
И мне надо его просто расшифровать»

Уоррен Уивер



Джорджтаунский эксперимент

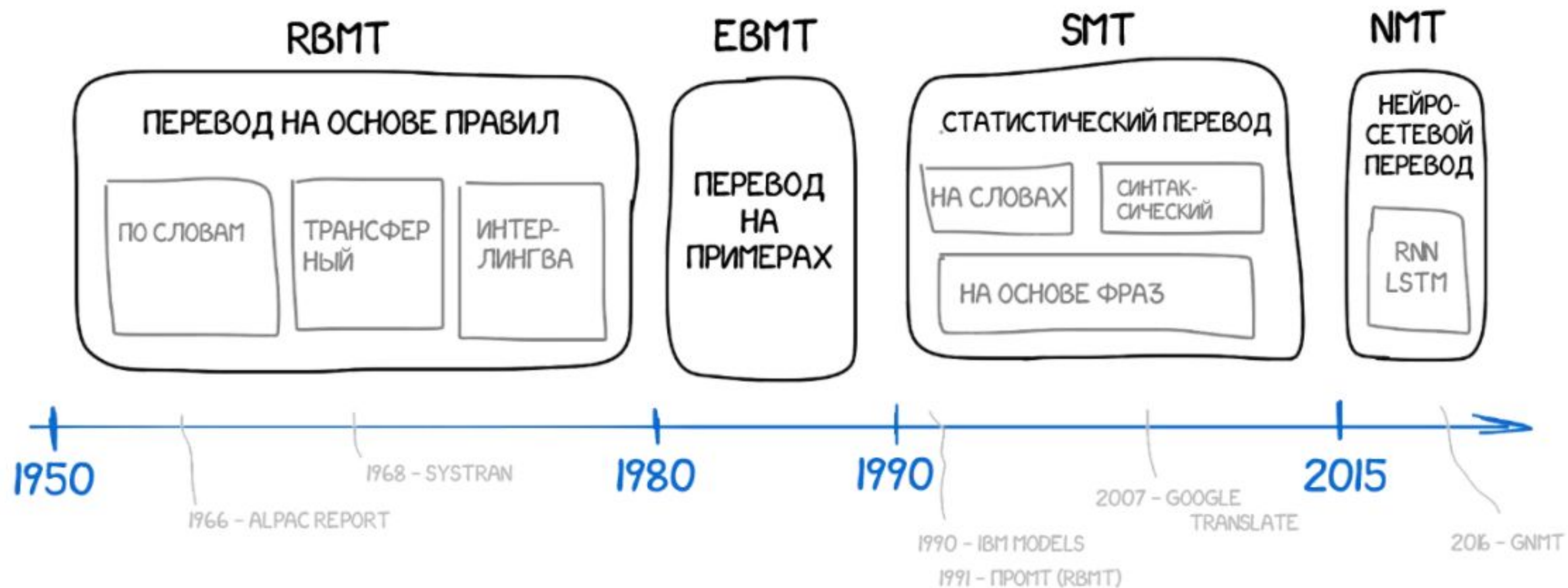


Джорджтаунский эксперимент — демонстрация возможностей машинного перевода, состоявшаяся 7 января 1954 года в Нью-Йорке, в штаб-квартире корпорации IBM. Подготовлена Джорджтаунским Университетом совместно с IBM.

В ходе эксперимента был продемонстрирован полностью автоматический перевод более 60 предложений с русского языка на английский.

В его основе лежала довольно простая система: она была основана всего на 6 грамматических правилах, а словарь включал 250 записей. Система была специализированной: в качестве предметной области для перевода была выбрана органическая химия, также был добавлен ряд предложений общей тематики.

Подходы к машинному переводу





+



+



ДОХЛЫЕ ЛИНГВИСТЫ



ГЛУПЫЕ МАШИНЫ



трансферные системы (исходное предложение → препарированное исходное предложение → предложение на другом языке)



системы пословного перевода (слова исходного текста → слова перевода)



интерлингвистические системы (исходный текст → описание его смысла на универсальном языке-посреднике → текст перевода)

Типы
перевода
на основе
правил



системы пословного перевода (слова исходного текста → слова перевода)

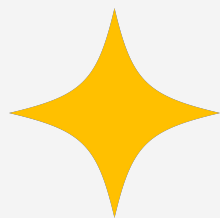
Я		ХОЧУ		СОРОК		КИЛОГРАММ		ХУРМЫ
↓		↓		↓		↓		↓
I		WANT		FORTY		KILOGRAM		PERSIMMONS





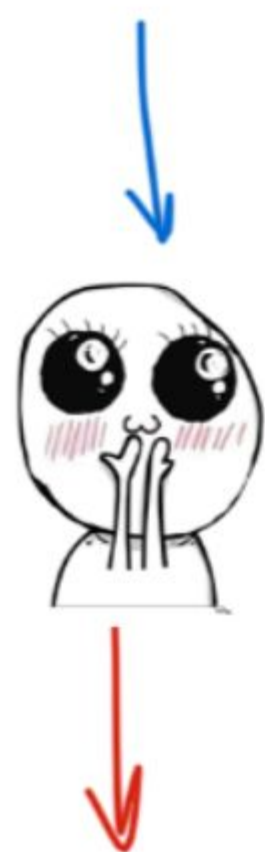
трансферные системы (исходное предложение → препарированное
исходное предложение → предложение на другом языке)





интерлингвистические системы (исходный текст → описание его смысла на универсальном языке-посреднике → текст перевода)

Я ХОЧУ СОРОК КИЛОГРАММ ХУРМЫ



I WANT FORTY KILOGRAMS OF PERSIMMONS



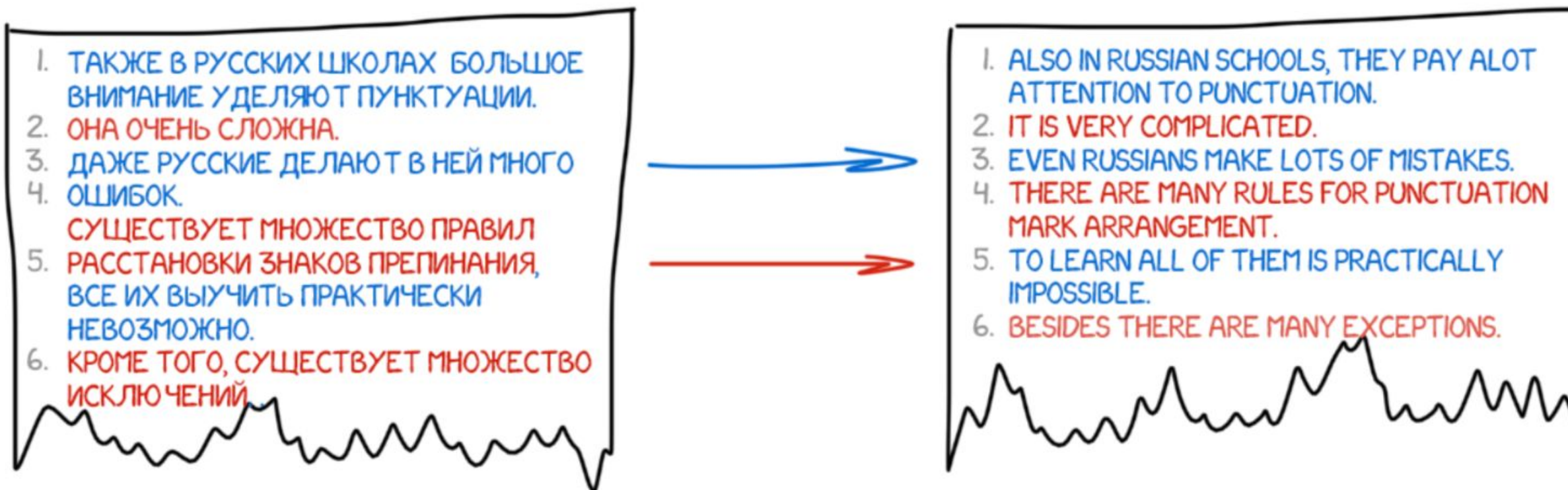
(УЖЕ ЗНАКОМЫЙ НАМ ПРИМЕР)

Я ИДУ В ТЕАТР = I'M GOING TO THE THEATER
Я ИДУ В МАГАЗИН^{???} = I'M GOING TO THE STORE

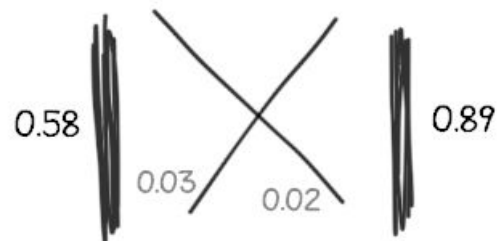
→ STORE →



ПАРАЛЛЕЛЬНЫЙ КОРПУС

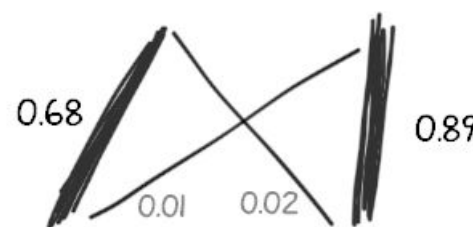


THE HOUSE



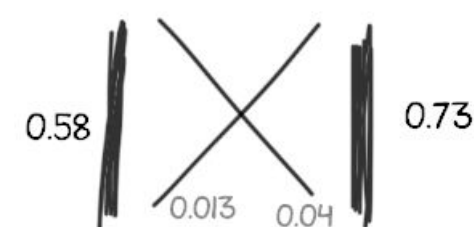
DAS HAUS

BLUE HOUSE



BLAUES HAUS

THE CAR



DAS AUTO

английский немецкий русский ▼ Перевести

leave ✓

☆ 📄 🔊 ⏪ ✎ Предложить исправление

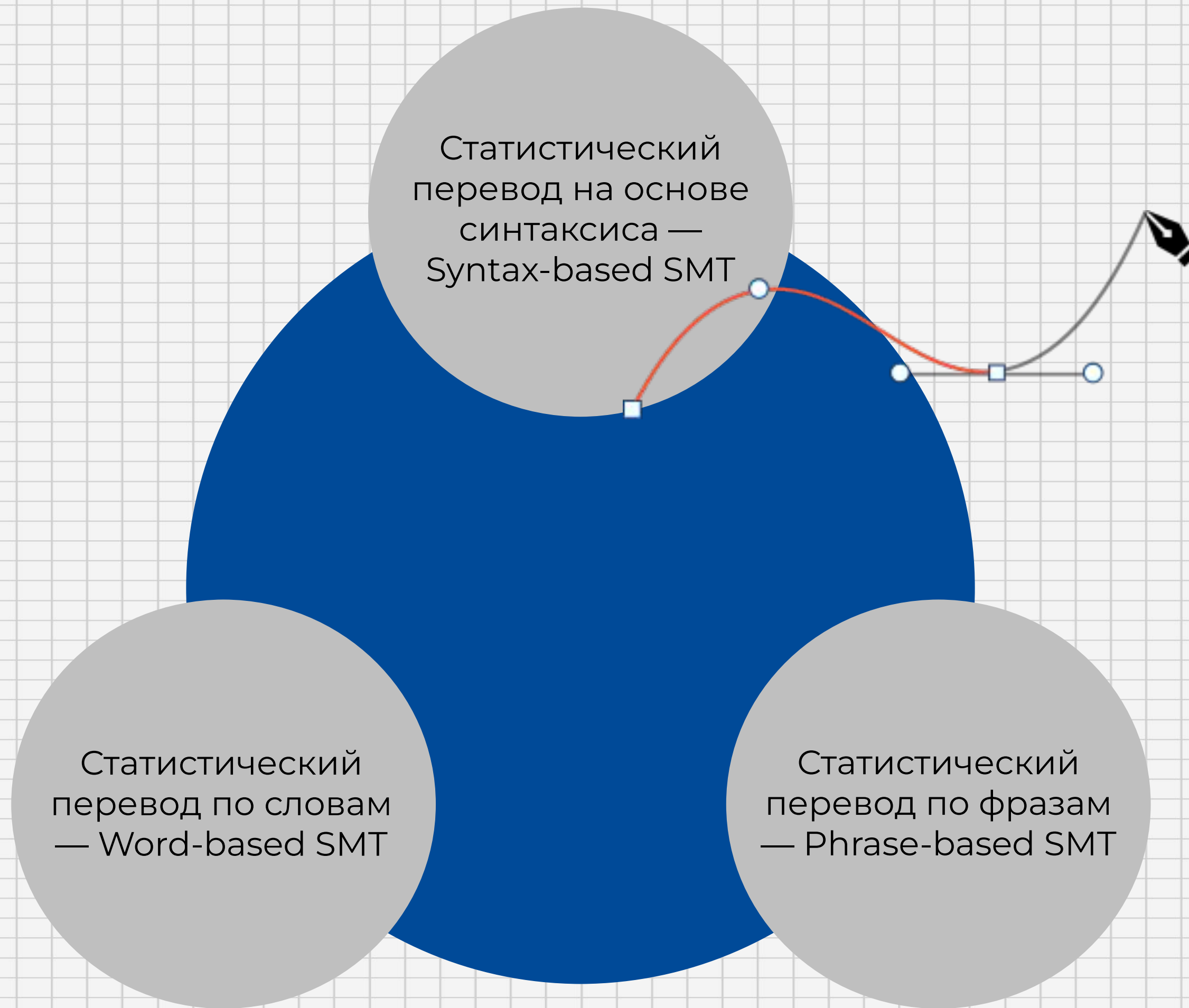
уходить: варианты перевода

глагол

■ go	идти, ехать, ходить, ездить, уходить, проходить
■ go away	уходить, отойти, разойтись
■ depart	отступать, отклоняться, уходить, отправляться, уезжать, отбывать
■ leave	оставлять, покидать, уходить, уезжать, бросать, удаляться
■ exit	выходить, уходить, умереть
■ get out	выходить, выбираться, вылезать, убираться, уходить, выкарабкиваться
■ get away	избежать, уходить, ускользнуть, удирать, выбираться, отрываться
■ walk away	уходить, унести, обходить стороной, уводить, украсть
■ walk off	уходить, унести, одержать легкую победу, украсть, уводить
■ go off	уходить, сходить, выстрелить, уезжать, сбежать, проходить
■ withdraw	изымать, отзываться, уходить, забирать, увести, удаляться
■ take away	увезти, забрать, забирать, отнимать, унести, уходить
■ come away	уходить, отламываться, отходить
■ retire	уходить, удаляться, уходить в отставку, увольняться, отступать, уединяться
■ retreat	отступать, отходить, уходить, удаляться, отбросить назад
■ walk out	уходить, демонстративно покинуть, выходить
■ be off	уходить
■ buzz off	уходить, удаляться, смыться, улизнуть
■ scam	уходить, удирать
■ be away	отсутствовать, уходить, не быть дома
■ shove off	убираться, уходить, отталкиваться
■ toddle off	уходить
■ toddle	ковылять, учиться ходить, прогуливаться, бродить, уходить
■ draw away	отвлекать, удаляться, уводить, уходить, отрываться от противника
■ split	раскалывать, разбивать, расщепить, раскалываться, расщепляться, уходи

Статистический перевод в Google Translate. Google не только показывает вероятности, но и считает обратную статистику







Я БОЛЬШЕ НЕ ХОЧУ ХУРМЫ

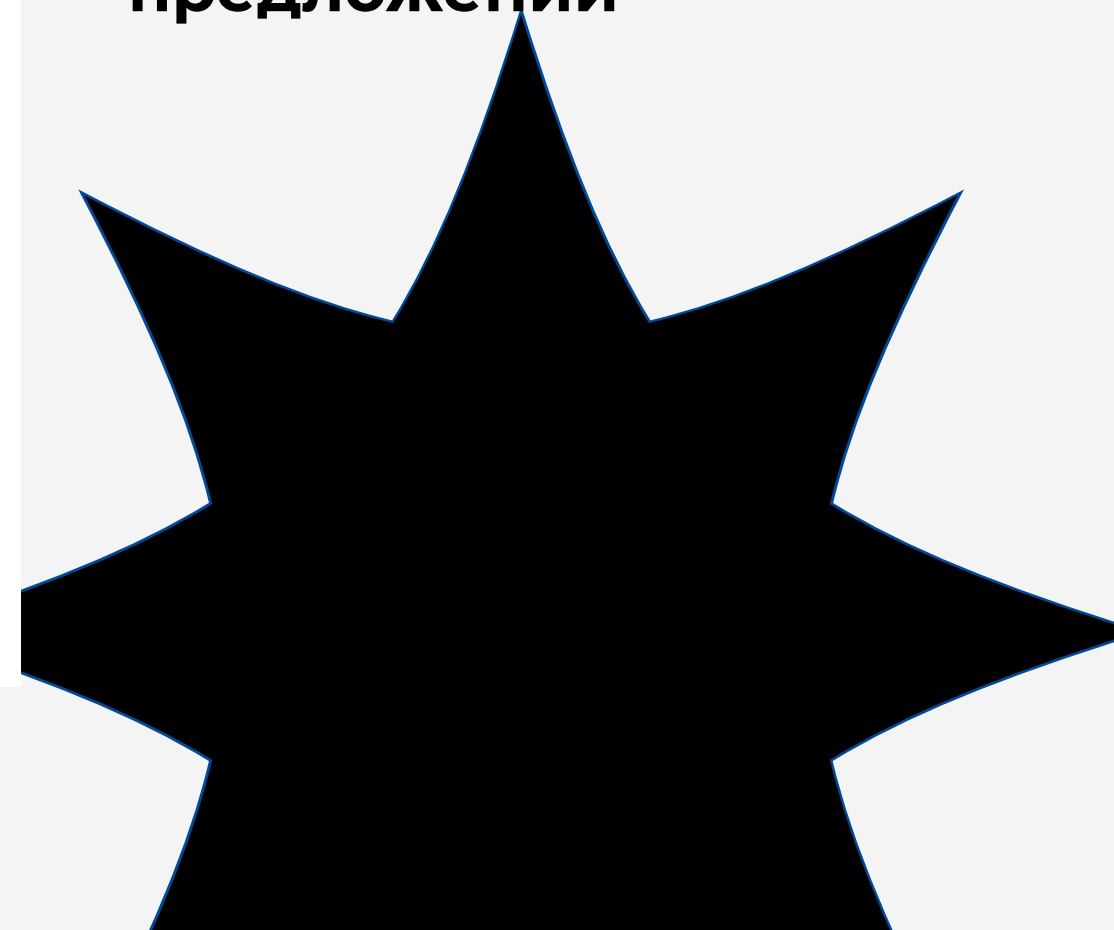
| 0.88 / 0.45 / 0.79 / 0.81 | 0.91

I MORE NOT WANT PERSIMMONS

| |

I NOT WANT MORE PERSIMMONS

Model 1: учет слов
Model 2: учет порядка слов в предложении





Я БОЛЬШЕ НЕ ХОЧУ ХУРМЫ

0.88

0.45

0.79

0.81

0.91

I NOT WANT MORE PERSIMMONS

I **NULL** NOT WANT MORE PERSIMMONS

I DO NOT WANT MORE PERSIMMONS

Model 3: добавление отсутствующих слов

1. Вставка маркеров (NULL-слов) на те места, где машина подозревает необходимость нового слова.
2. Подбор нужного артикля, частицы или глагола под каждый маркер.





FULL SUPERIORITY OF PERSIMMONS

ПЕРЕВОД ПО СЛОВАМ
(ХОРОШО, НО ДОСЛОВНО)

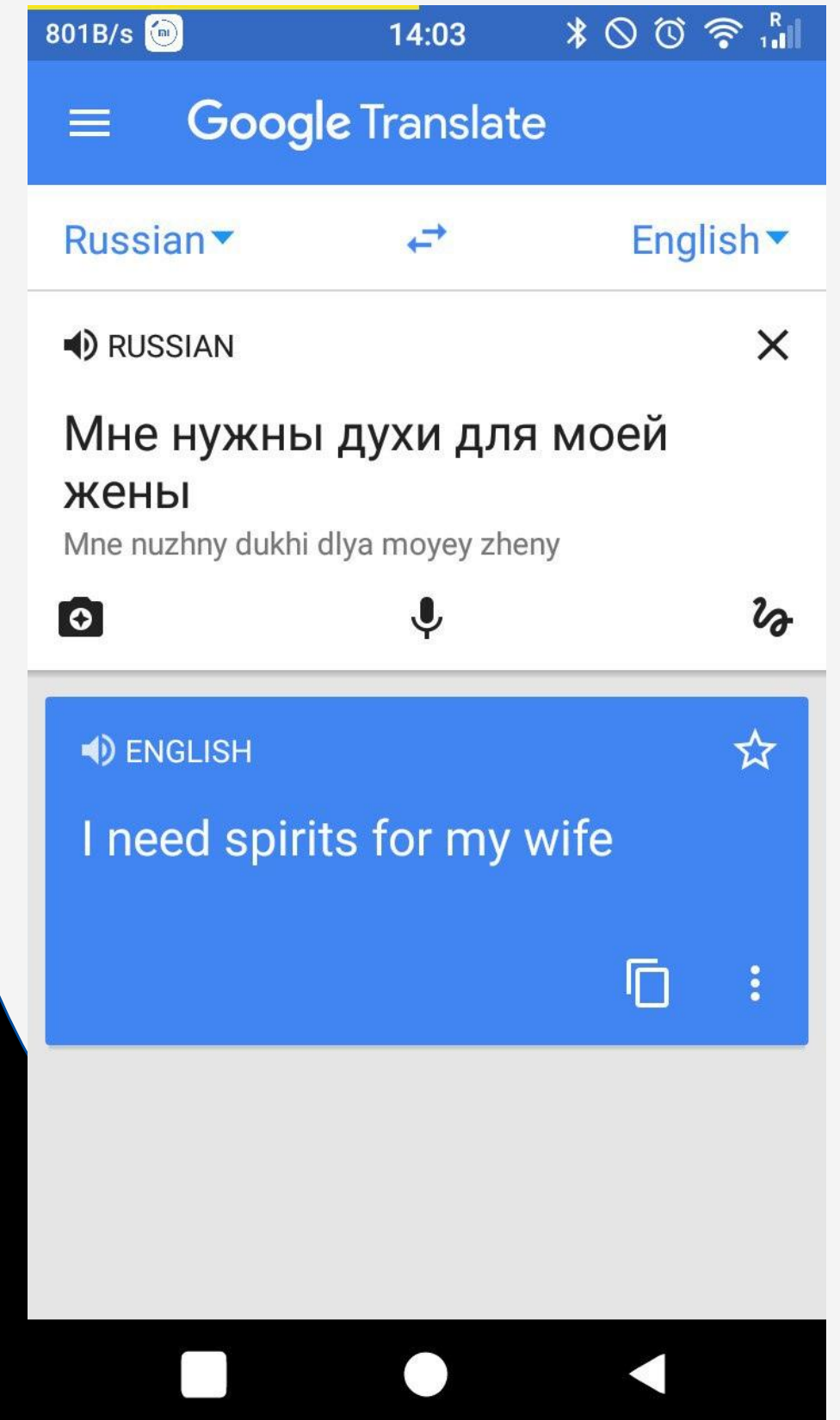
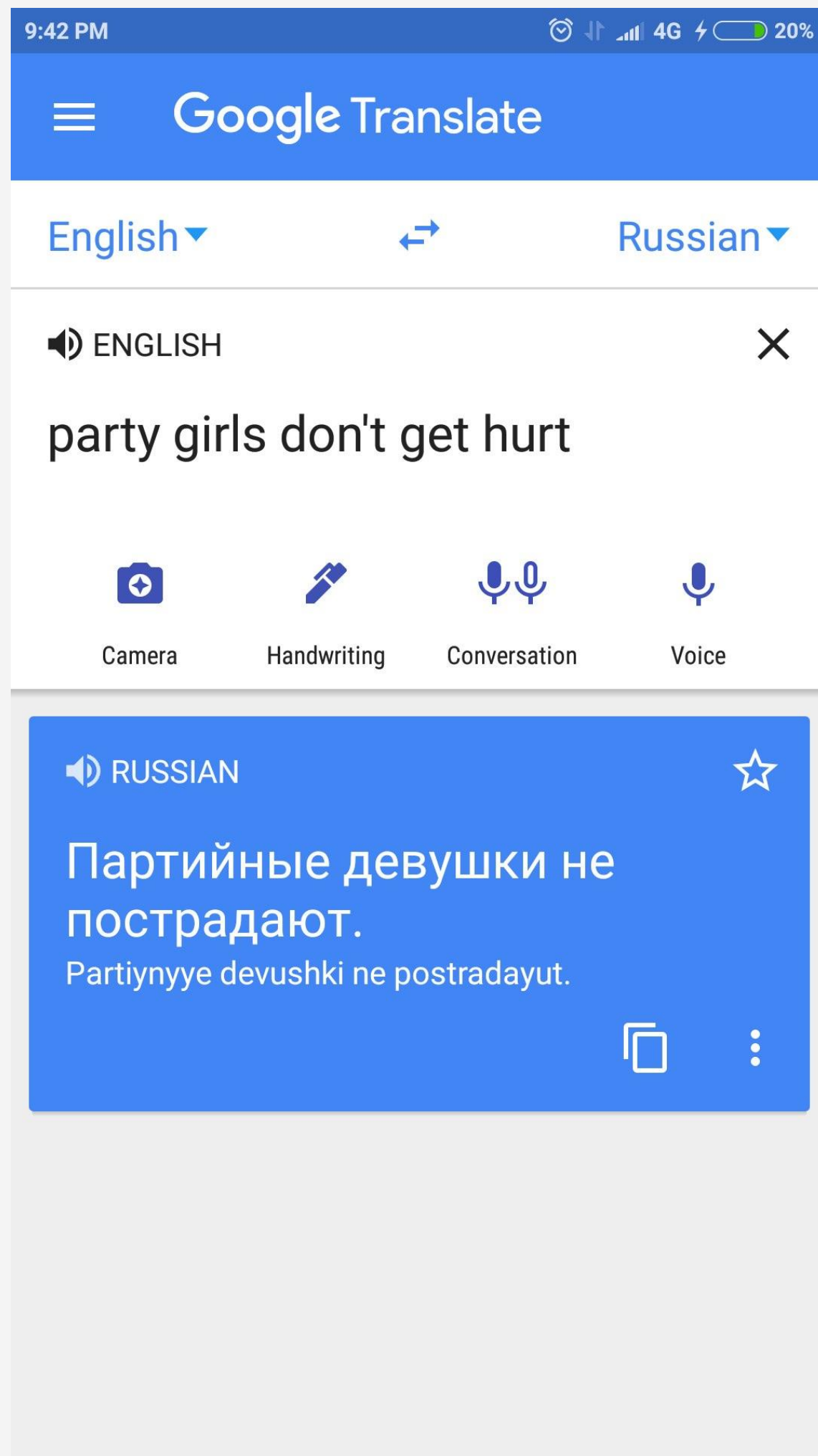
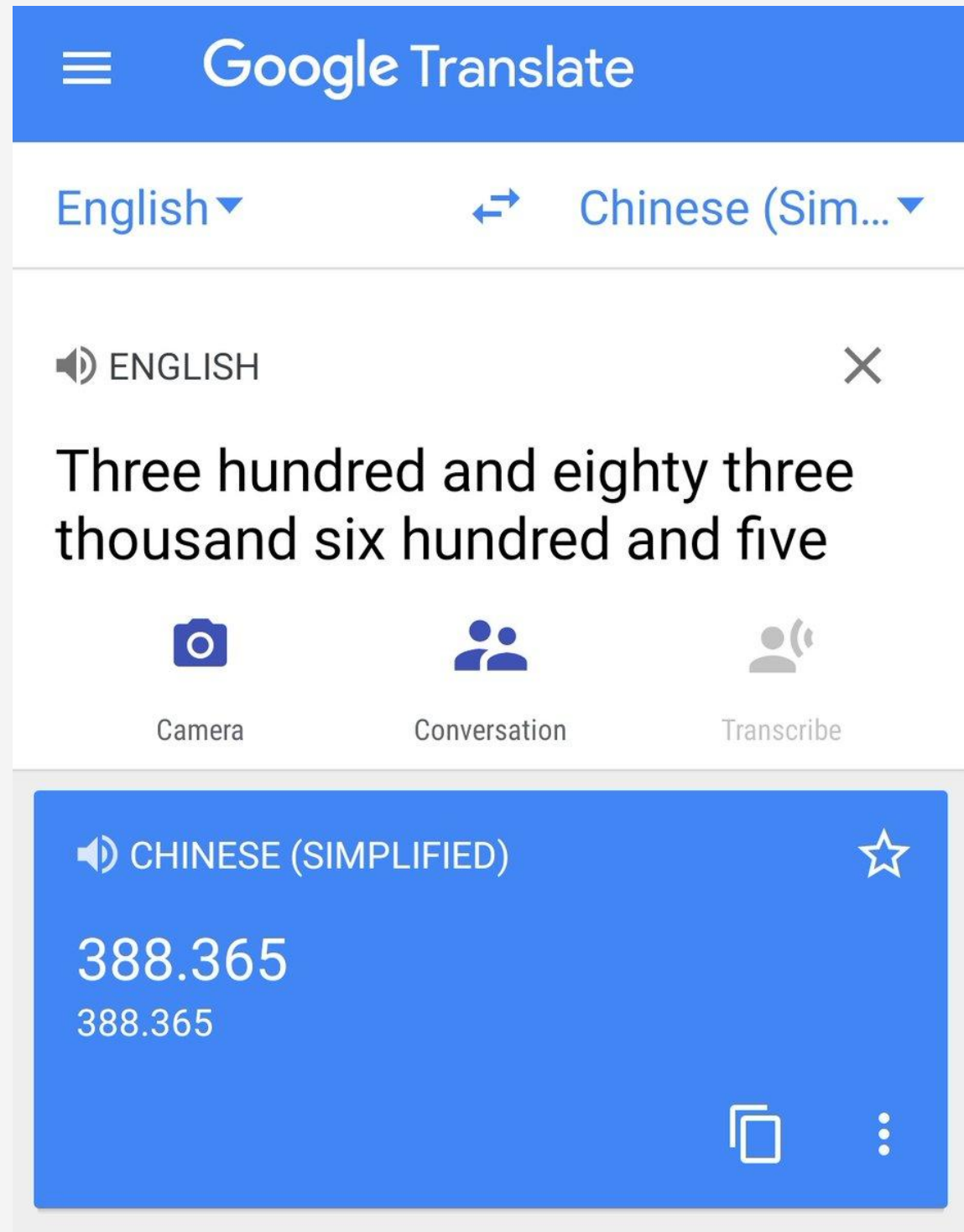
ПОЛНОЕ ПРЕВОСХОДСТВО ХУРМЫ

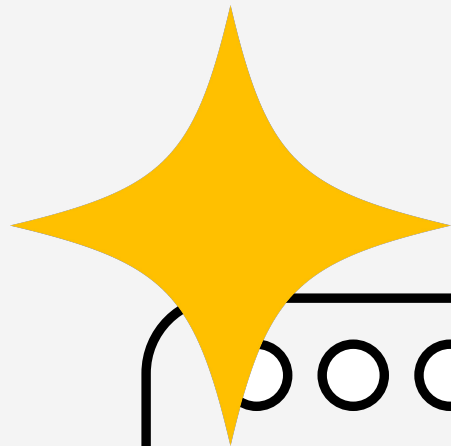
COMPLETE SUPERIORITY

PERSIMMON SUPERIORITY

ПЕРЕВОД ПО ФРАЗАМ
(УЧИТЫВАЕТ КОНТЕКСТ
СОСЕДНИХ СЛОВ)

COMPLETE PERSIMMON SUPERIORITY





○○○



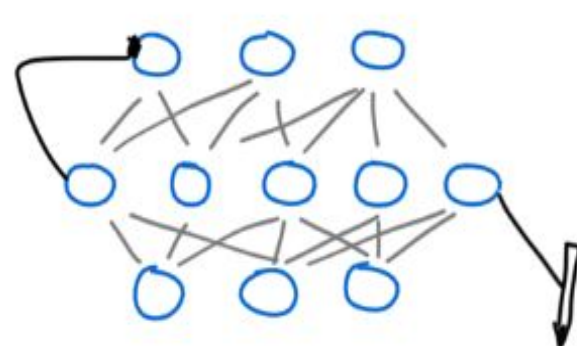
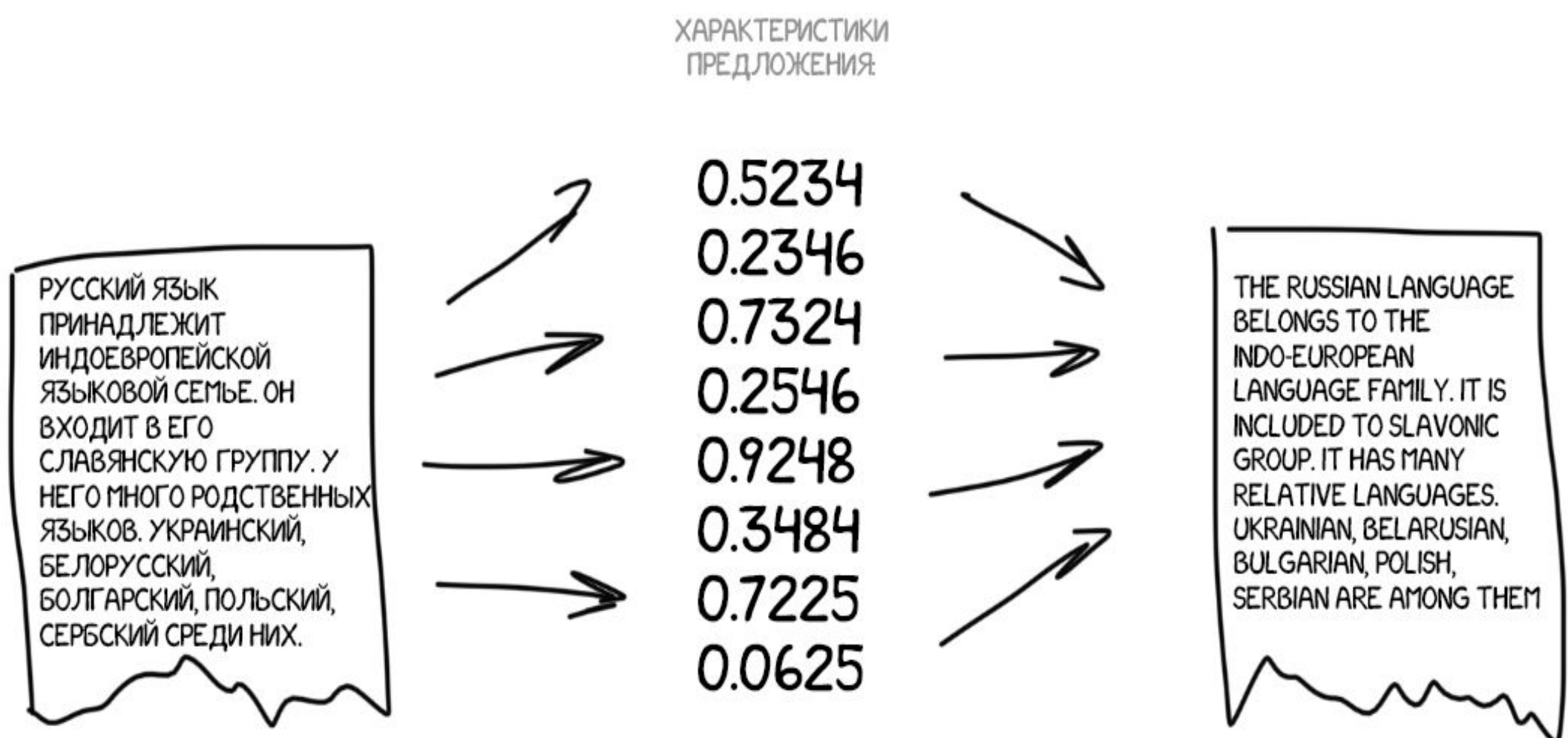
Нейронный машинный перевод —

Neural Machine Translation (NMT)



2016- XXXX





Рекуррентные нейросети (RNN)



РУССКИЙ

Вдохновения дали живописные дали



32 / 10000

АНГЛИЙСКИЙ



Inspirations gave picturesque gave

Перевести в [Google Bing](#)



РУССКИЙ

Вдохновения дали живописные дали.



33 / 10000

АНГЛИЙСКИЙ



Inspirations were given by picturesque distances.

Перевести в [Google Bing](#)

«Эти типы стали есть на складе»

