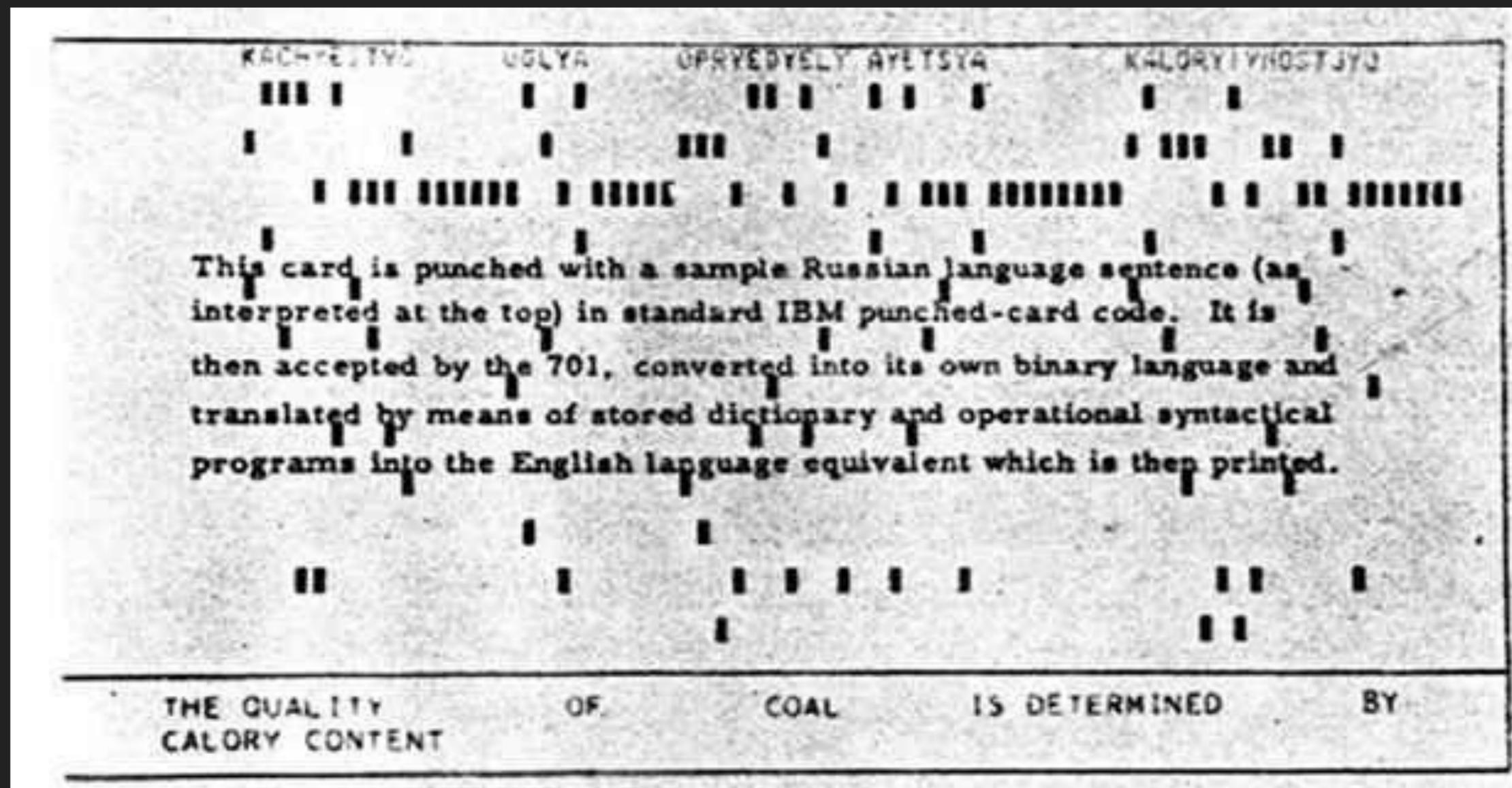


PROMT

ДИАЛЕКТИКА НЕЙРОННОГО
МАШИННОГО ПЕРЕВОДА

**ПЕРЕРАСТАЕТ ЛИ
КОЛИЧЕСТВО В
КАЧЕСТВО**

ВО ВСЕМ ВИНОВАТА ХОЛОДНАЯ ВОЙНА



7 января 1954 года

Georgetown University & IBM

Русско-английский переводчик: 6 грамматических правил, 250 лексических единиц; было переведено 200 фраз

ОТ ДЖОРДЖТАУНА ДО НЕЙРОННЫХ СЕТЕЙ

- ▶ 1966 год. Отчёт ALPAC (Automatic Language Processing Advisory Committee): развитие машинного перевода нецелесообразно
- ▶ Развитие персональных компьютеров дало толчок к появлению первых коммерческих систем машинного перевода в 80-90 годах
- ▶ Декабрь 1997 года — первый в мире онлайн-сервис перевода Babel Fish на сайте AltaVista
- ▶ Март 1998 года — первый в Рунете онлайн-сервис перевода Translate.Ru от PROMT
- ▶ Апрель 2006 года — онлайн-сервис перевода Google Translate, следом Bing Translator от Microsoft, Яндекс.Переводчик, Baidu Translator, Alibaba Group
- ▶ Сентябрь 2016 года — Neural Machine Translation (Google, Microsoft, Systran)

NEURAL MACHINE TRANSLATION: ПОЧЕМУ ТАК ДОЛГО?

Динамика популярности

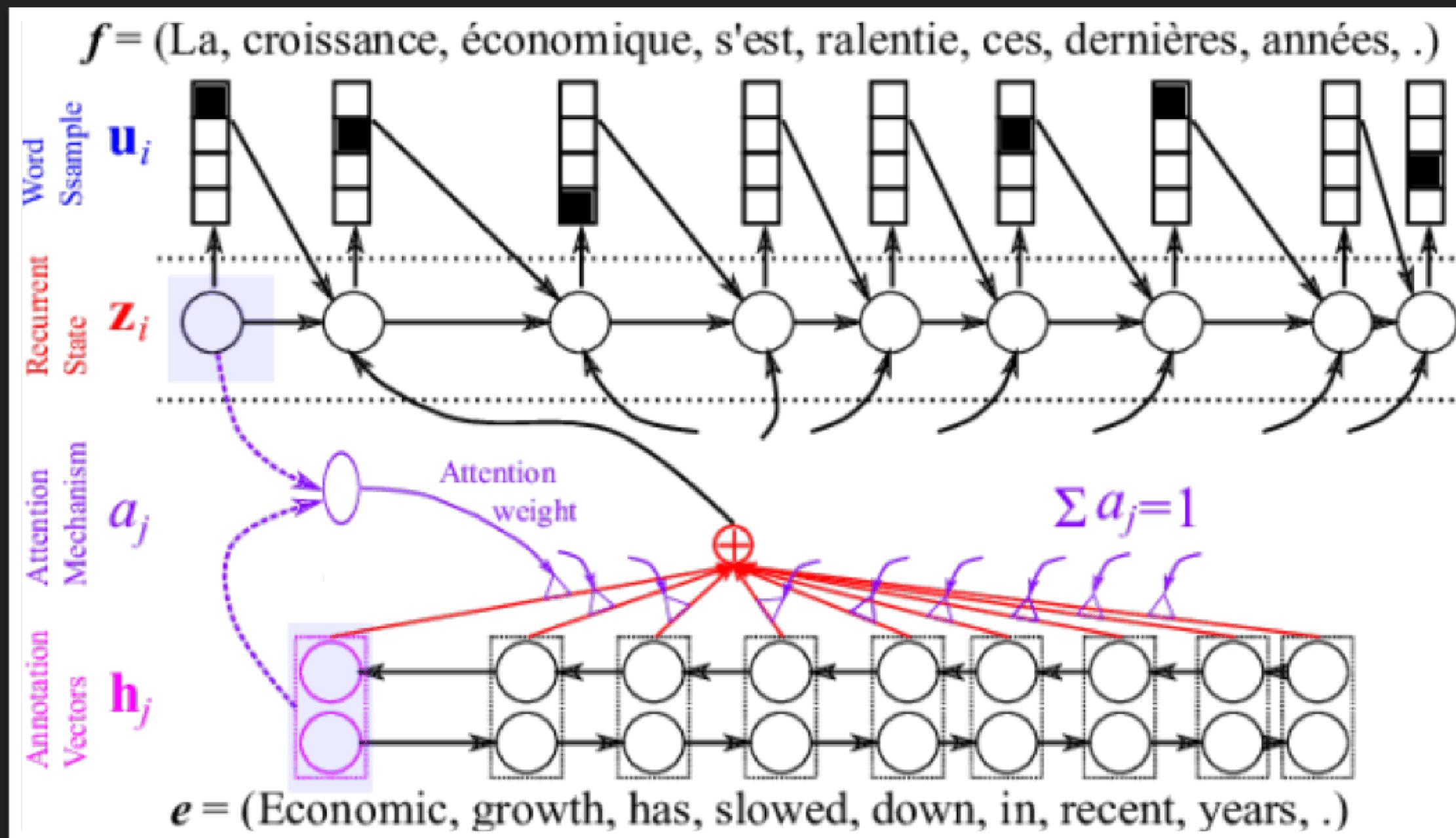
Google Trends

● Neural Network ● Neural Machine Translation



По всему миру. 2008 – настоящее время.

NEURAL MACHINE TRANSLATION: ЧТО ВНУТРИ



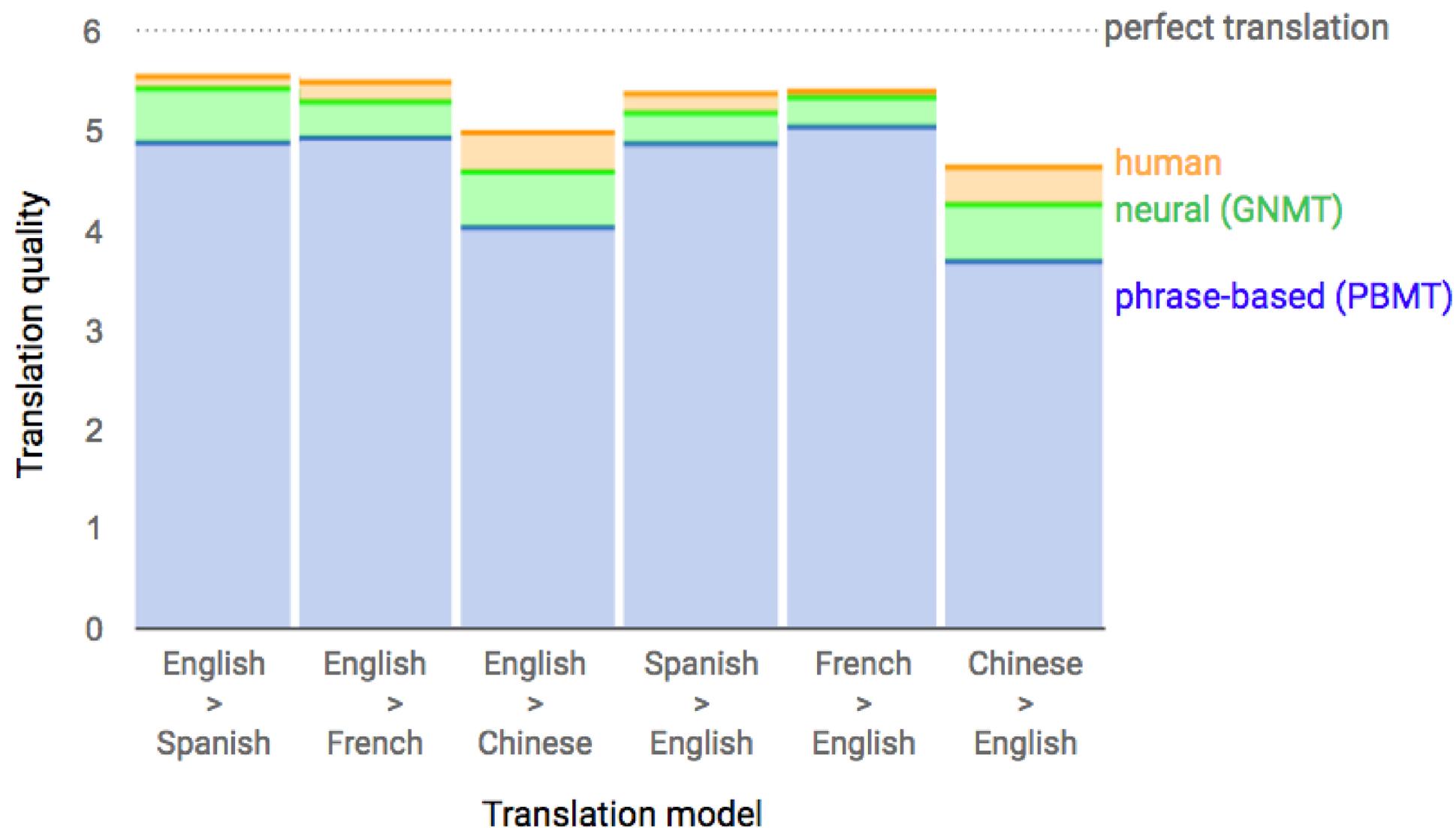
- ▶ Bidirectional Recurrent Neural Networks
- ▶ Матричные вычисления
- ▶ Тренировка модели на GPUs и TPUs (в 7-15 раз быстрее, чем CPU)
- ▶ Обучение одной модели занимает от 1 до 3 недель (SMT требует 1-3 дня)

КАЧЕСТВО ПЕРЕВОДА: ЧЕЙ BLEU SCORE ТОЛЩЕ?

Mean of side-by-side scores on production data

	PBMT	GNMT	Human	Relative Improvement						
				A>=B	RE	NM	BI	GO	NA	V8
English→Spanish	4.885	5.428	5.55							
English→French	4.932	5.295	5.496	RE		61.8	81.2	70	74.7	72.3
English→Chinese	4.035	4.594	4.987	NM	46.4		74.5	73.9	72	63.1
Spanish →English	4.872	5.187	5.372	BI	29	39		44.4	54.6	44
French →English	5.046	5.343	5.404	GO	38.2	40.2	79.3		72.1	56.1
Chinese → English	3.694	4.263	4.636	NA	34.7	40.6	73.4	51.6		44.5
				V8	35.6	49.9	77.7	68.4	79.2	

КАЧЕСТВО ПЕРЕВОДА: ЕСТЬ ЛИ ПРОРЫВ?



- ▶ Improvement всего лишь Relative
- ▶ BLEU-score +10%
- ▶ Тестирование проводилось на вполне определенных простых текстах Wikipedia

КАЧЕСТВО ПЕРЕВОДА: ЕСТЬ ЛИ ПРОРЫВ?

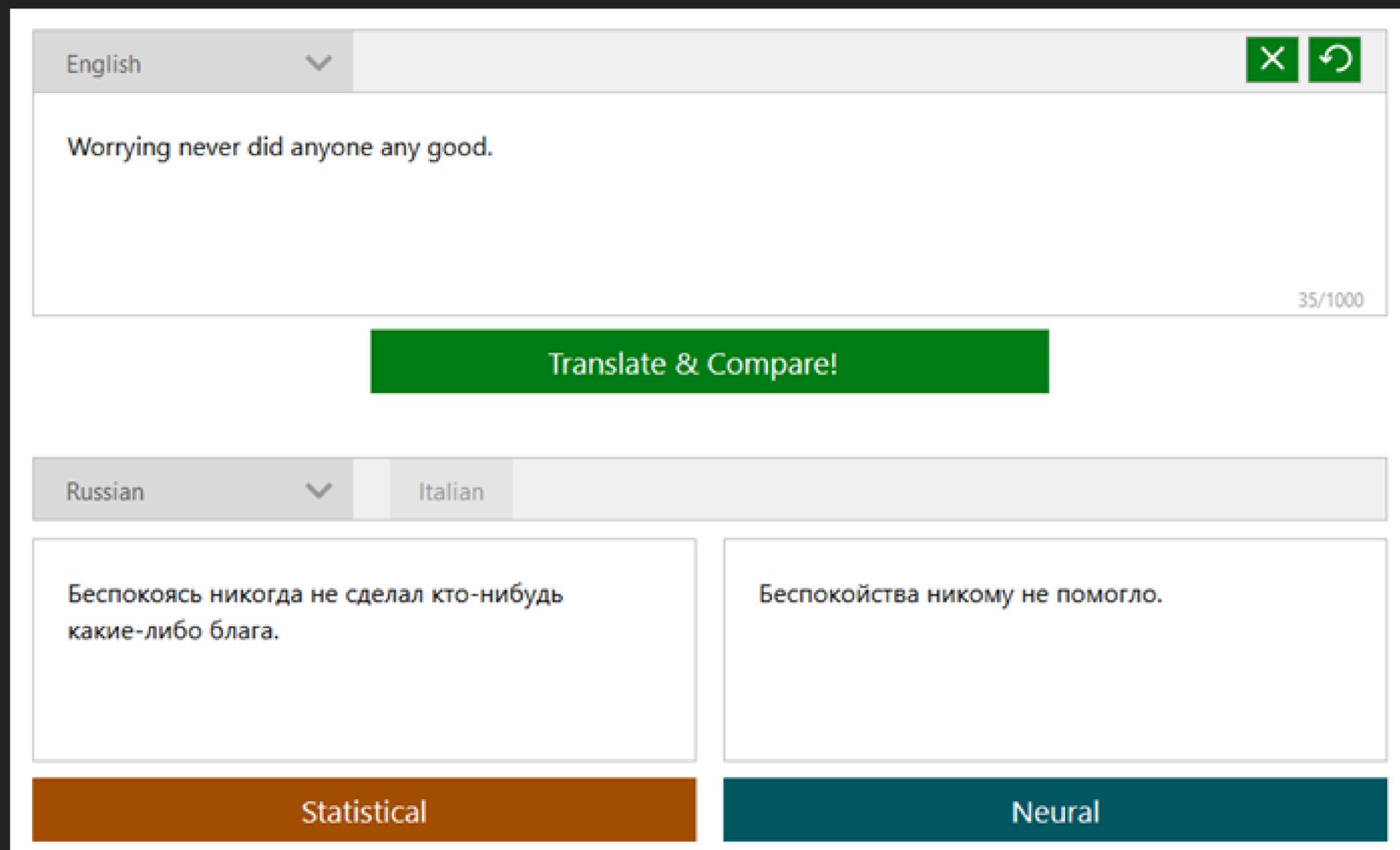
 Worrying never did anyone any good

 Не беспокоясь не делал никому ничего хорошего

 Беспокойство никогда никому не помогало

 Волнение никогда не приносило никому пользы

КАЧЕСТВО ПЕРЕВОДА: ЕСТЬ ЛИ ПРОРЫВ?

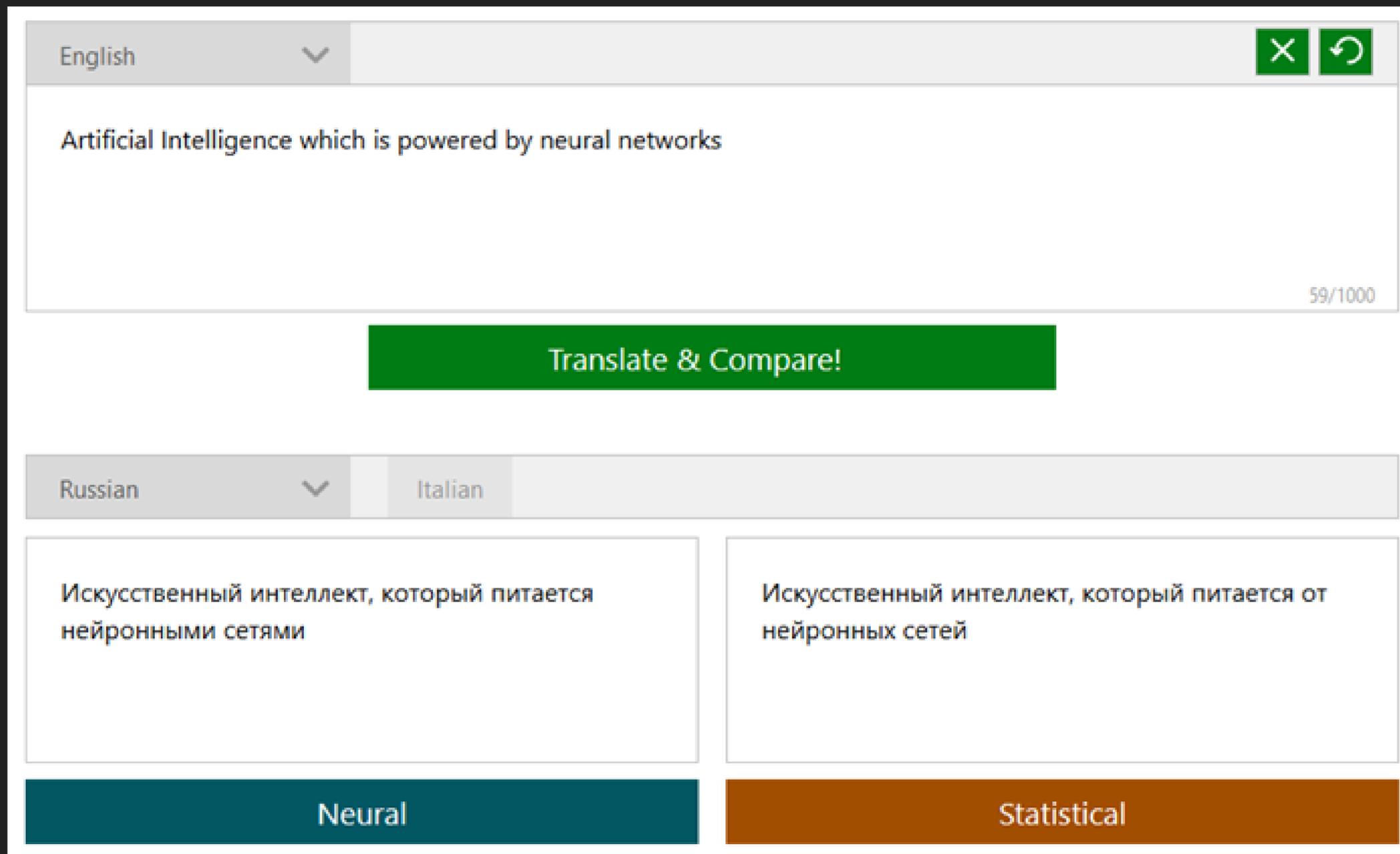


The screenshot shows the Microsoft Translator Neural interface. At the top, the source language is set to English. The input text is "Worrying never did anyone any good." Below the input is a green button labeled "Translate & Compare!". The target language is set to Russian. Two translation results are shown side-by-side: "Беспокоясь никогда не сделал кто-нибудь какие-либо блага." (Statistical) and "Беспокойства никому не помогло." (Neural).

<https://translator.microsoft.com/neural/>

Проверяйте сами!

КАЧЕСТВО ПЕРЕВОДА: ЕСТЬ ЛИ ПРОРЫВ?



The screenshot shows the PROMT translation interface. At the top, the source language is set to English. The input text is "Artificial Intelligence which is powered by neural networks". A green button labeled "Translate & Compare!" is positioned below the input. The target language is set to Russian. Below this, two columns are shown: "Neural" (teal background) and "Statistical" (orange background). The Neural translation is "Искусственный интеллект, который питается нейронными сетями", and the Statistical translation is "Искусственный интеллект, который питается от нейронных сетей".

- ▶ Прогресс есть, и он действительно заметный
- ▶ Высокий потенциал для развития технологии
- ▶ Походит не для всех типов текстов и задач

МАШИННЫЙ ПЕРЕВОД: В ЧЕМ ЗАДАЧИ

УНИВЕРСАЛЬНЫЙ ПЕРЕВОДЧИК:

- ▶ волшебная машинка из фантастических фильмов (Automagically)
- ▶ быстрый перевод слов и коротких текстов для различных целей
- ▶ автоматический перевод в процессе общения на форумах, в социальных сетях, мессенджерах
- ▶ автоматический перевод при чтении новостей, статей Wikipedia
- ▶ переводчик в путешествиях (mobile)

ПЕРЕВОДЧИК ДЛЯ БИЗНЕСА:

- ▶ деловая переписка с клиентами, партнерами, инвесторами, иностранными сотрудниками
- ▶ локализация сайтов, интернет-магазинов, описаний продуктов, инструкций
- ▶ перевод пользовательского контента (отзывы, форумы, блоги)
- ▶ интеграция перевода в бизнес-процессы и программные продукты и сервисы
- ▶ точность перевода с соблюдением терминологии, конфиденциальность и безопасность

КЕЙС: AMADEUS

- ▶ Amadeus - одна из крупнейших в мире глобальных систем дистрибуции авиабилетов
- ▶ Задача - локализация условий применения тарифов (Fare Rules), формирующихся в системе бронирования автоматически из разных источников
- ▶ Понятный перевод, передающий смысл тарифных правил, с учетом типичных терминов и аббревиатур
- ▶ Автоматический перевод интегрирован в систему бронирования Amadeus
- ▶ PROMT Cloud API

```

TST00002                                TKT-125-4616937
|
|      01.SMITH/JOHN MR
|
|      ORIGIN TRVL CURRENCY
|      OLD BASE FARE      GBP      526.00
|      NEW BASE FARE      GBP      674.00
|      FARE BALANCE
|
|      PAYMENT CURRENCY
|      OLD TAX            GBP      400.75
|      NEW TAX            GBP      502.25
|      TAX BALANCE
|
|      TICKET DIFFERENCE
|      RESIDUAL VALUE      GBP      0.00
|
|      TST ADD COLL
|      PENALTY
|      TOTAL ADD COLL
|      GRAND TOTAL
    
```

КЕЙС: AMADEUS



ROUND TRIP INSTANT PURCHASE FARES



ТАРИФЫ МГНОВЕННОЙ ПОКУПКИ РЕЙСА ТУДА И ОБРАТНО



КРУГЛЫЕ ПОКУПКИ

КЕЙС: TRIPADVISOR

- ▶ 165,600 новых отзывов ежедневно*
- ▶ Задача - перевод отзывов туристов с английского языка на русский
- ▶ Сложность: типичные особенности пользовательского контента (тексты с ошибками, опечатками, пропусками слов)
- ▶ Автоматическая оценка качества перевода перед публикацией на сайте TripAdvisor
- ▶ Технология PROMT DeepHybrid, включающая статистическое постредактирование, позволяющая получить более понятный перевод



КЕЙС: TRIPADVISOR



We ate there last night on a whim and it was a lovely meal.
The service was attentive without being over bearing.



Мы ели там в последний вечер случайно, и это была прекрасная еда. Персонал был внимательным, но не властным.



Мы ели там прошлой ночью по прихоти, и это была прекрасная еда. Сервис был внимательным, не перегружая подшипников.

В ЧЕМ ПРОБЛЕМЫ ДЛЯ БИЗНЕСА

- ▶ "Универсальный" переводчик не всегда дает приемлемое качество и не может поддерживать специфическую терминологию
- ▶ Нет достаточных объемов параллельных текстов для того, чтобы иметь возможность обучать нейронную сеть на своей тематике (100 - 500 млн. токенов)
- ▶ Нужен механизм или алгоритмы автоматической оценки качества получаемого результата
- ▶ Нужны достаточные вычислительные мощности
- ▶ Нужно быть готовым к временным затратам на обучение нейронной сети
- ▶ Непонятно, что делать с конфиденциальностью

ВЫВОДЫ

- ▶ В общем случае нейронный автоматический перевод дает результат более высокого качества, чем «чисто» статистический подход
- ▶ Автоматический перевод через нейронную сеть – лучше подходит для решения задачи «универсального перевода»
- ▶ Ни один из подходов к МП сам по себе не является идеальным универсальным инструментом для решения любой задачи перевода
- ▶ Для решения задач по переводу в бизнесе только специализированные решения могут гарантировать соответствие всем требованиям



**ВЫБИРАЙТЕ ТОТ ПЕРЕВОДЧИК,
КОТОРЫЙ БОЛЬШЕ ПОДХОДИТ
К ВАШИМ ЗАДАЧАМ**



СПАСИБО!

Борис Тихомиров

boris@promt.ru

