

Секция «Философия техники (совместно с кафедрой философии техники Дрезденского технического университета, Германия)»

**Сравнение алгоритмов машинного перевода в современных технологиях, основанного на гипотезе лингвистической относительности, и выявление лучшего из онлайн-переводчика.**

**Научный руководитель – Яковлев Алексей Александрович**

***Дудникова Варвара Эдуардовна***

*Студент (бакалавр)*

Российский университет дружбы народов, Факультет физико-математических и естественных наук, Москва, Россия

*E-mail: var.dudnikova@gmail.com*

В наше время, в эпоху компьютерных технологий, всё более востребованным становится перевод с одного языка на другой, так как большое количество информации в интернете, и не только, представлено на иностранных языках, например, иностранный фильм, инструкция какого-либо импортного товара или обычная Web-страница требуют знания. Однако далеко не каждый обладает знаниями какого-либо иностранного языка, и не желая отказываться от получения интересующей его информации, прибегает к помощи систем машинного перевода.

Хотя, качество машинного перевода оставляет желать лучшего: в большинстве случаев, текст, переведённый машиной, получается не складный, значения некоторых слов подобраны не совсем корректно.

Это можно объяснить работами Э.Сепира, который рассматривал язык как прежде всего строго организованную систему, все компоненты которой - такие, как звуковой состав, грамматика, словарный фонд, - связаны жесткими иерархическими отношениями. Связь между компонентами системы отдельно взятого языка строится по своим внутренним законам, в результате чего спроецировать систему одного языка на систему другого, не исказив при этом содержательных отношений между компонентами, оказывается невозможным. Понимая лингвистическую относительность именно как невозможность установить покомпонентные соответствия между системами разных языков, Сепир ввел термин «несоизмеримость» (incommensurability) языков.

Приведем цитату из статьи Сепира *Статус лингвистики как науки* (1928): "Два разных языка никогда не бывают столь схожими, чтобы их можно было считать средством выражения одной и той же социальной действительности".

Поэтому любой машинный перевод не переведёт текст с одного языка на другой лучше человека! Хотя и человеку это в полной мере не подвластно. Несмотря на всё это, использование электронных систем перевода увеличивается.

Информация, перевод которой требуется, с каждым днём возрастает, соответственно возрастает актуальность автоматизированных переводов. (но и это не означает, что потребность в переводе человеком уменьшается).

А сейчас перейдём к самому машинному переводу. Углубимся в историю: машинный перевод 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 изобре-

тателем 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, так и способы работы с грамматическими ролями между двумя языками на основе Эсперанто. Данная система представляла собой три этапа: первый заключался в следующем - редактор, носитель языка, должен был связать слова исходного языка (ИЯ) в логические формы в соответствии с синтаксическими функциями; на втором этапе машина должна была «перевести» все эти формы на язык перевода (ПЯ); а на третьем этапе носитель языка перевода занимался редактированием полученного результата. Его схема оставалась неизвестной до конца 50-х годов, когда появились ЭВМ.

Категории машинного перевода

Машинный перевод делится на два принципиально разных подхода к построению алгоритмов перевода: [U+202F] основанный на правилах [U+202F] (rule-based) и статистический, или [U+202F] основанный на статистике [U+202F] (statistical-based). Первый подход является традиционным и используется большинством разработчиков систем машинного перевода.

Также гипотезой лингвистической относительности занимался Б. Уорф, ученик Сепира. Уорф считал, что в концептуализацию входит не только выделяемые в тексте формальные единицы - такие, как отдельные слова и грамматические показатели, - но и избирательность языковых правил, т.е. то, как те или иные единицы могут сочетаться между собой, какой класс единиц возможен, а какой не возможен в той или иной грамматической конструкции и т.д. На этом основании Уорф предложил различать открытые и скрытые грамматические категории: одно и то же значение может в одном языке выражаться регулярно с помощью фиксированного набора грамматических показателей, т.е. быть представленным открытой категорией, а другом языке обнаруживаться лишь косвенно, по наличию тех или иных запретов, и в этом случае можно говорить о скрытой категории.

Два алгоритма машинного перевода, получается, основаны как раз на теории господина Уорфа.

Перевод в итоге сводится к тому, что информация (фраза, предложение) из одного языка и его представления в определенной модальности сжимается в представление другого языка с абсолютно другой модальностью. Под модальностью я имею в виду языковые рамки, семантические единицы, грамматика, словарный фонд, звуковой состав, категория определенности/неопределенности (переносное значение, общепринятые нормы, значения слова/фразы)

Проблемы машинного перевода связаны именно с наличием смысловой, качественной уникальности актов речевой деятельности.

Изучив структуру машинного перевода, его настоящее происхождение, в понимании гипотезы лингвистической относительности, а также после выявления наилучшего онлайн-переводчик является, важно всё-таки учитывать и помнить, что электронный переводчик создан не для того, чтобы заменить живых переводчиков, а для того чтобы помочь преодолеть языковые барьеры, обусловленные незнанием иностранных языков, поэтому лучшим переводчиком является человек, несмотря на все гипотезы и теории!

## Источники и литература

- 1) Tomović, Andrija; Janičić, Predrag; Kešelj, Vlado (2006). "n-Gram-based classification

and unsupervised hierarchical clustering of genome sequences". Computer Methods and Programs in Biomedicine. Philipp Koehn (2009). [U+202F] Statistical Machine Translation Ванников Ю.В. Языковая сложность текста как фактор трудности перевода (Методическое пособие). – М.: Всесоюзный центр переводов, 1988. Рецкер Я.И. Теория перевода и переводческая практика. – М.: Международные отношения, 1974. Федоров А.В. Введение в теорию перевода. – М.: изд-во литер. на иностр. языках, 1958. Федоров А.В. Основы общей теории перевода. – М.: Высш. шк., 1968. [http://www.promt.ru/images/ainl\\_molchanov\\_promt.pdf](http://www.promt.ru/images/ainl_molchanov_promt.pdf) [https://en.wikipedia.org/wiki/Bayesian\\_inference](https://en.wikipedia.org/wiki/Bayesian_inference) <https://www.mitpressjournals.org/doi/pdf/10.1162/coli.2006.32.4.527> <https://en.wikipedia.org/wiki/N-gram> <http://festvox.org/bsv/x1407.html> <http://web.stanford.edu/class/cs224n/syllabus.html> (<http://web.stanford.edu/class/cs224n/lectures/lecture8.pdf>) [https://ru.wikipedia.org/wiki/%D0%9C%D0%B0%D1%88%D0%B8%D0%BD%D0%BD%D1%8B%D0%B9\\_%D0%BF%D0%B5%D1%80%D0%B5%D0%B2%D0%BE%D0%B4\\_%D0%BD%D0%B0\\_%D0%BE%D1%81%D0%BD%D0%BE%D0%B2%D0%B5\\_%D0%BF%D1%80%D0%B0%D0%B2%D0%B8%D0%BB](https://ru.wikipedia.org/wiki/%D0%9C%D0%B0%D1%88%D0%B8%D0%BD%D0%BD%D1%8B%D0%B9_%D0%BF%D0%B5%D1%80%D0%B5%D0%B2%D0%BE%D0%B4_%D0%BD%D0%B0_%D0%BE%D1%81%D0%BD%D0%BE%D0%B2%D0%B5_%D0%BF%D1%80%D0%B0%D0%B2%D0%B8%D0%BB) <https://geektimes.ru/post/290133/> [https://r-datascience.ru/n-gram\\_word\\_prediction/](https://r-datascience.ru/n-gram_word_prediction/) <https://en.wikipedia.org/wiki/N-gram> Звегинцев В.А. Теоретико-лингвистические предпосылки гипотезы Сепира-Уорфа. – Новое в лингвистике, вып. I. М., 1960 Уорф Б.Л. Статьи. – Новое в лингвистике, вып. I. М., 1960 Сепир Э. Избранные труды по языкознанию и культурологии. М., 1993 Апресян Ю.Д. Интегральное описание языка и системная лексикография. М., 1995 Булыгина Т.В., Шмелев А.Д. Языковая концептуализация мира (на материале русской грамматики). М., 1997