

# Проблема эквивалентности параллельных предложений в тестовом корпусе для русско-татарского машинного переводчика

Хакимов Б.Э., Шаехов М.Р.

Институт прикладной семиотики  
Академии Наук Республики Татарстан



# ПРОБЛЕМАТИЗАЦИЯ

- ❖ Составление обучающих и тестовых корпусов для систем машинного перевода предполагает поиск **ГОТОВЫХ параллельных текстов**
- ❖ Предложения из оригинальных текстов и их опубликованных официальных переводов **считаются априори эквивалентными**, тождественными друг другу.
- ❖ Однако подробный качественный анализ таких предложений показывает, что это **не всегда соответствует** задачам разработки систем машинного перевода.



# ЭКВИВАЛЕНТНОСТЬ

- ❖ - относительная характеристика, зависящая от различных факторов, среди которых выделяются как собственно **лингвистические**, так и **экстралингвистические**.
- ❖ В.Н.Комиссаров: максимально возможная лингвистическая близость перевода и оригинала.
- ❖ Л.С. Бархударов: семантическая категория, реализующаяся в смысловом совпадении текстов на языке оригинала и текста на языке перевода.



# ЭКВИВАЛЕНТНОСТЬ В КОРПУСЕ

- ❖ Основное **отличие**: отношение к контексту на уровне целого текста или его фрагмента.
- ❖ Параллельные тестовые и обучающие корпуса представляют собой **выборку** отдельных предложений без сохранения контекстных связей.
- ❖ Многие факторы эквивалентности перестают существовать после **извлечения** этого предложения из конкретного контекста
- ❖ Предложения - отдельные независимые единицы (**микротексты**)



# ЗАТРУДНЕНИЯ



## Приблизительный, контекстуальный перевод, требующий больших трудозатрат для приведения к параллельности

Русский текст	Татарский текст
А Вахита уже обступили соседи, здоровались с ним и уверяли солдата, что они соскучились по нем.	Болар өстенә күршеләре йөгереп чыгып, үзләренең сагынуларын белдертө торган төрле-төрле сүзләр әйтеп, Вахит белән күрешә башладылар.
Вот и этот рассказ ты всегда приберегал на будущее.	Менә бу вакыйганы да Гарри языласы хикәя дип гел казага калдыра килде.

**РЕШЕНИЕ:** перемещение в резерв



## Перевод имен собственных, заключенных в кавычки: названия произведений, фирм

Русский текст	Татарский текст
То же самое - с «Алисой в Стране чудес».	Шул ук хэл - «Алиса в Стране чудес» белән.
Почетной грамотой Рослесхоза награждена мастер леса ГБУ «Бугульмалес» Любовь Хикматова.	«Бөгелмә урманы» ДБИнең урман мастеры Любовь Хикмәтова Рослесхозның Мактау грамотасы белән бүләкләнде.

**РЕШЕНИЕ:** применение общепринятых норм



## Передача имен числительных в цифровом написании с окончаниями на татарском языке

Русский текст	Татарский текст
30 марта с 09.00 до 16.00 будет произведено плановое отключение светофоров на пересечении улицы Шоссейная и улицы Герцена.	30 март көнне 09.00 дан алып 16.00 га кадәр Шоссейная һәм Герцен урамнары чатындагы светофорлар планлы рәвештә сүндерелеп торачак.
Чистые комиссионные доходы выросли на 14,7%, составив 55,4 млрд рублей.	Чиста комиссия керемнәр 14,7 %-ка үсте, 55,4 млрд сум тәшкил итте.

**РЕШЕНИЕ:** после цифр ставить пробел, после знаков - дефис



## Передача сокращений

Русский текст	Татарский текст
<p>На этой неделе в рамках реализации ФЦП «Развитие мелиорации земель сельскохозяйственного назначения России на 2014-2020 годы» ФГБУ «Управление «Татмелиоводхоз» введены две крупные оросительные системы в Тукаевском и Сабинском муниципальных районах РТ.</p>	<p>Бу атнада «2014-2020 елларга Россиянең авыл хужалыгы билгеләнешендәге жирләрэн сугаруны үстерү» федераль максатчан программасын гамәлгә ашыру барышында «Татмелиоводхоз» идарәсе» ФДБУ тарафыннан Татарстан Республикасының Саба һәм Тукай муниципаль районнарында ике зур сугару системасы файдалануга кертелде.</p>

**РЕШЕНИЕ:** принять общепринятые нормы

The background is a solid orange color. In the top-left corner, there are three vertical bars of varying heights, each composed of several overlapping semi-transparent circles. In the bottom-right corner, there are four vertical bars of varying heights, also composed of overlapping semi-transparent circles.

# Типы и примеры исправлений с точки зрения эквивалентности



## Исправление стилистической ошибки

Русский текст	Татарский текст	Исправленное предложение
<p>Убежище в январе в ЕС получили только 13 граждан Грузии, а 1288 гражданам дали отказ.</p>	<p>Гыйнварда ЕСта <b>сыенуны</b> Грузиянең 13 гражданы гына алган, ә 1288 <b>гражданга кире кагылган.</b></p>	<p>Гыйнварда ЕСта <b>сыену</b> <b>мөмкинлеген</b> Грузиянең 13 гражданы гына алган, ә 1288 <b>гражданның үтенече кире кагылган.</b></p>



## Добавление варианта при наличии в одном из языков синонима

Русский текст	Татарский текст	Исправленное предложение
Приглашать в школы героев войны.	Мәктәпкә сугыш геройларын чакырырга.	1) Мәктәпкә сугыш <b>геройларын</b> чакырырга. 2) Мәктәпкә сугыш <b>каһарманнарын</b> чакырырга.
В теории его сосчитать несложно	Теориядә аны санарга авыр түгел	1) Теориядә аны <b>санарга</b> авыр түгел 2) Теориядә аны <b>санау</b> авыр түгел



## Корректировка аффикса, присоединяемого к аббревиатуре

Русский текст	Татарский текст	Исправленное предложение
<p>28 октября 2017 года в ФГБОУ ВПО «Казанский государственный аграрный университет» состоится День открытых дверей.</p>	<p>2017 елның 28 октябрәндә «Казан дәүләт аграр университеты» югары һөнәри белем бирү <b>ФДБМУ</b>ендә Ачык ишекләр көне узачак.</p>	<p>2017 елның 28 октябрәндә «Казан дәүләт аграр университеты» югары һөнәри белем бирү <b>ФДБМУ</b>ндә Ачык ишекләр көне узачак.</p>



## Удаление слова, не имеющего прямого соответствия в параллельном предложении

Русский текст	Татарский текст	Исправленное предложение
Вспомнила Суок из сказки «Три толстяка».	“Три толстяка” экиятендәге Суокны <b>да</b> искә төшерде.	“Три толстяка” экиятендәге Суокны искә төшерде.
По словам Тяминава Артура, картофель хорошо хранится, <b>спрос на него велик</b> .	Тяминов Артур әйтүенчә, бәрәңге <b>бик</b> әйбәт саклана, аны <b>рәхәтләнеп саталар</b> , <b>Казан ашханәләренә дә илтәләр</b> икән.	Тяминов Артур әйтүенчә, бәрәңге әйбәт саклана, <b>аны теләп алалар</b> .

## Стилистическая корректировка перевода

Русский текст	Татарский текст	Исправленное предложение
Они должны за подписью руководителя присылать в Роскомнадзор бумагу со словами:	Алар җитәкче имзасы белән "Роскомнадзор" га <b>мондый сүзләр белән</b> хәбәр җибәрергә тиешләр	Алар җитәкче имзасы белән Роскомнадзорга <b>мондый эчтәлекле</b> кәгазь җибәрергә тиешләр:
События на Украине и в Крыму вновь всколыхнули разговоры о необходимости появления подобных подразделений.	Украинада һәм Кырымда мондый бүлекләрдән чыгуның кирәклеген турында сөйләшүләр яңадан <b>ишетелгәләде.</b>	Украинадагы һәм Кырымдагы вакыйгалардан соң кабат мондый хәрби бүлекләр кирәклеген турында <b>сүзләр кузгалды.</b>



## Исправление варианта перевода на контекстно-независимый

Русский текст	Татарский текст	Исправленное предложение
Бушуева <b>сидела</b> в кабине второй машины.	Бушуева икенче машинаның кабинасында <b>бара</b> иде.	Бушуева икенче машинаның кабинасында <b>утыра</b> иде.
Все знают - и никто не <b>понимает</b> .	Барысы да белә - һәм беркем <b>аңлата</b> алмый.	Барысы да белә - һәм беркем дә <b>аңламый</b> .



## Исправление грамматической формы

Русский текст	Татарский текст	Исправленное предложение
Минздрав ответил на предложение вернуть курилки <b>в аэропорты</b> .	Сәламәтлек саклау министрлыгы тәмәке тарту урыннарын <b>аэропортка</b> кайтару тәкъдименә жавап бирде.	Сәламәтлек саклау министрлыгы тәмәке тарту урыннарын <b>аэропортларга</b> кайтару тәкъдименә жавап бирде.



## Исправление оригинального текста

Русский текст	Татарский текст	Исправленное предложение
<p>Я, <u>Абузар Гиреевич</u>, знаете... после съезда впервые так остро почувствовала всю прелесть жизни... ну, и настоящее величие, красоту человека...</p>	<p>Мин, <u>Әбүзәр абзый</u>, съезддан дөньяның ямен, кешенең зурлыгын, матурлыгын тирәнрәк тоеп кайттым.</p>	<p>Мин, <u>Әбүзәр Гәрәй улы</u>, беләсезме... съезддан соң дөньяның бөтен ямен... һәм кешенең чын бөеклеген, матурлыгын беренче тапкыр шулкадәр тирән тойдым...</p>



# РЕЗУЛЬТАТЫ

- ❖ Аутентичные тексты и их переводы даже при их полной эквивалентности **требуют** значительного количества лексических, грамматических и других **трансформаций**.
- ❖ Для повышения качества и точности машинного перевода необходимо обеспечение **большей** «чистоты» и точности наборов данных.
- ❖ Необходимо соблюдение **баланса** между буквализмом и полной параллельностью

**Игътибарыгыз  
өчен  
рәхмәт!**

A decorative pattern at the bottom of the page consisting of numerous vertical bars of varying heights and shades of teal, creating a textured, bar-like effect.