

УДК 81'32

ББК 81.2

DOI: <https://doi.org/10.17308/lic/1680-5755/2025/3/6-15>

## **КОНТЕКСТУАЛЬНОЕ РАСПОЗНАВАНИЕ СЕМАНТИКИ МНОГОЗНАЧНЫХ СЛОВ СИСТЕМАМИ МАШИННОГО ПЕРЕВОДА**

**А. В. Новиков**

*Орловский государственный университет имени И. С. Тургенева*

## **CONTEXTUAL RECOGNITION OF POLYSEANTIC WORDS SEMANTICS BY MACHINE TRANSLATION SYSTEMS**

**A. V. Novikov**

*Orel State University named after I. S. Turgenev*

**Аннотация:** в рамках этой статьи рассмотрены основные принципы, на которых строится работа систем машинного перевода, а также представлен сравнительный анализ перевода текстов разных стилей, чтобы проследить, как системы машинного перевода справляются с определением контекстуального значения слова, использованного в принципиально разных контекстах. Для проведения этого исследования выбраны четыре системы машинного перевода – Google Translate, Яндекс.Переводчик, DeepL и PROMT. В начале статьи перечислены основные подходы к машинному переводу, после чего представлено краткое описание принципов работы, выбранных нами систем машинного перевода. Далее приведены примеры употребления многозначных слов в разных контекстах, а также варианты их перевода на русский язык вышеупомянутыми системами машинного перевода, после чего проведен сравнительный анализ полученных переводов. В случае неправильно распознанных значений в программу-переводчик добавлялся расширенный контекст (абзац вместо предложения), так как предполагалось, что это может улучшить распознавание контекста, и как следствие, количество случаев корректного перевода значений выбранных нами слов. Гипотезой исследования было, что расширение контекста в случае нераспознанного значения должно помочь системе его распознать. Методом сплошной выборки мы отобрали сорок предложений, в которых двадцать одних и тех же слов использованы в разных контекстах и обладают разным значением. Выбирая примеры предложений для анализа, мы не придерживались конкретных тематик. Главным принципом для отбора слов послужило наличие у них многозначности. В заключении статьи представлен вывод и изложены основные аспекты настоящего исследования. Мы считаем, что описанный в этой работе эксперимент послужит хорошим подспорьем для дальнейших исследований в этой области, в частности для понимания того, каким образом осуществляется учет контекста и как возможно улучшить работу систем машинного перевода.

**Ключевые слова:** машинный перевод, контекст, векторное представление слова, нейронная сеть, статистический перевод, исходный язык, язык перевода.

**Abstract:** this paper discusses the basic principles underlying the work of machine translation systems and presents a comparative analysis of texts of different styles to see how machine translation systems cope with finding the contextual meaning of a word used in fundamentally different contexts. Four machine translation systems, Google Translate, Яндекс.Переводчик, DeepL and PROMT were chosen for this study. The article presents main approaches to machine translation, a brief description of the operation principles of the four mentioned above machine translation systems, examples of using the same words in different contexts, as well as variants of their translation into Russian by the above-mentioned machine translation systems. In case of incorrectly recognized meanings an extended context (paragraph instead of sentence) was added to the translator



program, as it was assumed that this could improve context recognition and, as a consequence, the number of correct translations of different meanings of the selected words. The hypothesis was that expanding the context in case of an unrecognized meaning should help the system to recognize it. Using a continuous sampling method we have selected forty sentences in which twenty of the same words but with different meaning are used in different contexts. Selecting the sample sentences for analysis we did not stick to specific topics. The main principle of selecting the words presented in the article was polysemy. At the end of the article we discuss the comparative analysis results and summarize main aspects of this study. We believe that the experiment described in this paper can serve as a good reference for further research in this area, in particular to understand how context is taken into account and how machine translation systems can be improved.

**Key words:** machine translation, context, contextual word meaning, neural network, statistical translation, source language, target language.

## Введение

Активное развитие программ машинного перевода позволяет быстро выполнять обработку больших массивов текстов на иностранных языках в максимально короткие сроки и при минимальных финансовых затратах, что является одним из основных требований сегодняшней жизни.

Одной из ключевых проблем перевода является многозначность большого количества слов, и, соответственно, выбор правильного варианта перевода таких слов по контексту. Если говорить о системах машинного перевода, необходимо отметить, что в некоторых случаях такие системы испытывают трудности с определением тематики текста, что ведет к неправильному толкованию многозначных слов и, как следствие, подбору неподходящего контекстуального значения на ПЯ к слову на ИЯ [1, с. 196]. Стоит отметить, что, осуществляя поиск литературы по вышеописанной проблеме, мы пришли к выводу, что этот аспект в исследовании машинного перевода недостаточно разработан, в связи с чем мы решили изучить и проанализировать его посредством проведения эксперимента, суть которого будет описана ниже. Таким образом, **актуальность** исследования обуславливается всем вышесказанным.

Цель нашего исследования состоит в проведении сравнительно-сопоставительного анализа переводов, выполненных системами машинного перевода и определении того, насколько успешно современные машинные переводчики распознают значение одного и того же слова в зависимости от его окружения – контекста. Мы сравнили работу четырех распространенных систем машинного перевода Google Translate, Яндекс. Переводчик, DeepL и PROMT.

Перед началом эксперимента необходимо было понять, на каких базовых принципах основывается работа данных систем.

Существует четыре самых распространенных типа систем машинного перевода:

- 1) основанные на правилах – Rule-Based Machine Translation, или RBMT;
- 2) опирающиеся на статистические принципы – Statistical Machine Translation, или SMT;

3) работающие на базе нейронных сетей – Neural Machine Translation, или NMT;

4) гибридные системы машинного перевода, т. е. системы, включающие в себя сразу несколько, чаще всего два, подхода к машинному переводу [2].

## Принцип работы системы Google Translate

До 2016 г. система Google Translate была построена исключительно на статистическом принципе, т. е. на поиске соответствий на языке перевода (ПЯ) для текста на исходном языке (ИЯ) и использовании огромного массива данных, который содержит в себе слова, ранее вносимые пользователем в процессе перевода [3].

В 2016 г. компания Google представила полностью модернизированную систему нейронного машинного перевода – Google Neural Machine Translation (GNMT), принцип работы которой основывается на использовании искусственной нейронной сети, что позволило заметно улучшить качество перевода. В отличие от статистического метода, при котором осуществление трансформации текста производилось пословно с учетом лишь грамматики, GNMT работает с целым предложением и учитывает контекст. В процессе преобразования текста с ИЯ на ПЯ каждое «предложение разбивается на словарные сегменты, после чего посредством специальных декодеров программа устанавливает “вес” каждого сегмента текста» [3]. Затем производятся вычисления наиболее вероятных значений и осуществляется перевод сегментов текста. На финальном этапе переведенные сегменты складываются воедино с учетом грамматических правил [3].

«GNMT базируется на принципе работы рекуррентных двунаправленных нейронных сетей, благодаря которому система способна распознавать значение слова или фразы на основе предыдущих значений в последовательности, что позволяет системе учитывать контекст и выбирать необходимое значение слова из разных вариантов его перевода. Двунаправленность означает, что нейросеть разделена на два потока – анализирующий и синтезирующий. Каждый поток состоит из восьми слоев, которые проводят векторный анализ. Первый поток разбивает предло-

жение на смысловые элементы и анализирует их, а второй вычисляет наиболее вероятный вариант перевода, исходя из контекста и модулей внимания. Важно то, что анализирующая сеть “читает” предложение не только слева направо, но и справа налево, что позволяет в полной мере понять контекст» [3].

### **Принцип работы системы Яндекс.Переводчик**

Если говорить о принципе функционирования Яндекс.Переводчика, то он основан на гибридной модели машинного перевода, которая сочетает в себе два подхода – нейросетевой и статистический. В Яндекс.Переводчике статистический подход реализован следующим образом: в процессе перевода предложения на ИЯ членятся на слова и словосочетания, которые переводятся отдельно, так как для каждого из них в матрице подбирается возможный вариант на ПЯ, после чего система сравнивает несколько полученных вариантов одного предложения и, опираясь на статистику, выбирает лучший вариант, основываясь на сочетаемости слов в естественном языке. «Преимущество статистического подхода состоит в способности запоминать и выдавать перевод коротких, устоявшихся словосочетаний и редких слов; однако есть и существенный недостаток, который заключается в отсутствии каких-либо связей между фразами, т. е. контекст при переводе совсем не учитывается» [4]. Однако так как Яндекс.Переводчик является гибридной системой, для компенсации этого недостатка используется принцип нейросетей.

Принцип нейронного машинного перевода в Яндекс.Переводчике, так же как и статистический принцип, основывается на анализе массива параллельных текстов, поиске в них закономерностей и создании картотеки всех используемых в естественном языке слов и словосочетаний. Однако в отличие от статистического подхода, в нейросетевом существует понятие векторного представления слова (*word embedding*), т. е. чисел, которые характеризуют слово с точки зрения семантики. В процессе перевода каждое предложение на ИЯ переводится целиком, а не делится на отдельные слова и словосочетания, как это происходило при статистическом подходе. Значительным преимуществом принципа нейронного машинного перевода является возможность учитывать контекст, благодаря чему достигается связность текста перевода. Недостаток этого принципа связан с дефицитом информации по нечасто встречающимся словам, например, малораспространенным именам собственным или топонимам, для которых система не смогла создать оптимальное векторное представление [4].

Что касается совместной работы вышеописанных подходов, выбора варианта перевода и оценки его качества: после ввода текста на ИЯ в систему, Яндекс.

Переводчик передает этот текст и нейронной сети, и статистическому переводчику. Затем выданный обеими системами результат оценивается алгоритмом, который основывается на методе обучения CatBoost – алгоритме машинного обучения, который использует градиентное усиление на деревьях решений. При оценивании принимаются во внимание множество аспектов – начиная с длины предложения, и заканчивая синтаксисом. После произведения оценки обоих вариантов перевода по всем аспектам алгоритм выбирает лучший и показывает текст на ПЯ пользователю [4].

### **Принцип работы системы DeepL**

Если говорить о системе машинного перевода DeepL, то она функционирует на основе нейросетей. В такие сети внедряется большое количество уже существующих верных переводов, и на основе загруженных текстов система машинного перевода DeepL осуществляет поиск связей между ними. Благодаря этой процедуре система определяет, каким образом должны переводиться новые введенные в нее тексты, что и является процессом «машинного обучения». Для «обучения» системы машинного перевода DeepL используются специальные краулеры (crawlers) – поисковые роботы, применяемые для нахождения новых страниц в Интернете. Такие краулеры системы машинного перевода DeepL осуществляют целенаправленный автоматический поиск особых обучающих данных, а именно – переводов, а также производят оценку их качества. Кроме этого, для обучения машинного переводчика DeepL используется метод обучения с учителем, в процессе которого нейронная сеть выдает различные примеры перевода и каждый раз сравнивает свои собственные переводы с переводами из набора обучающих данных. В случае обнаружения расхождений между переводом, выполненным системой, и уже имеющимся переводом в корпусе данных вносятся соответствующие исправления в текст на ПЯ, который выдала система машинного перевода. Это означает, что в процессе обучения DeepL используется постредактирование, т. е. текст на ИЯ преобразуется машиной, а человек в свою очередь вносит определенные корректировки в выданный системой результат – текст на ПЯ [5].

### **Принцип работы системы PROMT**

Принцип работы системы машинного перевода PROMT сегодня основан на гибридном подходе. Изначально система машинного перевода PROMT была построена на базе принципа, основанного на правилах – Rule-based machine translation (RBMT), в основе которого лежало лингвистическое описание двух естественных языков, т. е. обширные базы данных, двуязычные словари, которые располагают морфоло-

гической, грамматической и семантической информацией, а также формальные грамматики и алгоритмы перевода. Качество выполненного перевода зависело от полноты словарей и глубины описания естественных языков. В системе машинного перевода PROMT принцип RBMT позже был модернизирован благодаря внедрению в него семантических сетей, и был назван аналитическим машинным переводом PROMT. В системе машинного перевода PROMT также применялся статистический подход, основу которого составляли корпуса параллельных текстов. При таком подходе качество перевода напрямую зависело от количества данных, также важную роль играли их полнота и качество. В 2019 г. компания PROMT представила усовершенствованную технологию машинного перевода, которая базируется на нейронных сетях – PROMT Neural. Сегодня PROMT представляет собой систему машинного перевода, построенную на базе гибридного принципа и включающую в себя нейросетевой подход – Neural Machine Translation (NMT) и подход на основе правил (RBMT). Алгоритмы нейронной сети PROMT осуществляют анализ текста и выбирают, какой из вышеперечисленных принципов лучше подходит для трансформации конкретного фрагмента текста. Этот процесс выбора системой PROMT подходящего подхода для перевода способствует получению наиболее качественного результата [6].

### **Практическая часть исследования: сравнение переводов, выполненных системами машинного перевода Google, Яндекс, DeepL и PROMT**

Суть нашего эксперимента состоит в следующем: методом сплошной выборки мы отобрали сорок предложений, в которых двадцать одних и тех же слов использованы в разных контекстах и, следовательно, обладают разным значением. Выбирая примеры предложений для анализа, мы не придерживались конкретных тематик, поэтому в настоящем исследовании представлены фрагменты текстов на самые разные темы, такие как медицина, спорт, военное дело, автомобилестроение и др. Целью было сравнить и понять, как программы машинного перевода обучены разным предметным сферам. Главным принципом для отбора конкретных слов, представленных ниже, послужило наличие у них многозначности. Для проверки и уточнения возможных контекстуальных значений слов мы использовали электронные словари: англо-английский Cambridge Online Dictionary и переводной онлайн-словарь Multitran. После отбора предложений на английском языке для сравнительно-сопоставительного анализа их перевода на русский язык мы прогнали их через выбранные нами системы машинного перевода и посчитали количество

правильно распознанных случаев перевода значений. Если значения распознавались неправильно, в программу-переводчик был добавлен расширенный контекст (абзац вместо предложения). Предполагалось, что это может улучшить распознавание значения слова и, как следствие, количество случаев корректного перевода выбранных нами слов. Гипотезой исследования было, что расширение контекста в случае нераспознанного значения должно помочь системе его распознать.

**Научная новизна** нашего исследования состоит в том, что ранее проблема распознавания контекстуального значения одного и того же слова системами машинного перевода с учетом расширения контекста не подвергалась глубокому изучению. В связи с этим, говоря о научно-практической значимости этого исследования, мы считаем, что описанный выше эксперимент может служить хорошим подспорьем для дальнейших исследований в этой области, в частности для понимания того, каким образом осуществляется учет контекста и как возможно улучшить работу систем машинного перевода.

Перейдем непосредственно к эксперименту по сравнению четырех вышеописанных систем машинного перевода. В данной статье представлены не все примеры, что объясняется требованиями к ее объему. Таблицы с итоговой статистикой даются в конце статьи.

Рассмотрим слово “pocket” в двух абсолютно разных контекстах. В представленных ниже предложениях оно должно быть переведено: в первом случае как «луз» – термином из бильярда, а во втором случае – военным термином «котел» или «окружение». В итоге получились следующие варианты переводов:

Оригинал: Sean Murphy is snookered behind the yellow on most balls, which forces him into playing on the red sitting just past the left middle pocket [7].

Google: Шон Мёрфи находится в снукере за желтым на большинстве шаров, что вынуждает его играть на красном, сидячем сразу за левой средней лузой.

Яндекс: Шон Мёрфи в большинстве мячей находится за желтой линией, что вынуждает его играть на красной, расположенной сразу за левой средней лузой.

DeepL: На большинстве шаров Шон Мерфи оказывается за желтым, что вынуждает его играть на красном, сидячем сразу за левой средней лузой.

PROMT: Шон Мерфи снует за желтым на большинстве мячей, что заставляет его играть на красном, сидячем прямо мимо левого среднего кармана.

Из приведенных выше примеров видно, что системы машинного перевода Google, Яндекс и DeepL справились с поставленной задачей – правильно распознали тематику предложения и перевели английское слово “pocket” контекстуально подходящим эк-

вивалентом на русский язык. Однако системой PROMT контекстуальное значение слова “*pocket*” подобрано при переводе неверно и не соответствует данному контексту. В связи с этим мы расширили контекст предложения до абзаца, предположив, что системе машинного перевода, базирующейся на принципе нейронных сетей, может понадобиться более широкий контекст для генерирования эквивалентного и адекватного перевода. В итоге мы получили следующий результат.

Оригинал (расширенный контекст): He's snookered behind the yellow on most balls, which forces him into playing on the red sitting just past the left middle pocket. The direct approach won't do, however, so Murphy uses his excellent knowledge of side and angles to come off the left cushion first with lots of right-hand side, causing the white to bounce back towards the left cushion and nudge the red into the middle pocket [7].

PROMT (расширенный контекст): Он снует за желтым на большинстве мячей, что заставляет его играть на красном, сидящем прямо за левым средним карманом. Прямой подход, однако, не сделает, поэтому Мерфи использует свои отличные знания стороны и углов, чтобы сойти с левой подушки сначала с большим количеством правой стороны, заставляя белых отскочить к левой подушке и подтолкнуть красный в средний карман.

Из полученного варианта перевода следует, что расширение контекста предложения до абзаца не дало ожидаемого результата, так как системе машинного перевода PROMT также не удалось распознать тематику абзаца и выдать необходимое контекстуальное значение на русском языке – «луга».

Перейдем к рассмотрению второго примера со словом “*pocket*”:

Оригинал: A deep salient is vulnerable to being "pinched out" across the base, forming a pocket in which the defenders of the salient become isolated [8].

Google: Глубокий выступ уязвим для «зажемления» основания, образуя карман, в котором защитники выступа оказываются изолированными.

Яндекс: Глубокий выступ уязвим для того, чтобы быть «зажатым» поперек основания, образуя карман, в котором защитники выступа оказываются изолированными.

DeepL: Глубокий выступ уязвим для того, чтобы быть «зажатым» по всей базе, образуя карман, в котором защитники выступа оказываются изолированными.

PROMT: Глубокий отблеск уязвим для того, чтобы быть «зажатым» по основанию, образуя карман, в котором изолируются защитники отблеска.

Из представленных примеров видно, что в этом предложении все исследуемые нами системы машинного перевода перевели военный термин “*pocket*” как

«карман», что является неверным переводом, не соответствующим данному контексту, так как в русском языке есть специальный военный термин «котел» или «окружение».

Исходя из предположения, что системам машинного перевода, построенным на основе нейронных сетей, может быть нужен более широкий контекст, чтобы выдать эквивалентный и адекватный перевод, мы расширили контекст и дали возможность обеим системам машинного перевода перевести весь абзац, откуда было взято это предложение. В итоге получили следующий результат:

Оригинал (расширенный контекст): A salient is a battlefield feature that projects into enemy territory. The salient is surrounded by the enemy on three sides, making the troops occupying the salient vulnerable. The enemy's line facing a salient is referred to as a re-entrant (an angle pointing inwards). A deep salient is vulnerable to being "pinched out" across the base, forming a pocket in which the defenders of the salient become isolated [8].

Google: Выступ – это особенность поля боя, которая проецируется на вражескую территорию. Выступ окружен противником с трех сторон, что делает войска, занимающие выступ, уязвимыми. Линия противника, обращенная к выступу, называется повторным входом (угол, направленный внутрь). Глубокий выступ уязвим для «зажемления» основания, образуя карман, в котором защитники выступа оказываются изолированными.

Яндекс: Выступ – это объект на поле боя, который выступает на вражескую территорию. Выступ окружен противником с трех сторон, что делает войска, занимающие выступ, уязвимыми. Линия противника, обращенная к выступу, называется обратным входом (угол, направленный внутрь). Глубокий выступ уязвим для того, чтобы быть «зажатым» поперек основания, образуя карман, в котором защитники выступа оказываются изолированными.

DeepL (расширенный контекст): Залп – это участок поля боя, выступающий на территорию противника. Салиент окружен противником с трех сторон, что делает войска, занимающие салиент, уязвимыми. Линия противника, обращенная к выступу, называется ре-ентрантом (угол, направленный внутрь). Глубокий выемка уязвим для того, чтобы быть «отжатым» по всей базе, образуя карман, в котором защитники выемки оказываются изолированными.

PROMT (расширенный контекст): Заметной является особенность поля боя, которая проецируется на вражескую территорию. Видное окружено противником с трех сторон, что делает войска, занявшие видное, уязвимыми. Линия противника, обращенная к явному, упоминается как повторный вход (угол, направленный внутрь). Глубокий отблеск уязвим для того, чтобы быть «зажатым» по основа-

нию, образуя карман, в котором изолируются защитники отблеска.

Полученный перевод показывает, что расширение контекста снова не дало результата. Все четыре системы машинного перевода не смогли подобрать подходящий эквивалент на русском языке английскому военному термину “*rocket*” ни на уровне предложения, ни на уровне абзаца.

Рассмотрим пример со словом gauge. В первом предложении это слово необходимо перевести термином, связанным с железнодорожным транспортом – «ширина колеи» или «колея». Во втором предложении то же слово “*gauge*” должно переводиться как «манометр» – термином, связанным с нефтегазовым или насосным производством.

Оригинал: Because the cost of construction and operation of a rail line is greater or less depending on the gauge, much controversy has surrounded decisions in respect to it, and a proliferation of gauges has developed throughout the world [9].

Google: Поскольку стоимость строительства и эксплуатации железнодорожной линии больше или меньше в зависимости от ширины колеи, решения в отношении нее вызывают много споров, и во всем мире наблюдается распространение ширины колеи.

Яндекс: Поскольку стоимость строительства и эксплуатации железнодорожной линии больше или меньше в зависимости от колеи, решения в отношении нее вызвали много споров, и по всему миру появилось множество колей.

DeepL: Поскольку стоимость строительства и эксплуатации железнодорожной линии больше или меньше в зависимости от ширины колеи, решения в отношении нее вызывают много споров, и в мире появилось большое количество калибров.

PROMT: Поскольку стоимость строительства и эксплуатации железнодорожной линии в большей или меньшей степени зависит от колеи, многие споры окружают решения в отношении нее, и по всему миру появилось множество датчиков.

Из этого примера мы видим, что Google Translate и Яндекс.Переводчик справились с определением контекстуального значения слова “*gauge*” и подобрали подходящие к нему эквиваленты на русском языке – «ширина колеи» и «колея». Если говорить о переводах, выполненных системами DeepL и PROMT, то в первой части предложения обе системы машинного перевода справились с подбором необходимого контекстуального значения к английскому слову “*gauge*”. Однако во второй части предложения им это не удалось. Обе системы в рамках одного предложения перевели одно и то же слово по-разному (во втором случае употребления – неверно: DeepL перевел слово “*gauges*” как «калибров», а PROMT – как «датчиков»). В связи с этим мы также решили рас-

ширить контекст предложения до абзаца для обеих систем. В результате получился следующий перевод:

Оригинал (расширенный контекст): Because the cost of construction and operation of a rail line is greater or less depending on the gauge, much controversy has surrounded decisions in respect to it, and a proliferation of gauges has developed throughout the world. A narrow gauge has, in addition to cost advantages, a capability for sharper curvature; among its disadvantages are reduced lateral stability and consequent loss of operating speed [9].

DeepL (расширенный контекст): Поскольку стоимость строительства и эксплуатации железнодорожной линии больше или меньше в зависимости от ширины колеи, решения, принимаемые в отношении нее, вызывают много споров, и во всем мире появилось множество видов ширины колеи. Узкая колея, помимо преимуществ в стоимости, имеет возможность более резкого искривления; к ее недостаткам относится снижение боковой устойчивости и, как следствие, потеря рабочей скорости.

PROMT (расширенный контекст): Поскольку стоимость строительства и эксплуатации железнодорожной линии в большей или меньшей степени зависит от колеи, многие споры окружают решения в отношении нее, и по всему миру появилось множество датчиков. Узкая колея имеет, помимо ценовых преимуществ, возможность более резкой кривизны; среди его недостатков – снижение поперечной устойчивости и последующая потеря рабочей скорости.

Из полученных переводов мы видим, что в случае с системой машинного перевода DeepL расширение контекста до абзаца помогло ей распознать английское слово “*gauges*” во втором случае в первом предложении абзаца и подобрать верное контекстуальное значение на русском языке – «ширины колеи» вместо «калибров». В случае с системой машинного перевода PROMT мы видим, что расширение контекста не дало положительного результата, так как система не смогла подобрать необходимое значение к английскому слову “*gauges*” во втором случае его употребления в первом предложении абзаца и снова перевела его как «датчики».

Рассмотрим второй пример со словом “*gauge*”:

Оригинал: Gauges are required for the set-up and tuning of fluid power machines and are indispensable in troubleshooting them [10].

Google: Манометры необходимы для установки и настройки гидравлических силовых машин и незаменимы при устранении их неисправностей.

Яндекс: Датчики необходимы для настройки гидравлических силовых машин и незаменимы при их устранении.

DeepL: Манометры необходимы для настройки и регулировки жидкостных силовых машин и незаменимы при поиске и устранении неисправностей.

PROMT: Манометры необходимы для настройки и настройки гидравлических силовых машин и необходимы для их устранения.

Если говорить о втором примере, то здесь мы наблюдаем следующую картину: Google Translate, DeepL и PROMT правильно определили тему предложения и в соответствии с ней подобрали контекстуально подходящий эквивалент на русском языке английскому термину “gauge” – перевели его как «манометр». В случае с Яндекс.Переводчиком необходимо отметить, что подобранное этой системой машинного перевода контекстуальное значение – «датчик» – не подходит данному контексту, так как оно является слишком общим и не дает нам понятия, о каком именно «датчике» идет речь в предложении. Однако стоит обратить внимание на то, что если кликнуть на слово «датчики» в полученном с помощью Яндекс.Переводчика предложении на русском языке, то мы увидим, что вторым вариантом перевода термина “gauge” будет «манометр». Из этого мы можем сделать вывод, что система машинного перевода Яндекс.Переводчик имела в своем словаре подходящий вариант перевода для настоящего контекста, но не смогла осуществить выбор. Кроме этого, помимо неверного перевода слова “gauge” Яндекс.Переводчик допустил ошибку при передаче значения сложносоставного слова “troubleshooting”, в результате чего полностью нарушился смысл целого предложения. В связи с этим, так же как и со словом “rocket”, мы решили дать Яндекс.Переводчику более широкий контекст, чтобы посмотреть, сможет ли он подобрать подходящее контекстуальное значение на русском языке к английскому термину “gauge”. Результат получился следующий.

Оригинал (расширенный контекст): A pressure gauge is a fluid intensity measurement device. Gauges are required for the set-up and tuning of fluid power machines and are indispensable in troubleshooting them. Without pressure gauges, fluid power systems would be both unpredictable and unreliable. Gauges help to ensure there are no leaks or pressure changes that could affect the operating condition of the hydraulic system [10].

Яндекс: Манометр – это устройство для измерения плотности жидкости. Датчики необходимы для настройки гидравлических силовых машин и незаменимы при их устранении. Без манометров гидравлические системы питания были бы непредсказуемыми и ненадежными. Датчики помогают убедиться в отсутствии утечек или перепадов давления, которые могли бы повлиять на рабочее состояние гидравлической системы.

Из полученного перевода можно сделать вывод, что расширенный контекст не помог системе машинного перевода Яндекс.Переводчик определить необходимое контекстуальное значение на русском языке английского слова “gauge”. Однако мы обратили внимание на то, что в двух предложениях оригинального текста было использовано не просто слово “gauge”, а словосочетание “pressure gauge”. В этом случае Яндекс.Переводчик справился с подбором контекстуального значения на русском языке и перевел его как «манометр». Из этого следует, что здесь система «оттолкнулась» от словосочетания, чему, вероятно, способствовал статистический модуль, и благодаря этому выдала необходимое контекстуальное значение.

Таким же образом мы проанализировали еще семнадцать примеров других слов, которые также имеют разные значения в зависимости от контекста. Результаты эксперимента представлены в табл. 1–2.

Таблица 1

Результаты экспериментального исследования (Google Translate и Яндекс.Переводчик)

№ п/п	Слово	Google Translate				Яндекс.Переводчик			
		предложения: № 1 / № 2		абзац		предложения: № 1 / № 2		абзац	
1	2	3	4	5	6	7	8	9	10
1	pocket	+	–		–	+	–		–
2	gauge	+	+			+	–		–
3	jaw	+	–		–	+	–		–
4	socket	+	+			+	–		–
5	vessel	+	+			+	+		
6	capacity	+	+			+	+		
7	spring	+	+			+	+		
8	bed	+	+			+	+		
9	cell	+	+			+	+		
10	lens	+	+			+	+		
11	frame	+	+			+	+		

Окончание табл. 1

1	2	3	4	5	6	7	8	9	10
12	crown	+	+			+	+		
13	shell	+	+			+	+		
14	torque	+	+			+	-		-
15	gear	+	+			+	+		
16	calf	+	+			+	+		
17	tap	+	+			+	+		
18	bar	+	+			+	+		
19	ratio	+	+			+	+		
20	hearing	+	+			+	+		

Таблица 2

## Результаты экспериментального исследования (DeepL и PROMT)

№ п/п	Слово	DeepL				PROMT			
		предложения: № 1 / № 2		абзац		предложения: № 1 / № 2		абзац	
1	pocket	+	-		-	-	-	-	-
2	gauge	+/-	+	+/-		+/-	+	+/-	
3	jaw	+	-		-	+	-		-
4	socket	+	+			+	-		-
5	vessel	+	+			+	+		
6	capacity	+	+			+	+		
7	spring	+	+			+	+		
8	bed	+	+			+	-		-
9	cell	+	+			+	+		
10	lens	+	+			+	-/+		-/+
11	frame	+	+			+	+		
12	crown	+	+			+	-		-
13	shell	+	+			+	+		
14	torque	+	+			+	-		-
15	gear	+	+			+	+		
16	calf	+	+			-	+		-
17	tap	+	+			-	+		-
18	bar	-	+	+		-	-/+	-	-/+
19	ratio	+	-		+	+	-		+
20	hearing	+	+			+	+		

В соответствии с табл. 1 мы получаем следующие цифры: Google справился с 38 предложениями из 40, что составляет 95 %, а Яндекс справился с 35 предложениями из 40, что составляет 87,5 %. Стоит отметить, что в случае с этими системами расширение контекста до абзаца для получения корректного варианта перевода многозначного слова в случае с Google понадобилось для 2 предложений, а с Яндексом – для 5 предложений, и ни в одном из случаев не дало результата. Судя по результатам в табл. 2 мы видим, что DeepL справился с 38 предложениями из 40, что составляет 95 %. С 3 из 5 предложений он

справился после расширения контекста предложения до абзаца, что составляет 60 %. PROMT справился с 27 предложениями из 40, что составляет 67,5 %. С 1 из 14 предложений он справился после расширения контекста до абзаца, что составляет 7 %.

## Вывод

Исходя из этих данных, мы делаем вывод, что все четыре рассматриваемые нами системы машинного перевода в целомправляются с определением контекста и подбором к одному и тому же слову на ИЯ необходимого эквивалента на ПЯ, однако Google и

DeepL справились с подбором контекстуального значения немного лучше, чем Яндекс и PROMT.

Тем не менее существуют случаи, когда возникают определенные трудности с подбором системой необходимого перевода, когда даже после ввода в систему машинного перевода расширенного контекста обеим системам не удается распознать необходимое контекстуальное значение слова. Такой пример можно наблюдать во втором предложении со словом “pocket”. Это может быть обусловлено тем, что слово “pocket” в данном значении является довольно специфичным, что подтверждает отсутствие такой его дефиниции в англо-английских словарях, и следовательно, статистически не так часто употребимо. Кроме этого, военный контекст может быть менее частотным, т. е. встречаться реже, чем спортивный, юридический и т. д. Соответственно и нейросеть не обучается / меньше обучается его распознаванию. Кроме этого, важно отметить, что те случаи, в которых системы машинного перевода не справились с подбором необходимого контекстуального значения, тематически были разные. Полагаем, что системы Google и Яндекс не реагируют на расширенный контекст, а работают только в пределах предложения. Также необходимо отметить, что во втором примере со словом “gauge” система машинного перевода Яндекс.Переводчик неточно перевела этот термин на русский язык, однако в возможных вариантах перевода присутствовало необходимое значение, что позволяло выбрать именно его. В случаях с системами машинного перевода DeepL и PROMT мы наблюдали следующее: расширение контекста предложения до абзаца помогло системе DeepL корректно подобрать значение слова в трех случаях из пяти изначально неверно переведенных слов, а системе PROMT – только в одном случае из четырнадцати. В соответствии с полученными данными мы можем предположить, что в отличие от систем Google и Яндекс, системам DeepL и PROMT расширение контекста с предложения до абзаца в некоторых случаях может способствовать правильноому переводу многозначных слов.

## ЛИТЕРАТУРА

1. Новиков А. В. Основные подходы к преобразованию текста в системах машинного перевода // Актуальные вопросы лингвистики и лингводидактики в контексте межкультурной коммуникации : сб. материалов III Всерос. науч.-практ. онлайн-конф. (30 марта 2023 г.) / под ред. О. Ю. Ивановой. Орел : ОГУ им. И. С. Тургенева, 2023. С. 195–203.

2. Козина А. В., Черепков Е. А., Белов Ю. С. Обзор подходов к машинному переводу // Междунар. студ. науч. вестник. 2018. № 6. URL: <https://eduherald.ru/ru/article/view?id=19408>

3. Как работает нейросеть Google Translate. URL: <https://www.cossa.ru/152/196086/>

4. О машинном переводе. Руководство разработчика. URL: <https://yandex.ru/dev/translate/doc/dg/concepts/how-works-machine-translation.html>

5. Как работает DeepL Переводчик. URL: <https://www.deepl.com/ru/blog/how-does-deepl-work>

6. Молчанов А. Статистические и гибридные методы перевода в технологиях компании PROMT // Control Engineering Россия – Август 2013. URL: <https://controlengrussia.com/innovatsii/statisticheskie-i-gibridnye-metody-perevoda-v-tehnologiyah-kompanii-promt>

7. Top 10 Snooker Shots of All Time: Jaw Dropping, Pure Magic, Surreal. URL: <https://scorum.com/en-us/snooker/@jodcarey/top-10-snooker-shots-of-all-time-jaw-dropping-pure-magic-surreal>

8. Salients, re-entrants and pockets. URL: [https://military-history.fandom.com/wiki/Salients,\\_re-entrants\\_and\\_pockets](https://military-history.fandom.com/wiki/Salients,_re-entrants_and_pockets)

9. Britannica, Gauge. Railroad track. URL: <https://www.britannica.com/technology/gauge-railroad-track>

10. Understanding Pressure Gauge. URL: <https://studentlesson.com/definition-uses-diagram-types-working-of-a-pressure-gauge/>

## REFERENCES

1. Novikov A. V Basic approaches to text transformation in machine translation systems. In: *Current issues of linguistics and linguodidactics in the context of intercultural communication: proceedings of the III All-Russian Scientific and Practical Online Conference (March 30, 2023)*. Edited by O. U. Ivanova. Orel: OSU named after I. S. Turgenev, 2023. Pp. 195–203.

2. Kozina A. V., Cherepkov E. A., Belov U. S. Review of approaches to machine translation. In: *International Student Scientific Journal*. 2018. No. 6. Available at: <https://eduherald.ru/ru/article/view?id=19408>

3. How the Google Translate neural network works. Available at: <https://www.cossa.ru/152/196086/>

4. On machine translation. Developer's Guide. Available at: <https://yandex.ru/dev/translate/doc/dg/concepts/how-works-machine-translation.html>

5. How DeepL Translator works. Available at: <https://www.deepl.com/ru/blog/how-does-deepl-work>

6. Molchanov A. Statistical and hybrid translation methods in PROMT technologies. In: *Control Engineering Russia – August 2013*. Available at: <https://controlengrussia.com/innovatsii/statisticheskie-i-gibridnye-metody-perevoda-v-tehnologiyah-kompanii-promt>

7. Top 10 Snooker Shots of All Time: Jaw Dropping, Pure Magic, Surreal. Available at: [https://scrum.com/en-us/snooker/@jodcarey/top-10-snooker-shots-of-all-time-jaw-dropping-pure-magic-surreal](https://scorum.com/en-us/snooker/@jodcarey/top-10-snooker-shots-of-all-time-jaw-dropping-pure-magic-surreal)

8. Salients, re-entrants and pockets. Available at: [https://military-history.fandom.com/wiki/Salients,\\_re-entrants\\_and\\_pockets](https://military-history.fandom.com/wiki/Salients,_re-entrants_and_pockets)

9. Britannica, Gauge. Railroad track. Available at: <https://www.britannica.com/technology/gauge-railroad-track>

10. Understanding Pressure Gauge. Available at: <https://studentlesson.com/definition-uses-diagram-types-working-of-a-pressure-gauge/>

*Орловский государственный университет имени  
И. С. Тургенева*

*Новиков А. В., аспирант кафедры английской филологии*

*E-mail: artemnovikov57@yandex.ru*

*Поступила в редакцию 24 декабря 2024 г.*

*Принята к публикации 26 мая 2025 г.*

**Для цитирования:**

Новиков А. В. Контекстуальное распознавание семантики многозначных слов системами машинного перевода // Вестник Воронежского государственного университета. Серия: Лингвистика и межкультурная коммуникация. 2025. № 3. С. 6–15. DOI: <https://doi.org/10.17308/lic/1680-5755/2025/3/6-15>

*Orel State University named after I. S. Turgenev  
Novikov A. V., Post-graduate Student of the English  
Philology Department*

*E-mail: artemnovikov57@yandex.ru*

*Received: 24 December 2024*

*Accepted: 26 May 2025*

**For citation:**

Novikov A. V. Contextual recognition of polysemantic words semantics by machine translation systems. *Proceedings of Voronezh State University. Series: Linguistics and Intercultural Communication*. 2025. No. 3. Pp. 6–15. DOI: <https://doi.org/10.17308/lic/1680-5755/2025/3/6-15>