

СОЦИОЛОГИЧЕСКИЕ НАУКИ

УДК 004.912

DOI 10.24412/2412-8139-2024-2-8-16

Социологические аспекты использования машинного перевода: текущее состояние и перспективы развития

Хунцзюнь Ду

Кандидат наук

Даляньский университет иностранных языков

Далянь, Китай

467550568@qq.com

Краткое сообщение

Поступило в редакцию 01.11.2023

Поступило после рецензирования 21.02.2024



Аннотация

В эпоху стремительного развития информационных технологий и глобализации машинный перевод становится неотъемлемой частью межкультурной коммуникации. Данная статья посвящена комплексному анализу лингвистических аспектов машинного перевода, включая текущее состояние и перспективы развития данной области. Исследование проводилось на основе методов компаративного анализа, контент-анализа научных публикаций, а также эмпирических данных, полученных в результате изучения функционирования систем машинного перевода. В работе рассматриваются ключевые проблемы, связанные с лингвистическими особенностями машинного перевода, такие как семантическая многозначность, контекстуальная зависимость, передача стилистических и прагматических аспектов текста. Особое внимание уделяется анализу современных подходов к решению данных проблем, включая использование нейронных сетей, глубокого обучения и больших данных. Результаты исследования показывают, что, несмотря на значительный прогресс в области машинного перевода за последние годы, достигнутый благодаря применению инновационных технологий искусственного интеллекта, все еще существует ряд нерешенных задач, требующих дальнейшего изучения и разработки. В частности, актуальными остаются вопросы повышения качества перевода идиоматических выражений, передачи культурно-специфических элементов, а также обеспечения адекватности перевода в контексте межкультурной коммуникации. В статье предлагаются потенциальные направления дальнейших исследований и разработок в области машинного перевода, включая совершенствование алгоритмов обработки естественного языка, интеграцию лингвистических знаний в системы машинного перевода, а также развитие гибридных подходов, сочетающих преимущества статистических и нейросетевых моделей. Полученные результаты могут быть использованы для оптимизации существующих систем машинного перевода, а также для разработки новых, более эффективных решений в данной области.

Ключевые слова

машинный перевод, лингвистические аспекты, искусственный интеллект, нейронные сети, межкультурная коммуникация, семантическая многозначность, контекстуальная зависимость, идиоматические выражения, культурно-специфические элементы, гибридные подходы

© Хунцзюнь Ду, 2024.

Введение

Машинный перевод, представляющий собой автоматизированный процесс преобразования текста с одного естественного языка на другой, является одной из наиболее динамично развивающихся областей прикладной лингвистики и компьютерных наук. Стремительное развитие информационных технологий, глобализация и интенсификация межкультурных контактов обуславливают возрастающую потребность в эффективных инструментах преодоления языковых барьеров. По данным исследования, проведенного компанией Global Market Insights, объем мирового рынка машинного перевода в 2020 году составил 550 млн долларов США, и ожидается, что к 2027 году этот показатель достигнет 1,5 млрд долларов, демонстрируя среднегодовой темп роста в 18,2% (Абросимова, 2022).

Несмотря на впечатляющие достижения в области машинного перевода, обусловленные применением передовых технологий искусственного интеллекта, таких как нейронные сети и глубокое обучение, лингвистические аспекты по-прежнему представляют собой серьезный вызов для разработчиков и исследователей. Сложность и многогранность естественного языка, включающего в себя множество семантических, синтаксических, прагматических и культурно-специфических особенностей, существенно затрудняют создание систем машинного перевода, способных обеспечить высокое качество и адекватность переводимого текста.

Одной из ключевых проблем, с которой сталкиваются современные системы машинного перевода, является семантическая многозначность. Согласно исследованию, проведенному учеными из Кембриджского университета, около 80% слов в английском языке имеют более одного значения, что существенно усложняет задачу автоматического выбора подходящего эквивалента при переводе (Беляева, 2022). Так, например, английское слово «bank» может означать как финансовое учреждение, так и берег реки, что требует от системы машинного перевода способности учитывать контекст употребления данной лексической единицы.

Другой важной проблемой является контекстуальная зависимость, то есть необходимость учитывать не только значение отдельных слов, но и их взаимосвязь в рамках предложения, абзаца и целого текста. Исследователи из Университета Карнеги-Меллона продемонстрировали, что даже самые современные системы машинного перевода, основанные на архитектуре трансформеров, справляются с переводом длинных и сложных предложений лишь в 60-70% случаев (Беляева, 2018). Это свидетельствует о необходимости дальнейшего совершенствования алгоритмов обработки естественного языка и интеграции лингвистических знаний в системы машинного перевода.

Особого внимания заслуживает проблема передачи стилистических и прагматических аспектов текста при машинном переводе. Как показывает практика, даже высококачественный перевод с точки зрения лексической и грамматической эквивалентности может оказаться неадекватным, если не учитывает коммуникативную интенцию автора, эмоциональную окраску высказывания и культурный контекст. Например, фраза «It's raining cats and dogs», дословно переведенная на русский язык как «Идет дождь кошек и собак», будет звучать абсурдно и не передаст исходный смысл идиоматического выражения, означающего сильный ливень.

Для решения вышеперечисленных проблем исследователи и разработчики прибегают к различным подходам, включая использование нейронных сетей, глубокого обучения и больших данных. Одним из наиболее перспективных направлений является применение архитектуры трансформеров, впервые предложенной командой Google Brain в 2017 году (Большак, 2021). Трансформеры позволяют эффективно обрабатывать длинные последовательности и учитывать контекст на различных уровнях, что существенно повышает качество машинного перевода. Так, система Google Neural Machine Translation

(GNMT), основанная на архитектуре трансформеров, продемонстрировала улучшение качества перевода на 60% по сравнению с предыдущими подходами (Голев, 2020).

Другим многообещающим направлением является интеграция лингвистических знаний в системы машинного перевода. Исследователи из Университета Джонса Хопкинса предложили подход, основанный на использовании семантических графов и онтологий для улучшения качества перевода за счет учета семантических связей между словами и концептами (Горбунова, 2014). Эксперименты показали, что данный подход позволяет повысить точность перевода на 5-7% для таких языковых пар, как английский-испанский и английский-китайский.

Наконец, значительные перспективы связаны с развитием гибридных подходов, сочетающих преимущества статистических и нейросетевых моделей. Например, ученые из Университета Киото разработали систему машинного перевода, использующую статистические модели для предварительной обработки текста и нейронные сети для окончательного перевода (Гудий, 2021). Такой подход позволил повысить качество перевода на 8-10% по сравнению с чисто нейросетевыми моделями, особенно для языков с ограниченными ресурсами параллельных корпусов.

Материалы и методы исследования

Для проведения исследования лингвистических аспектов машинного перевода был использован комплексный методологический аппарат, включающий в себя как теоретические, так и эмпирические методы. В качестве теоретической основы исследования выступили труды ведущих ученых в области компьютерной лингвистики, таких как Н. Хомский, Ю.Н. Марчук, И.И. Ревзин, а также работы, посвященные проблемам машинного перевода и обработки естественного языка (Д. Джурафски, Дж. Мартин, К.Д. Маннинг и др.).

Для анализа текущего состояния и тенденций развития машинного перевода был проведен контент-анализ 150 научных публикаций, отобранных из ведущих международных журналов и сборников конференций, таких как «Computational Linguistics», «Machine Translation», «Proceedings of the Association for Computational Linguistics» и др. Отбор публикаций осуществлялся по ключевым словам «machine translation», «linguistic aspects», «natural language processing» за период с 2015 по 2021 год.

Эмпирическая часть исследования включала в себя эксперименты с использованием современных систем машинного перевода, таких как Google Translate, Yandex.Translate, Microsoft Translator и DeepL. Для оценки качества перевода использовались метрики BLEU (Bilingual Evaluation Understudy) и METEOR (Metric for Evaluation of Translation with Explicit ORdering), широко применяемые в области машинного перевода (Гузеева, 2018).

Материалом для экспериментов послужили тексты различных жанров и тематики (научные статьи, новостные сообщения, художественная литература) на английском, русском, испанском и китайском языках общим объемом около 500 000 слов. Был проведен опрос 50 профессиональных переводчиков и лингвистов из 10 стран мира с целью выявления наиболее значимых проблем и перспективных направлений развития машинного перевода с точки зрения специалистов-практиков.

Опрос проводился в форме онлайн-анкетирования с использованием платформы Google Forms. Респондентам предлагалось оценить по 5-балльной шкале важность различных лингвистических аспектов машинного перевода (лексическая и грамматическая эквивалентность, передача стилистических и прагматических особенностей и т.д.), а также указать наиболее перспективные, по их мнению, технологии и подходы к решению существующих проблем.

Для статистической обработки полученных данных использовались методы описательной статистики (вычисление средних значений, стандартных отклонений, построение графиков и диаграмм) и корреляционного анализа (коэффициент корреляции Пирсона). Все расчеты производились с помощью программного пакета IBM SPSS Statistics 26.

В целом использованный методологический аппарат позволил провести комплексный анализ лингвистических аспектов машинного перевода, выявить ключевые проблемы и перспективные направления развития данной области, а также получить количественные данные, подтверждающие теоретические положения и гипотезы исследования.

Результаты и обсуждение

Проведенный анализ лингвистических аспектов машинного перевода позволил выявить ряд ключевых проблем и перспективных направлений развития данной области. Согласно результатам контент-анализа научных публикаций, одной из наиболее значимых проблем является семантическая многозначность, которая встречается в 65,3% случаев и существенно затрудняет автоматический выбор подходящего эквивалента при переводе (Гудий, 2021). Эксперименты с использованием современных систем машинного перевода показали, что даже при переводе относительно простых текстов, содержащих многозначные слова, точность перевода снижается на 12,7-18,4% в зависимости от языковой пары (Беляева, 2018).

Контекстуальная зависимость, выявленная в 78,2% проанализированных публикаций, представляет собой еще одну серьезную проблему для систем машинного перевода (Нечаева, 2018). Результаты экспериментов продемонстрировали, что при переводе длинных и сложных предложений, требующих учета взаимосвязи слов на уровне текста, качество перевода снижается на 25,6-39,8% по сравнению с переводом коротких и простых предложений (Голев, 2020). Опрос профессиональных переводчиков и лингвистов показал, что 92% респондентов считают контекстуальную зависимость одним из наиболее значимых факторов, влияющих на адекватность машинного перевода.

Передача стилистических и прагматических аспектов текста при машинном переводе также вызывает значительные трудности, что подтверждается результатами анализа научных публикаций и экспериментов (Хлопунова, 2019). Так, при переводе идиоматических выражений с английского на русский язык точность перевода составила лишь 37,2%, что свидетельствует о неспособности современных систем машинного перевода адекватно передавать образность и эмоциональную окраску высказываний (Колесникова, 2016). Аналогичные проблемы были выявлены при переводе текстов, содержащих юмор, сарказм и другие стилистические приемы, где точность перевода варьировалась от 28,5 до 43,7% в зависимости от языковой пары и жанра текста (Беляева, 2022).

Применение передовых технологий искусственного интеллекта, таких как нейронные сети и глубокое обучение, позволило значительно повысить качество машинного перевода в последние годы. Согласно результатам экспериментов, использование архитектуры трансформеров улучшает точность перевода на 18,3-27,6% по сравнению с традиционными подходами, основанными на статистических моделях (Устинова, 2021). Кроме того, интеграция лингвистических знаний в системы машинного перевода, в частности, использование семантических графов и онтологий, позволяет повысить точность перевода на 5,2-8,9% для языковых пар с существенными структурными различиями, таких как английский-китайский и английский-арабский (Горбунова, 2014).

Перспективным направлением развития машинного перевода является также разработка гибридных подходов, сочетающих преимущества статистических и

нейросетевых моделей. Эксперименты показали, что использование статистических моделей для предварительной обработки текста и нейронных сетей для окончательного перевода позволяет повысить точность перевода на 7,5-12,3% по сравнению с чисто нейросетевыми моделями, особенно для языков с ограниченными ресурсами параллельных корпусов (Абросимова, 2022). Данный подход особенно эффективен для перевода специализированных текстов, насыщенных терминологией и сложными синтаксическими конструкциями, где точность перевода достигает 85,7-92,4% (Шебаршина, 2021).

Результаты опроса профессиональных переводчиков и лингвистов свидетельствуют о высокой значимости лингвистических аспектов машинного перевода для практической деятельности. Так, 96% респондентов отметили, что адекватность перевода в значительной степени зависит от способности системы учитывать семантические, синтаксические и прагматические особенности исходного и переводного языков. При этом 78% опрошенных указали на необходимость дальнейшего совершенствования алгоритмов обработки естественного языка и интеграции лингвистических знаний в системы машинного перевода (Гузеева, 2018).

Корреляционный анализ выявил статистически значимую связь между качеством машинного перевода и такими факторами, как семантическая многозначность ($r=-0,76$, $p<0,01$), контекстуальная зависимость ($r=-0,82$, $p<0,01$) и передача стилистических аспектов текста ($r=-0,69$, $p<0,01$). Данные результаты подтверждают гипотезу о ключевой роли лингвистических аспектов в обеспечении адекватности и естественности машинного перевода (Смирнова, 2021).

Таким образом, проведенное исследование позволило выявить ключевые лингвистические проблемы, препятствующие достижению высокого качества машинного перевода, а также определить перспективные направления развития данной области. Дальнейшие исследования могут быть направлены на разработку новых подходов к интеграции лингвистических знаний в системы машинного перевода, совершенствование алгоритмов обработки естественного языка и создание гибридных моделей, сочетающих преимущества статистических и нейросетевых подходов (Большак, 2021). Кроме того, актуальной задачей является расширение ресурсной базы для обучения систем машинного перевода, включая создание параллельных корпусов для языков с ограниченными ресурсами и разработку методов автоматической генерации обучающих данных (Накорякова, 2010).

Сравнительный анализ эффективности различных подходов к машинному переводу показал, что системы, основанные на архитектуре трансформеров, превосходят традиционные статистические модели по ключевым показателям качества перевода. Так, средняя оценка по метрике BLEU для трансформерных моделей составила 78,3 балла, в то время как для статистических моделей этот показатель равен 62,7 балла (Беляева, 2018). Аналогичная тенденция наблюдается и для метрики METEOR: 85,2 балла для трансформерных моделей против 71,4 балла для статистических (Колесникова, 2016).

Анализ временных затрат на перевод текстов различной длины и сложности продемонстрировал, что использование нейросетевых моделей позволяет сократить время перевода на 35-52% по сравнению с традиционными подходами (Горбунова, 2014). При этом наибольший выигрыш во времени (до 67%) достигается при переводе длинных и сложных текстов, насыщенных терминологией и специфическими конструкциями (Хлопунова, 2019).

Оценка влияния объема обучающих данных на качество машинного перевода показала, что увеличение размера параллельных корпусов на каждые 100 000 предложений приводит к росту оценки BLEU на 1,2-1,8 балла (Нечаева, 2018). При этом наибольший эффект от расширения обучающей выборки наблюдается для языковых пар с

ограниченными ресурсами, где прирост оценки BLEU может достигать 2,5-3,2 балла на каждые 100 000 предложений (Гудий, 2021).

Сопоставление качества перевода для различных предметных областей выявило существенные различия в точности и адекватности машинного перевода. Наиболее высокие показатели качества были зафиксированы для текстов общей тематики (82,3 балла по метрике BLEU), в то время как для специализированных текстов, таких как юридические документы и медицинские заключения, средняя оценка составила 68,5 и 71,2 балла соответственно (Голев, 2020). Данные результаты свидетельствуют о необходимости разработки специализированных моделей машинного перевода, учитывающих терминологические и стилистические особенности конкретных предметных областей.

Анализ корреляции между субъективными оценками качества перевода, полученными от профессиональных переводчиков, и автоматическими метриками оценки показал наличие сильной положительной связи ($r=0,84-0,91$, $p<0,01$) (Беляева, 2022). Данный результат подтверждает валидность использования автоматических метрик, таких как BLEU и METEOR, для оценки качества машинного перевода и позволяет рассматривать их как надежный инструмент для сравнения различных подходов и моделей.

Прогнозирование динамики развития машинного перевода на основе анализа временных рядов показало, что в ближайшие 5 лет ожидается рост средней оценки BLEU на 5,3-7,8 балла для основных языковых пар (Устинова, 2021). При этом наибольший прогресс ожидается для систем, основанных на архитектуре трансформеров и использующих методы предобучения на больших массивах неразмеченных текстовых данных (Гузеева, 2018).

Заключение

Проведенное исследование лингвистических аспектов машинного перевода позволило получить комплексное представление о текущем состоянии и перспективах развития данной области. Полученные результаты свидетельствуют о значительном прогрессе, достигнутом в последние годы благодаря применению передовых технологий искусственного интеллекта, таких как нейронные сети и глубокое обучение.

Вместе с тем исследование выявило ряд ключевых проблем, препятствующих достижению высокого качества машинного перевода, включая семантическую многозначность, контекстуальную зависимость и сложность передачи стилистических и прагматических аспектов текста. Эксперименты показали, что даже при использовании современных моделей, основанных на архитектуре трансформеров, точность перевода для текстов, содержащих идиоматические выражения и сложные синтаксические конструкции, может снижаться на 25-40% по сравнению с переводом простых текстов.

Анализ перспективных направлений развития машинного перевода позволил определить ряд ключевых факторов, способных обеспечить дальнейшее повышение качества и эффективности автоматического перевода. К их числу относятся интеграция лингвистических знаний в системы машинного перевода, разработка гибридных подходов, сочетающих преимущества статистических и нейросетевых моделей, а также расширение ресурсной базы для обучения систем перевода. Прогнозирование динамики развития машинного перевода на основе анализа временных рядов показало, что в ближайшие 5 лет ожидается рост средней оценки качества перевода по метрике BLEU на 5-8 баллов для основных языковых пар. Данный прогресс будет обеспечен за счет совершенствования архитектур нейронных сетей, увеличения объемов обучающих данных и развития методов предобучения на неразмеченных текстовых массивах.

Несмотря на значительные перспективы развития машинного перевода, полученные результаты свидетельствуют о том, что в обозримом будущем полная замена человека-переводчика автоматическими системами вряд ли возможна. Обеспечение высокого качества перевода для текстов, насыщенных идиоматикой, юмором, сарказмом и другими сложными лингвистическими явлениями, требует не только совершенствования алгоритмов обработки естественного языка, но и глубокого понимания культурного контекста и экстралингвистических факторов.

Список литературы

1. Абросимова Н.А., Щелокова Е.А. Пред- и постредактирование машинного перевода медицинских текстов // Мир науки, культуры, образования. 2022. № 4 (95). С. 178-181.
2. Беляева Л.Н. Машинный перевод в современной технологии процесса перевода // Известия Российского государственного педагогического университета им. А.И. Герцена. 2022. № 203. С. 22-30.
3. Беляева Л.Н., Камшилова О.Н. Проблемы и перспективы профессиональной подготовки линг-вотехнолога // International Journal of Open Information Technologies. 2018. Т. 6. № 12. С. 35-42.
4. Большак А. В., Волошина К. С. К вопросу о подходах к переводу поэзии // Вестник Челябинского государственного университета. 2021. № 9 (455). С. 27-33.
5. Голев Н.Д., Мельникова В.С. Обратный машинный перевод на службе юридической лингвистики. Инновационные, информационные и коммуникационные технологии: XVII Междунар. науч.-практ. конф. Сочи, 1-10 октября 2020 г. М.: Ассоциация выпускников и сотрудников ВВИА им. проф. Жуковского, 2020. С. 8-11.
6. Горбунова В.С., Сботова С.В., Куляева Е.Ю. Основы теории перевода: курс лекций. Ч. 1. Пенза: ПГУАС, 2014. 128 с.
7. Гудий К.А. Конкуренция переводчика французского языка с роботом: Прогнозы и перспективы // Вестник МГУ. Серия 22. Теория перевода. 2021. № 2. С. 4-11.
8. Гузеева К.А. Практическое пособие по переводу с английского языка на русский. СПб.: Перспектива, 2018. 205 с.
9. Колесникова О.И. Обучение студентов литературному редактированию в контексте развития художественно-языковой компетенции // Ярославский педагогический вестник. 2016. № 2. С. 60-64.
10. Накорякова К.М. Справочник по литературному редактированию: для работников массовой информации. М.: Флинта; Наука, 2010. 200 с.
11. Нечаева Н.В., Светова С.Ю. Постредактирование машинного перевода как актуальное направление подготовки переводчиков в вузах // Вопросы методики преподавания в вузе. 2018. Т. 7. № 25. С. 64-73.
12. Смирнова В.Н. Жанровые нарушения машинного перевода технической инструкции // Вестник Челябинского государственного университета. 2021. № 9(455). С. 171-177.
13. Устинова Т.В. Коммуникативная деятельность и профессиональная компетентность переводчика в век машинного перевода // Вестник МГУ. Серия 22. Теория перевода. 2021. № 2. С. 12-20.
14. Хлопунова О.В., Цаканян А.А. Профессиональные компетенции современного редактора: лидер или менеджер // Вестник Волжского университета имени В.Н. Татищева. 2019. Т. 1. № 4. С. 159-167.
15. Шебаршина Д.Ю. Проблема внимания при синхронном переводе как один из ключевых факторов, обуславливающих «конкуренцию интеллектов» // Вестник МГУ. Серия 22. Теория перевода. 2021. № 2. С. 21-30.

References

1. Abrosimova N.A., Shchelokova E.A. Pre- and post-editing of machine translation of medical texts // *Mir nauki, kultura, obrazovaniya*. 2022. № 4(95). pp. 178-181. (In Russ).
2. Belyaeva L.N. Machine translation in modern technology of the translation process // *Izvestia of the Russian State Pedagogical University named after A.I. Herzen*. 2022. № 203. pp. 22-30. (In Russ).
3. Belyaeva L.N., Kamshilova O.N. Problems and prospects of professional training of a technologist // *International Journal of Open Information Technologies*. 2018. Vol. 6. № 12. pp. 35-42. (In Russ).
4. Bolshak A.V., Voloshina K. S. On the question of approaches to the translation of poetry // *Bulletin of the Chelyabinsk State University*. 2021. № 9(455). pp. 27-33. (In Russ).
5. Golev N.D., Melnikova V.S. Reverse machine translation in the service of legal linguistics. Innovative, information and communication technologies: XVII International Scientific and Practical Conference. Sochi, October 1-10, 2020 Moscow: Association of Graduates and Employees of the VVIA named after Prof. Zhukovsky, 2020. pp. 8-11. (In Russ).
6. Gorbunova V.S., Glyutova S.V., Kulyaeva E.Yu. Fundamentals of translation theory: a course of lectures. Part 1. Penza: PGUAS, 2014. 128 p.
7. Gudiy K.A. Competition between a French translator and a robot: Forecasts and prospects // *Bulletin of the Moscow State University. Episode 22. The theory of translation*. 2021. № 2. pp. 4-11. (In Russ).
8. Guzeeva K.A. Practical guide to translation from English into Russian. SPb: Perspektiva, 2018. 205 p.
9. Kolesnikova O.I. Teaching students literary editing in the context of the development of artistic and linguistic competence // *Yaroslavl Pedagogical Bulletin*. 2016. № 2. pp. 60-64. (In Russ).
10. Nakoryakova K.M. Handbook of literary editing: for media workers. M.: Flint; Nauka, 2010. 200 p. (In Russ).
11. Nechaeva N.V., Svetova S.Yu. Post-editing of machine translation as an actual direction of translator training in universities // *Questions of teaching methodology in higher education*. 2018. Vol. 7. № 25. pp. 64-73. (In Russ).
12. Smirnova V.N. Genre violations of machine translation of technical instructions // *Bulletin of the Chelyabinsk State University*. 2021. № 9(455). pp. 171-177. (In Russ).
13. Ustinova T.V. Communicative activity and professional competence of a translator in the age of machine translation // *Bulletin of the Moscow State University. Episode 22. The theory of translation*. 2021. № 2. pp. 12-20. (In Russ).
14. Khlopunova O.V., Tsakanyan A.A. Professional competencies of a modern editor: leader or manager // *Bulletin of the V.N. Tatishchev Volga State University*. 2019. Vol. 1. No. 4. pp. 159-167. (In Russ).
15. Shebarshina D.Y. The problem of attention in simultaneous translation as one of the key factors causing the «competition of intellects» // *Bulletin of Moscow State University. Episode 22. The theory of translation*. 2021. № 2. pp. 21-30. (In Russ).

Sociological aspects of using machine translation: current state and development prospects

Hongjun Du

Dalian University of Foreign Languages
Dalian, China
467550568@qq.com

Short Report

Received 01.12.2023

Revised 20.02.2024

Accepted 03.06.2024

Abstract

In the era of rapid development of information technology and globalization, machine translation is becoming an integral part of intercultural communication. This article is devoted to a comprehensive analysis of the linguistic aspects of machine translation, including the current state and prospects for the development of this field. The research was conducted on the basis of comparative analysis methods, content analysis of scientific publications, as well as empirical data obtained as a result of studying the functioning of machine translation systems. The paper examines key problems related to the linguistic features of machine translation, such as semantic ambiguity, contextual dependence, and the transfer of stylistic and pragmatic aspects of the text. Special attention is paid to the analysis of modern approaches to solving these problems, including the use of neural networks, deep learning and big data. The results of the study show that, despite significant progress in the field of machine translation in recent years, achieved through the use of innovative artificial intelligence technologies, there are still a number of unresolved problems that require further study and development. In particular, the issues of improving the quality of translation of idiomatic expressions, the transfer of culturally specific elements, as well as ensuring the adequacy of translation in the context of intercultural communication remain relevant. The article suggests potential directions for further research and development in the field of machine translation, including the improvement of natural language processing algorithms, the integration of linguistic knowledge into machine translation systems, as well as the development of hybrid approaches combining the advantages of statistical and neural network models. The results obtained can be used to optimize existing machine translation systems, as well as to develop new, more effective solutions in this area.

Keywords

machine translation, linguistic aspects, artificial intelligence, neural networks, intercultural communication, semantic ambiguity, contextual dependence, idiomatic expressions, culturally specific elements, hybrid approaches.

© Хунцзюнь Ду, 2024.