

УДК 004:912

ИЗУЧЕНИЕ РОЛИ КЛАССИФИКАЦИИ ЯЗЫКОВ В СОЗДАНИИ ПРОГРАММ-ПЕРЕВОДЧИКОВ

Махмуджанова Сайёра Яшин кизи

Преподаватель Университета мировой экономики и дипломатии

s.maxmudjanova@uwed.uz

Аннотация: *В этой статье анализируется роль и значение классификации языков при создании программ-переводчиков. Результаты исследования показали, что классификация, основанная на морфологических, синтаксических и семантических свойствах языков, повышает эффективность систем машинного перевода, повышает точность перевода и семантическую совместимость. Особые подходы важны, особенно для агглютинативных и синтетических языков. В исследовании анализируется место классификации языков в статистических, основанных на правилах и нейронных моделях перевода и показаны перспективы создания адаптивных и динамических систем перевода в будущем.*

Ключевые слова: *классификация языков, машинный перевод, агглютинативные языки, нейронный перевод, лингвистическая классификация, языки с ограниченными ресурсами.*

ВВЕДЕНИЕ

Ускорение процессов глобализации в современном информационном обществе и широкое внедрение цифровых технологий резко повысили потребность в быстром и качественном обмене информацией между различными языками. В частности, развитие алгоритмов, основанных на искусственном интеллекте, больших данных (Big Data) и глубоком обучении, вывело область машинного перевода (Machine Translation – MT) на новый уровень. Сегодня системы автоматического переводчика, такие как Google Translate, DeepL, Яндекс переводчик, активно используются в научном, техническом, образовательном и повседневном общении.

Тем не менее, качество машинного перевода не одинаково высокое для всех языков. Это состояние напрямую связано, прежде всего, с лингвистической структурой и типологическими особенностями языков. Языки существенно различаются по морфологической структуре (аналитическая, синтетическая, агглютинативная), синтаксической модели (порядок слов, построение предложения) и средствам семантического выражения. Эти различия создают важные проблемы для обеспечения грамматической согласованности в процессе перевода, предотвращения потери смысла и правильной интерпретации контекста.

При создании программ-переводчиков одним из важнейших условий решения этих задач является правильная классификация (классификация) языков на основе лингвистических критериев. Классификация языков позволяет адаптировать модели перевода, оптимизировать алгоритмы и разрабатывать модели с учетом грамматических и морфологических особенностей языка. В научных исследованиях отмечалось, что универсальные модели перевода, особенно для языков с агглютинативной структурой, включая узбекский, не дают достаточного результата.

В то время как подходы к статистическому, основанному на правилах и нейронному машинному переводу широко используются в последние годы, языковая классификация рассматривается как вторичный фактор в большинстве этих систем. В результате качество перевода для некоторых языков остается низким, особенно языки с ограниченными ресурсами страдают от этого больше. Поэтому глубокая интеграция классификации языков в системы машинного перевода представляется актуальной научно-практической проблемой.

В данной статье всесторонне анализируется роль классификации языков в создании программ-переводчиков, исследуются существующие классификационные подходы и их влияние на эффективность машинного перевода. Основная цель исследования-определить важность классификации языков для проектирования и оптимизации систем машинного перевода, а также обосновать эффективные подходы, обеспечивающие высокое качество перевода.

Методология

Настоящее исследование направлено на установление теоретических основ и практического значения классификации языков при создании систем машинного перевода, в процессе исследования был применен комплексный научный подход. Методология исследования основана на взаимной интеграции областей лингвистического анализа, компьютерной лингвистики и искусственного интеллекта.

На первом этапе исследования систематически изучались фундаментальные и современные научные ресурсы, посвященные вопросам машинного перевода, лингвистической типологии и классификации языков. В частности, были проанализированы научные статьи, материалы конференций и технические отчеты о статистических моделях и моделях нейронного машинного перевода, многоязычных системах перевода и алгоритмах, разработанных для языков с ограниченными ресурсами. Этот этап послужил обобщением существующих научных взглядов на классификацию языков в системах перевода.

На втором этапе был проведен углубленный анализ подходов к классификации языков, используемых в программах переводчика. В ходе

исследования с целью выявления языковых особенностей языков и их влияния на машинный перевод были использованы следующие научные методы.

Аналитический метод применялся как основной инструмент при изучении морфологической и синтаксической структуры языков. С помощью этого метода было установлено, что в аналитических языках грамматические отношения выражаются в основном через Порядок слов и вспомогательные слова, в то время как в агглютинативных языках то, как грамматическое значение выражается непрерывно через аффиксы, влияет на процесс перевода. Также научно обосновано, что сложные грамматические структуры синтетических языков создают дополнительные проблемы при машинном переводе.

Метод сравнения применялся для сравнения процесса перевода между языками, принадлежащими к разным языковым группам. В частности, на примере английского (аналитического), русского (синтетического) и узбекского (агглютинативного) языков сравнивались результаты перевода и анализировалось, как различия в структуре языков влияют на точность перевода, грамматическую согласованность и семантическую согласованность. Этот метод позволил выявить преимущества подходов, основанных на классификации языков.¹⁹

Системный подход был использован для комплексной оценки места классификации языков в статистических, основанных на правилах и нейронных моделях машинного перевода. С помощью этого подхода оценивалась эффективность моделирования языковых морфологических и синтаксических характеристик в системах перевода, использования параллельных корпусов, а также эффективность механизмов трансферного обучения в многоязычных нейронных моделях. В результате выяснилось, что модели, разработанные с учетом классификации языков, обладают высокой эффективностью.²⁰

В рамках исследования языки классифицировались на основе лингвистических критериев: по морфологическому строению (аналитический, синтетический, агглютинативный), по порядку слов (СВО, сов и другие структурные модели), а также по способу выражения грамматических категорий (падеж, время, лицо-число и другие грамматические показатели). Данная классификация послужила методологической основой при проектировании и оптимизации моделей машинного перевода.

В целом данный методологический подход позволил комплексно изучить теоретическую и практическую значимость классификации языков в программах переводчика, а также выявить основные факторы, влияющие на качество перевода и эффективность модели.

Результаты

¹⁹ Croft, W., & Cruse, D. A. (2004). *Cognitive Linguistics*. Cambridge University Press.

²⁰ Johnson, M., et al. (2017). *Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation*. Transactions of the Association for Computational Linguistics, 5, 339–351.

Результаты исследования показали, что программы-переводчики, созданные с учетом классификации языков, значительно улучшают качество перевода.

В частности, применение отдельных модулей морфологического анализа для агглютинативных языков (например, узбекского) обеспечивает несколько преимуществ в процессе перевода.

Например, в узбекском языке к одному слову может быть присоединено несколько аффиксов (вид: из наших книг — “из наших книг”), и необходимо, чтобы значение и грамматические категории этого слова были определены одновременно.

Модули морфологического анализа обеспечивают точный контекст для системы перевода, автоматически определяя эту структуру и разделяя сегменты слова kontekst, что приводит к грамматически и семантически Правильному переводу.

В моделях статистического машинного перевода (SMT) языковая классификация оптимизирует процесс выбора и сопоставления параллельных корпусов.

Например, когда мы сравниваем английский и русский языки при обучении модели перевода, ошибки перевода значительно уменьшаются при выборе корпусов, соответствующих синтаксическим и морфологическим характеристикам.

Это служит для повышения уровня совместимости слов и предложений в системе SMT, а также для повышения точности перевода на языках с ограниченными ресурсами.

Учет структурного сходства языков в системах нейронного машинного перевода (НМТ) дает высокую эффективность при построении многоязычных моделей.

Например, при использовании технологии трансферного обучения между языками, принадлежащими к одной морфологической группе, модель адаптирует параметры, изучаемые на одном языке, к другому языку.

В результате получается качественный перевод даже на малоресурсные языки. Например, при адаптации параметров нейронной модели, обучаемой англо-русскому переводу, к англо-узбекскому переводу уменьшаются семантические и грамматические ошибки, особенно при правильном использовании порядка слов и согласных.

Кроме того, подходы, основанные на классификации языков, уменьшают количество грамматических ошибок и улучшают семантическую согласованность перевода.

Например, грамматические категории, такие как спряжение, время, лицо и число, автоматически моделируются и передаются в систему перевода.

При этом правильно учитывается порядок слов и контекстные зависимости.

В результате перевод становится не только грамматически правильным, но и понятным и понятным читателю по смыслу.



В целом результаты показывают, что:

1. Классификация языков повышает качество и точность системы машинного перевода.

2. Модули, адаптированные к морфологическим и синтаксическим характеристикам, также эффективно работают на сложных языках.

3. Многоязычные нейронные модели и трансферное обучение обеспечивают высококачественный перевод даже на языки с ограниченными ресурсами.

4. Классификация языков сводит к минимуму грамматические и семантические ошибки, делая перевод понятным и лаконичным для читателя.

Таким образом, классификация языков оказалась важным стратегическим фактором в системах машинного перевода не только технически, но и лингвистически.

Обсуждение

Полученные результаты ясно показывают, что классификация языков имеет стратегическое значение для систем машинного перевода. Исследование подтвердило, что традиционные программы-переводчики часто полагаются на универсальную модель, то есть одна модель работает по общим правилам для всех языков.

Такой подход приводит к снижению качества перевода на некоторых языках, особенно на морфологически сложных или малоресурсных. Например, в агглютинативных языках грамматические значения непрерывно выражаются через аффиксы внутри слова; универсальная модель, с другой стороны, не учитывает полностью такую структуру, что приводит к потере значения словаря или грамматическим ошибкам.

В этом контексте лингвистически правильная модель группировки и классификации языков позволяет адаптироваться. Например, когда

агглютинативные, синтетические и аналитические языки рассматриваются как отдельная группа, морфологические и синтаксические характеристики каждой группы учитываются в моделях перевода. Это позволяет модели изучать грамматические и семантические правила, специфичные для языка, и повышает точность перевода.

Для языков с ограниченными ресурсами важность подходов, основанных на классификации языков, еще больше возрастает. В агглютинативных языках, таких как узбекский, необходимы специальные алгоритмы, модули морфологического анализа и разработка структур данных. Такой подход не только повышает качество перевода, но и позволяет эффективно осуществлять машинный перевод на малоресурсные языки. Однако отсутствие параллельных корпусов и учет морфологической сложности требуют дальнейшей оптимизации модели.

Одним из перспективных направлений будущих исследований является глубокая интеграция языковой классификации в нейронные сети, в частности внедрение механизмов адаптивной и динамической классификации. При этом модель может осуществлять передачу знаний между различными языками, обеспечивая высококачественные результаты перевода даже на языках с ограниченными ресурсами. Также такой подход повышает эффективность многоязычных систем машинного перевода и позволяет создавать инновационные алгоритмы, основанные на типологии языков.

В целом, исследование показывает, что классификация языков важна не только теоретически, но и имеет практическое значение при проектировании и оптимизации систем машинного перевода. Типологическую группировку и адаптацию языков следует рассматривать как стратегический инструмент повышения эффективности современных систем перевода.

Заключение

Это исследование ясно показало, что классификация языков является ключевым стратегическим фактором эффективности и точности систем машинного перевода. Классификация, основанная на морфологических, синтаксических и семантических свойствах языков, особенно агглютинативных и синтетических языков, значительно повышает качество перевода, обеспечивает грамматическую и семантическую согласованность и уменьшает ошибки перевода.

Результаты исследования показали, что классификация языков в статистических, основанных на правилах и нейронных моделях машинного перевода служит альтернативным и необходимым инструментом для создания высококачественных моделей перевода не только для эффективного использования ресурсов, но и для языков с ограниченными ресурсами. В то же время использование технологии трансферного обучения на основе типологической группировки языков и морфологических особенностей в

многоязычных нейронных системах значительно повышает точность перевода и гибкость системы.

В будущем глубокая интеграция классификации языков в системы машинного перевода, в частности внедрение механизмов адаптивной, динамической и автоматической классификации в нейронные сети, выведет качество перевода на новый уровень. Эти подходы являются важным с научной и практической точки зрения перспективным направлением в создании высокоэффективных систем перевода, особенно для малоресурсных и морфологически сложных языков.

Таким образом, классификацию языков следует признать ключевым элементом, обеспечивающим не только теоретическую основу, но и практическое и стратегическое преимущество в современных системах машинного перевода.

ЛИТЕРАТУРА:

1. Koehn, P. (2020). *Neural Machine Translation*. Cambridge University Press.
2. Popescu-Belis, A. (2019). *Multilingual Natural Language Processing: Challenges and Approaches*. Springer.
3. Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. *International Conference on Learning Representations (ICLR)*.
4. Comrie, B. (1989). *Language Universals and Linguistic Typology: Syntax and Morphology*. University of Chicago Press.
5. Croft, W., & Cruse, D. A. (2004). *Cognitive Linguistics*. Cambridge University Press.
6. Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., Dean, J. (2017). Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5, 339–351.
7. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems (NeurIPS)*, 30.
8. Och, F. J., & Ney, H. (2003). A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1), 19–51