



УДК 004.9:519.7655

© 2022 г. А.А. Животова,

В.Д. Бердоносков, канд. техн. наук

(Комсомольский-на-Амуре государственный университет)

ПЕРСПЕКТИВНЫЕ НАПРАВЛЕНИЯ РАЗВИТИЯ СИСТЕМ МАШИННОГО ПЕРЕВОДА

Систематизированы знания о видах систем машинного перевода в контексте развития и повышения идеальности. Проведен глубокий анализ проблем существующих систем с использованием ТРИЗ-эволюционного подхода. Определены перспективные направления развития машинного перевода.

Ключевые слова: машинный перевод, гибридный перевод, интерактивный перевод, систематизация знаний, ТРИЗ-эволюция.

DOI: 10.22250/18142400_2022_72_2_116

Введение

В условиях, когда количество данных и скорость обмена знаниями постоянно увеличиваются, время генерирования информации сократилось с месяцев и недель до дней и часов, а ее объем экспоненциально растет, процесс получения и систематизации релевантных знаний стремительно усложняется. Требуются современные инструменты поиска, отбора и обработки релевантных данных, а также систематизации полученных знаний. Учитывая развитие технологий и глобализации, применение таких инструментов невозможно представить без использования перевода. Перевод – рутинная необходимость во многих отраслях, включая науку, производство, медицину и т.д., а с неуклонным ростом количества информации и скорости ее генерирования и воспроизведения потребность в повышении качества перевода, наряду с сокращением затрат на него, неуклонно увеличивается.

Естественно, что одним из способов решения данной задачи является автоматизация процессов перевода. Первые попытки такой автоматизации берут свое начало еще в 1950-х гг. С тех пор технологии машинного перево-

да (МП) прошли большой путь, но только в 2014-2017 гг. [1] произошел качественный скачок роста идеальности данных систем, который привел рынок лингвистических услуг к пониманию того, что внедрение и развитие данных технологий является одним из наиболее перспективных направлений развития отрасли [2]. Современные системы МП демонстрируют высокие показатели качества, кардинально изменив за 2019-2021 гг. структуру рынка лингвистических услуг России [3], вытесняя переводчиков в пользу пост-редакторов и корректоров машинного перевода.

С развитием технологий искусственного интеллекта интерес исследователей к теме машинного перевода значительно возрос. Так, согласно данным базы «Science Direct» [4], количество статей по ключевым словам «машинный перевод» (Machine Translation) и «качество машинного перевода» (Machine Translation Quality) в 2021 г. выросло по сравнению с 2017 г. на 96% и 115% соответственно.

Несмотря на широкую освещенность темы в литературе, многие из существующих статей, опубликованных в рецензируемых журналах, устарели. Технологии за последние годы совершили качественный скачок, но рецензируемых материалов, описывающих новые технологии в разрезе эволюции, очень мало, а описание перспектив и направлений развития ограничено и/или отсутствует их обоснование. Между тем систематизация информации о технологиях МП необходима для выявления ключевых направлений и перспектив развития и исследования.

В статье систематизированы знания о существующих системах машинного перевода и проведен ТРИЗ-эволюционный анализ с целью выявить и определить перспективные направления исследований в области машинного перевода (МП).

Систематизация знаний на основе ТРИЗ-эволюционного подхода

Развернувшаяся в последние десятилетия информационная революция обострила потребности создания и развития новых методов извлечения и систематизации знаний, которые обеспечивали бы возможность изучения и структурирования огромного объема информации за ограниченное время. Наметить пути разрешения данного противоречия позволяет ТРИЗ-эволюционный подход. ТРИЗ (теория решения изобретательских задач) – область знаний, исследующая механизмы развития искусственных систем с целью создания практических методов решения инновационных задач [5]. Благодаря своей высокой эффективности и универсальности ТРИЗ получила международное признание, успешно применяется и развивается во многих

областях деятельности человека, прежде всего в промышленном производстве, науке и образовании [6, 7].

Методология ТРИЗ-эволюционного подхода к систематизации знаний была подробно рассмотрена авторами на примере систем МП с построением ТРИЗ-эволюционной карты и детальным описанием исходного объекта ТРИЗ-эволюции систем МП «перевод слово в слово» [8]. Однако, как показала практика, одной лишь визуализации знаний за счет построения ТРИЗ-эволюционной карты недостаточно для глубокого анализа систематизируемой области. Предлагается дополнить ТРИЗ-эволюционный анализ следующими шагами: оценка развития главных параметров при переходе от одного объекта ТРИЗ-эволюции к другому; определение ключевых проблем, ограничивающих применение в каждой технологии; компактное, но емкое описание решений, за счет которых произошло преодоление выявленных ограничений.

В контексте МП такой анализ позволит получить представление о развитии систем МП, составить карту проблем, требующих решения, что сократит время на выявление актуальных проблем МП за счет систематизации, визуализации и структурирования ключевых данных, а определение релевантных путей совершенствования систем МП облегчит и ускорит постановку задач исследований и формулирования гипотез.

ТРИЗ-эволюционный анализ систем машинного перевода

В качестве объектов ТРИЗ-эволюции можно выделить следующие системы МП: дословный МП [9], трансферный МП [10], интерлингвистический МП [11], МП на примерах [12], интерактивный МП [13], статистический МП по словам [14], статистический МП по фразам [15], статистический МП на основе синтаксиса [16], нейронный МП [17], нейронный МП «без учителя» [18], адаптированный нейронный МП [19], гибридный МП [20]. Далее рассмотрим краткую характеристику каждого вида указанных систем МП.

В табл. 1 представлен перечень (выявленных в рамках анализа) главных параметров (ГП) систем МП.

Таблица 1

Обозначение	Описание
ГП1	Качество перевода (ГП1.1 – с точки зрения лексической точности; ГП1.2 – с точки зрения грамматической точности; ГП1.3 – с точки зрения стилистической точности; ГП1.4 – с точки зрения единообразия; ГП1.5 – с точки зрения смысловой точности; ГП1.6 – гладкость перевода; ГП1.7 – длинных предложений; ГП1.8 – коротких предложений; ГП1.9 – текстов узких тематик).

Обозначение	Описание
ГП2	Количество возможных смысловых значений одной лексической единицы.
ГП3	Работоспособность системы.
ГП4	Количество языковых пар.
ГП5	Время на подготовку обучающих данных (ГП5.1 – двуязычных словарей; ГП5.2 – параллельных корпусов; ГП5.3 – правил; ГП5.4 – моно-корпусов).
ГП6	Сложность структуры текста оригинала.
ГП7	Объем обучающих данных (ГП7.1 – правил; ГП7.2 – корпусов текста).
ГП8	Время на разработку системы.
ГП9	Трудозатраты на сопровождение системы.
ГП10	Количество вариантов сочетаний слов.
ГП11	Трудозатраты на пред-, постредактирование.
ГП12	Время на обучение системы.
ГП13	Вероятность грубых ошибок.
ГП14	Объем исходного текста.
ГП15	Время на поиск ошибок в переведенном тексте.
ГП16	Количество переводимых тематик.
ГП17	Сложность системы.

Для каждой из систем произведен подробный анализ с выявлением ключевых проблем, ограничивающих рост идеальности, на основании которых сформулированы противоречия. Спецификация выявленных проблем и противоречий (ПР) в системах МП представлена в табл. 2.

Таблица 2

Описание проблемы	Противоречие
Дословный МП	
Практически невозможно язык представить набором правил, следовало бы также учесть и все возможные исключения из правил (неправильные глаголы в английском, плавающие приставки в немецком, суффиксы, диалекты, сленг и т.д.).	ПР 2.1: при повышении качества перевода недопустимо увеличивается объем обучающих данных (правил).
Количество правил в каждом языке огромно и для качественной проработки необходимых обучающих данных требуется огромное количество человеко-часов.	ПР 2.2: при повышении качества перевода недопустимо увеличивается время на подготовку обучающих данных.
Требуется постоянно поддерживать лингвистическую базу в актуальном состоянии, так как язык – динамическая система.	ПР 2.3: при повышении качества перевода недопустимо увеличиваются трудозатраты на сопровождение системы.
Трансферный МП	
С одной стороны, можно задать общие правила переноса грамматической струк-	ПР 3.1: при сокращении времени на разработку системы недопустимо увеличивается

Описание проблемы	Противоречие
туры, что упрощает задачу перевода, с другой стороны, – сочетаний слов намного больше, чем самих слов, и каждый вариант почти невозможно учесть.	объем обучающих данных (правил).
	ПР 3.2: при сокращении количества правил недопустимо снижается качество перевода возможных сочетаний слов.
Интерлингвистический МП	
Сложность реализации и отсутствие методов и моделей поиска закономерностей и классификации атрибутов текста для создания унифицированного языка и его структуры.	ПР 4.1: при повышении качества перевода недопустимо увеличивается время на разработку системы.
МП на примерах	
Примеры содержат слова, словосочетания и даже предложения, но, фактически, мы находим дословно схожие части, не учитывая особенности синтаксиса, морфологии, грамматического строя и т.д. Чтобы учесть все возможные варианты, необходимо больше обучающих данных.	ПР 5.1: при повышении качества перевода объем обучающих данных (корпусов текста) недопустимо увеличивается.
	ПР 5.2: при увеличении количества вариантов сочетаний слов в тексте оригинала качество перевода с точки зрения единообразия недопустимо снижается.
Даже если объем обучающих данных достаточный, система не делит предложение на структурные части из-за чего, например, служебные части речи, влияющие на контекст, могут отразиться на качестве перевода.	ПР 5.3: при повышении качества с точки зрения единообразия недопустимо снижается качество передачи смысла.
Из-за разницы структуры языков перевода опущенные (нулевые) части предложений не учитываются в переводе либо переводятся части предложения, которые ввиду грамматических правил языка перевода должны быть опущены.	ПР 5.4: при повышении качества с точки зрения единообразия недопустимо снижается качество перевода грамматической точности.
Система не может работать автоматически без участия человека.	ПР 6.1: при повышении качества перевода недопустимо увеличиваются трудозатраты на пред-, постредактирование.
Статистический МП по словам	
При подготовке обучающих данных необходимо максимально точное соответствие оригинала и перевода, но не всегда перевод может быть строго формализован, есть еще литературные или вольные переводы, которые также необходимо учитывать.	ПР 7.1: при повышении качества перевода недопустимо увеличивается время на подготовку обучающих данных (корпусов).
Из-за отсутствия двуязычных словарей между некоторыми языками, система переводит текст сначала на английский, а	ПР 7.2: при повышении количества языковых пар недопустимо снижается качества перевода.

Описание проблемы	Противоречие
затем на язык перевода, из-за чего возникают «двойные потери» качества.	
Статистический МП по фразам	
Статистические аномалии.	ПР 8.1: при увеличении объема обучающих данных недопустимо снижается качество перевода с точки зрения смысловой точности.
Отдельные фразы плохо согласуются между собой, в итоге переведенное предложение – набор фраз, иногда не связанных по смыслу.	ПР 8.2: при повышении качества с точки зрения лексической точности недопустимо снижается гладкость перевода.
	ПР 8.3: при повышении качества с точки зрения лексической точности недопустимо снижается смысловая точность.
Статистический МП на основе синтаксиса	
Даже для корпуса с простейшими 2-3 уровневými деревьями время обучения слишком велико, а значит, на практике система не может быть применима.	ПР 9.1: при повышении качества перевода с точки зрения грамматической точности время на обучение системы недопустимо увеличивается.
Не для всех языков разработаны методы синтаксического разбора. Не для всех языков, для которых разработаны методы синтаксического анализа, они работают достаточно качественно.	ПР 9.2: при повышении количества языковых пар недопустимо снижается качество перевода.
Нейронный МП	
В целом «гладкий» перевод может содержать грубые лексические ошибки.	ПР 10.1: при повышении гладкости перевода недопустимо повышается вероятность в нем грубых ошибок.
Почти любой перевод требует понимания контекста нескольких предложений, иногда это имеет решающее значение для точного перевода с точки зрения используемой лексики. Нейронная система не может анализировать и хранить информацию о контексте текста большого объема и эффективно ее запоминать.	ПР 10.2: при увеличении объема исходного текста недопустимо снижается качество перевода с точки зрения смысловой точности.
Зависимость от состава обучающих данных: если тренировать нейронный перевод только на длинных парах предложений, система будет неспособна перевести корректно короткое предложение и даже отдельное слово.	ПР 10.3: при повышении качества перевода длинных предложений недопустимо снижается качество перевода коротких предложений.
Аномалии в переводе: пропущенные отрицания, отдельные слова или целые фразы.	ПР 10.4: при повышении гладкости перевода недопустимо увеличивается время на по-

Описание проблемы	Противоречие
Аномалии непредсказуемы и непоследовательны, что затрудняет их автоматическое выявление и исправление.	иск возможных ошибок.
Низкое качество перевода исходных текстов, которые сильно отличаются от данных, использованных для машинного обучения.	ПР 10.5: при увеличении количества тематик исходного текста недопустимо увеличивается время на подготовку обучающих данных.
	ПР 10.6: при увеличении количества тематик исходного текста недопустимо снижается качество перевода.
Адаптированный нейронный МП	
По узким тематикам чрезвычайно сложно производить сбор и обработку релевантных двуязычных корпусов для обучения нейронной сети в достаточном объеме.	ПР 11.1: при повышении качества перевода текстов узкой тематики и/или редкой языковой пары недопустимо увеличивается время на подготовку обучающих данных.
Нейронный МП «без учителя»	
Необходимость подготовки монокорпусов большого объема по редким языкам и/или тематикам в векторном представлении.	ПР 12.1: при повышении качества перевода редких языковых пар или тематик объем необходимых обучающих данных недопустимо увеличивается.
Гибридный МП	
При комбинировании разных систем достигается более высокое качество перевода, но это означает усложнение системы, которая наследует не только преимущества, но и недостатки систем, входящих в ее состав.	ПР 13.1: при повышении качества перевода недопустимо увеличивается сложность конечной системы.
	ПР 13.2: при увеличении количества тематик недопустимо увеличивается время на разработку системы.
	ПР 13.3: при увеличении количества языковых пар недопустимо увеличивается время на разработку системы.

В табл. 3 проведен анализ выявленных противоречий и решений (Р), при помощи которых были разрешены противоречия в описанных системах, описаны итерации ТРИЗ-эволюции систем МП.

Таблица 3

Итерация	Описание решения
Дословный МП	
ПР2.1: ГП1↑, ГП7.1↑ → Р11: ГП7.1↓	Переводить с погрешностью, предлагать пользователю скорректировать перевод с целью улучшения качества. <i>Реализовано в системах интерактивного МП.</i>
ПР2.1: ГП1↑, ГП7.1↑ → Р12: {ГП1.4↑, ГП1.3↑, ГП7.1↓}	Не описывать правила вообще, загрузить в систему большое количество примеров, из которых система будет брать готовые части перевода. <i>Реализовано в системах МП на примерах.</i>

Итерация	Описание решения
ПР2.2: ГП1↑, ГП5↑ → Р13: ГП5↓	Выделить ключевые синтаксические конструкции. Заложить в систему правила перевода каждого слова и подстановки в соответствии с синтаксическими конструкциями языка перевода. <i>Реализовано в системах трансферного МП.</i>
ПР2.2: ГП1↑, ГП5↑ → Р14: ГП5↓	Создать промежуточный алгоритм обучения системы: на основе статистических методов система будет выбирать наиболее статистически вероятный перевода, при этом нет необходимости описывать правила. <i>Реализовано в системах СМП по словам.</i>
ПР2.3: ГП1↑, ГП9↑ → Р15: ГП9↓	Периодически подгружать в систему новые параллельные корпуса текста, на которых система доучивается. <i>Реализовано в системах СМП по словам.</i>
Трансферный МП	
ПР3.1: ГП9↓, ГП7.1↑ → Р16: ГП7.1↓	Не описывать правила, загрузить в систему большое количество примеров, из которых система будет брать готовые части перевода. <i>Реализовано в системах МП на примерах.</i>
ПР3.1: ГП9↓, ГП7.1↑ → Р17: ГП7.1↓	Создать промежуточный алгоритм обучения системы: на основе статистических методов система будет выбирать наиболее статистически вероятный вариант перевода, при этом нет необходимости описывать правила. <i>Реализовано в системах СМП по словам.</i>
ПР3.2: ГП7.1↓, ГП1.1↓ → Р18: ГП1.1↑	Найти все возможные сочетания слов в тексте, при помощи статистических методов найти наиболее вероятный перевод. <i>Реализовано в системах СМП по фразам.</i>
ПР3.2: ГП7.1↓, ГП1.1↓ → Р19: ГП1.1↑	Предварительно обрабатывать оригинал для того, чтобы он был более простым и «понятным» машине, переводить с погрешностью, привлекая постредактора для повышения качества перевода. <i>Реализовано в системах интерактивного МП.</i>
Интерлингвистический МП	
ПР4.1: ГП1↑, ГП8↑ → Р20: ГП8↓	Для редких языков и/или тематик из-за сложности создания и разметки универсальной модели и сбора обучающих данных использовать английский язык в качестве промежуточного. <i>Реализовано в системах СМП по словам.</i>
ПР4.1: ГП1↑, ГП8↑ → Р21: {ГП1↑, ГП8↓}	Для кодировки текста в универсальное представление использовать многослойную нейронную сеть. <i>Реализовано в системах нейронного МП.</i>
МП на примерах	
ПР5.1: ГП1↑, ГП7.2↑ → Р22: ГП1.1↑	Перед копированием перевода выбрать из множества вариантов наиболее вероятный на основе статистических методов. <i>Реализовано в системах СМП по словам.</i>
ПР5.2: ГП10↑, ГП1.4↓ → Р23: ГП1.4↑	Записывать порядок слов в предложении относительно друг друга, при переводе выбирать наиболее вероятный вариант на основе статистических методов. <i>Реализовано в системах СМП по словам.</i>

Итерация	Описание решения
ПР5.4: ГП1.5↑, ГП1.2↓ → Р24:ГП1.2↑	Заранее промаркировать места, где необходимо подставить служебное слово, при переводе заполнить маркеры. <i>Реализовано в системах СМП по словам.</i>
ПР5.2: ГП10↑, ГП1.4↓ → Р25: {ГП1.4↑, ГП1.1↑}	Найти все возможные сочетания слов в тексте, при помощи статистических методов найти наиболее вероятный перевод. <i>Реализовано в системах СМП по фразам.</i>
ПР5.3: ГП1.4↑, ГП1.5↓ → Р26:ГП1↑	При помощи статистических методов находить среди множества вариантов обучающего корпуса наиболее вероятный перевод целых фраз, а не отдельных слов. <i>Реализовано в системах СМП по фразам.</i>
{ПР5.2: ГП10↑, ГП1.4↓, ПР5.4: ГП1.5↑, ГП1.2↓} → Р27: {ГП1.4↑, ГП1.2↑}	Перед переводом производить полный синтаксический разбор предложения, структурируя части текста при помощи деревьев. <i>Реализовано в системах СМП на основе синтаксиса.</i>
Интерактивный МП	
ПР6.1: ГП1↑, ГП11↑ → Р28: {ГП1↑, ГП11↓}	Использовать в качестве посредника между машинным переводом и человеком среду автоматизации переводов, объединив все цифровые инструменты, облегчающие перевод в одной среде. Использовать МП для тех фрагментов, для которых не нашлось совпадений по памяти перевода, при этом выбирать модель перевода исходя из особенности задачи. <i>Реализовано в системах гибридного МП.</i>
Статистический МП по словам	
ПР7.1: ГП1↑, ГП5.2↑ → Р29: ГП5.2↓	Объединить отдельные слова в предложении в фразы во всех возможных сочетаниях по n-слов в каждом и проанализировать перевод для каждого из них, выбрав впоследствии только статистически наиболее вероятные сочетания, т.е. учитывать контекст. <i>Реализовано в системах СМП по фразам.</i>
ПР7.2: ГП4↑, ГП1↓ → Р30: {ГП1↑, ГП4↑}	В качестве алгоритма вывода использовать нейронные сети, состоящие из энкодера и декодера. Энкодер сначала кодирует исходный текст в универсальное представление. Декодер распознает универсальное представление и переводит его на язык перевода. <i>Реализовано в системах нейронного МП.</i>
Статистический МП по фразам	
{ПР8.2: ГП1.1↑, ГП1.6↓, ПР8.3: ГП1.1↑, ГП1.5↓} → Р31: ГП1↑	Объединить в одной системе несколько типов систем, воспользовавшись преимуществами каждой из них, разделив задачи. Например: перевод отдельных слов по словарю, а «выравнивание» текста доверить СМП или НМП; НМП переводить длинные предложения, СМП переводить короткие фразы. <i>Реализовано в системах гибридного МП.</i>
{ПР8.2: ГП1.1↑, ГП1.6↓, ПР8.3: ГП1.1↑, ГП1.5↓} →	Учитывать в переводе контекст всего предложения, а не только отдельных фраз при помощи многослойных нейронных сетей с долгосрочной памятью. <i>Реализовано в системах нейронного</i>

Итерация	Описание решения
Р32: ГП1↑	<i>МП.</i>
ПР8.1: ГП7↑, ГП1.5↓ – Решение отсутствует	
Статистический МП на основе синтаксиса	
{ПР9.1: ГП1.2↑, ГП12↑, ПР9.2: ГП4↑, ГП1↓} → Р33: ГП1↑	Использовать многослойную нейронную сеть, не использовать методы синтаксического разбора и анализа. <i>Реализовано в системах нейронного МП.</i>
Нейронный МП	
ПР10.3: ГП1.7↑, ГП1.8↓ → Р34: {ГП1.7↑, ГП1.8↑}	Объединить в одной системе два типа систем, выбрав наиболее подходящую технологию отдельно для длинных предложений, отдельно – для коротких фраз. <i>Реализовано в системах гибридного МП.</i>
ПР10.5: ГП16↑, ГП5↑ → Р35: ГП5.2↓	Обучать сеть алгоритмом без учителя, исключив тем самым необходимость подготовки эталонной выборки данных. <i>Реализовано в системах НМП «без учителя».</i>
ПР10.6: ГП16↑, ГП1.9↓ → Р36: ГП1.9↑	Для каждой из тематик готовить корпуса обучающих данных отдельно. <i>Реализовано в системах адаптированного НМП.</i>
ПР10.1: ГП1.6↑, ГП13↑ – Решение отсутствует	
ПР10.2: ГП14↑, ГП1.5↓ – Решение отсутствует	
ПР10.4: ГП1.6↑, ГП15↑ – Решение отсутствует	
Адаптированный НМП	
ПР11.1: ГП1.9↑, ГП5.2↑ → Р37: ГП5.2↓	Обучать нейронную сеть алгоритмом «без учителя» на монокорпусах текста вместо параллельных корпусов. <i>Реализовано в системах НМП «без учителя».</i>
НМП «без учителя»	
ПР12.1: ГП1↑, ГП7↑ – Решение отсутствует	
Гибридный МП	
ПР13.1: ГП1↑, ГП17↑ – Решение отсутствует	
ПР13.2: ГП16↑, ГП8↑ – Решение отсутствует	
ПР13.3: ГП4↑, ГП8↑ – Решение отсутствует	

Анализ ТРИЗ-эволюции систем МП

Анализ ТРИЗ-эволюции систем МП показывает в первую очередь качественные скачки развития, которые позволили сформировать новые парадигмы: переход к трансферному МП привел к созданию парадигмы «перевод, основанный на правилах»; переход к СМП по словам – к созданию парадигмы «статистический МП»; переход к НМП – к созданию парадигмы «нейронный МП».

По улучшаемым параметрам в ходе развития систем МП мы видим, что оно происходило по следующим ключевым направлениям:

повышение качества перевода, причем с развитием МП сам параметр качества трансформировался и детализировался в подпараметры;

сокращение времени на сбор и подготовку обучающих данных;
развитие способов обработки оригинала с целью упрощения его структуры для более точного понимания системой семантики текста;
совершенствование технической реализации МП.

Часть противоречий существующих систем не разрешена (8.1, 10.1, 10.2, 10.4, 12.1, 13.1, 13.2, 13.3), и, следовательно, задачи по разрешению этих противоречий являются перспективными направлениями исследования в области МП.

На основании проведенного анализа можно заключить, что ключевыми направлениями перспективных исследований в области МП являются:

1) *развитие методов интерактивного перевода и нативного взаимодействия с пользователями МП.* Проблема заключается в том, что, несмотря на огромный прорыв в качестве нейросетевого машинного перевода, он все еще может содержать грубые ошибки, поскольку система не может достаточно точно определять контекст текста целиком, а не отдельного предложения, оттенки смысла и т.д.;

2) *разработка эффективных методов анализа и алгоритмов хранения нейросетью контекста всего текста, а не отдельных его частей.* Актуальным является исследование и совершенствование способов определения и хранения информации о контексте (смысле) не только соседних слов, но и предложений, а также оценка смысла всего текста;

3) *совершенствование архитектуры систем МП и нейросетей, используемых в задачах обработки естественного языка;*

4) *разработка и совершенствование методов обучения нейросетей на ограниченном объеме монокорпусов текста.* Современные системы нейросетевого перевода обучаются «с учителем» на двуязычных параллельных корпусах текста (оригинал и эталонный перевод). Не для всех языков и тематик возможно аккумулировать достаточный для качественного обучения объем таких корпусов. Интерес представляет развитие алгоритмов обучения нейросетей на монокорпусах (один язык, без перевода) «без учителя»;

5) *решение проблемы извлечения знаний и анализа семантики оригинальных текстов.* Проблема извлечения знаний актуальна для многих областей, связана она в том числе и с обработкой естественного языка;

6) *совершенствование методов синтаксического разбора текста для упрощения структуры оригинальных текстов при переводе.* Не для всех языков разработаны методы синтаксического разбора и не для всех, для которых разработаны такие, они достаточно качественные;

7) *совершенствование методов автоматической оценки и проверки*

качества переведенных текстов. Проблема заключается в том, что ни методы автоматической оценки качества, ни методы «ручной» оценки качества не являются надежными и достаточно достоверными. Первые не опираются на семантику (смысл) текста, а вторые зависят от компетенции проверяющего специалиста. При этом в обоих случаях высока вероятность пропуска грубых ошибок, учитывая общую «гладкость» нейронного машинного перевода.

Описанные направления развития систем МП относятся к задачам автоматизации извлечения знаний, обработки естественного языка, нейросетевого программирования.

Заключение

Предложено использовать ТРИЗ-эволюционный подход к выявлению направлений развития гибридных систем нейронного МП, сформулированы дополнительные положения ТРИЗ-эволюционного анализа, позволяющие систематизировать исследуемую область знаний с большей детализацией.

Применение ТРИЗ-эволюционного анализа систем МП позволило систематизировать данные об эволюции систем МП; определить ключевые этапы развития систем МП; выделить главные производственные параметры, определяющие направления развития систем нейронного МП.

Показано, что перспективным направлением является развития методов анализа, сбора и хранения нейросетью данных о семантических особенностях оригинального текста целиком, а не отдельных его частей.

Таким образом, представленное исследование позволит добиться повышения качества перевода при ограниченном объеме обучающих данных за счет развития методов и алгоритмов предобработки оригинального текста.

Создание модуля автоматизированного предредактирования согласно правилам, которые упрощают понимание текста системой МП и в полуавтоматическом режиме позволяют производить анализ семантики оригинальных текстов, повысит качество МП и дает возможность сделать МП более доступным для пользователей, владеющих только языком оригинального текста.

ЛИТЕРАТУРА

1. *Wu Y., Schuster M., Chen Z., et al.* Google's neural machine translation system: Bridging the gap between human and machine translation // Conference on Empirical Methods in Natural Language Processing. – 2017.
2. *Screen B.* What effect does post-editing have on the translation product from an end-user's perspective? // J. Spec. Transl. – 2019. – Vol. 31 – P. 133-157.
3. Сайт о развитии отрасли переводов [Электронный ресурс]. – Режим доступа: <https://translationrating.ru/>

4. Science Direct Journals & Books [Электронный ресурс]. – Режим доступа: <https://www.sciencedirect.com/>
5. *Altshuller G.* The innovation algorithm: TRIZ, systematic innovation, and technical creativity. – Worcester, Massachusetts: Technical Innovation Center, 1999.
6. *Gadd K.* TRIZ for Engineers: Enabling Inventive Problem Solving. – John Wiley & Sons, 2011.
7. Creativity, learning techniques and TRIZ / T. Bertonecelli, O. Mayer, M. Lynass // *Procedia CIRP*. – 2016. – Vol. 39. – P. 191-196.
8. *Zhivotova A.A., Berdonosov V.D., Redkolis E.V.* Machine translation systems analysis and development prospects // 2020 International Multi-Conference on Industrial Engineering and Modern Technologies. – Vladivostok, 2020.
9. *Brown A.* Automatic translation of languages // Automatic Translation of Languages: papers presented at NATO Summer School – Venice, 1966. – P. 27-54.
10. *Tosh L.W.* Data preparation for syntactic translation // International Conference on Computational Linguistics – New York, 1965.
11. *Tosh L.W.* Stratificational grammar and interlingual mapping for automatic translation // Actes du Xe Congrès International des Linguistes. – Bucarest, 1967, – P.1049-1059.
12. *Aramaki E., Kurohashi S.* Example-based machine translation using structural translation examples // International Workshop on Spoken Language Translation: Evaluation Campaign on Spoken Language Translation. – Japan, 2004. – P. 91-94.
13. *Nepveu L., Lapalme G., Langlais P.* Adaptive language and translation models for interactive machine translation // Conference on Empirical Methods in Natural Language Processing. – 2004.
14. A statistical approach to machine translation / P. Brown, J. Cocke, S. Della Pietra, et al // Computational linguistics. – 1990. – Vol. 16, No.2. – P.79-85.
15. *Junczys-Dowmunt M., Grundkiewicz R.* Phrase-based machine translation is state-of-the-art for automatic grammatical error correction // Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing – 2016. – P. 1546-1556.
16. Yamada K., Knight K. A syntax-based statistical translation model // Proceedings of the Conference of the Association for Computational Linguistics. – 2001. – P. 132-139.
17. Google's multilingual neural machine translation system: enabling zero-shot translation / M. Johnson, M. Schuster, Q. Le, et al // Transactions of the Association for Computational Linguistics. – 2017. – Vol. 5. – P. 339-351.
18. *Lample G., Conneau A., Denoyer L., et al.* Unsupervised machine translation using monolingual corpora only // ICLR 2018. – 2018.
19. *Chu C., Wang R.* A survey of domain adaptation for neural machine translation / Proceedings of the 27th International Conference on Computational Linguistics. – 2018. – P. 1304-1319.
20. *Costa-Jussà M., Fonollosa J.* Latest trends in hybrid machine translation and its applications / Computer Speech & Language. – 2014. – Vol. 32. – No. 1. – P. 3-10.

Статья представлена к публикации членом редколлегии Е.А. Шеленком.

E-mail:

Животова Алена Анатольевна – zhivotova.aa@gmail.com;

Бердоносков Виктор Дмитриевич – berd1946@gmail.com.