

Рублёва Ольга Сергеевна

доцент

ФГБОУ ВО «Вятский государственный университет»

г. Киров, Кировская область

DOI 10.21661/r-599162

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПРОМПТ-ИНЖЕНИРИНГА В НЕЙРОННОМ МАШИННОМ ПЕРЕВОДЕ

Аннотация: в статье предпринята попытка систематизировать теоретические основания промпт-инженеринга применительно к нейронному машинному переводу (НМП). Рассматриваются архитектурные особенности НМП, лингвистические факторы, влияющие на результативность промптов, а также некоторые приёмы их структурной оптимизации. Затрагиваются границы применимости данного подхода в деловой коммуникации, предлагается типология промптов. Цель работы – выявить и упорядочить те лингвистические параметры, которые требуют контроля при формулировке запроса, чтобы свести к минимуму переводческие ошибки в специализированных контекстах.

Ключевые слова: промпт-инженеринг, нейронный машинный перевод, лингвистическая оптимизация, синтаксическая структура, семантическая неоднозначность, прагматика, структурные методы.

Нейронный машинный перевод сегодня прочно занял позицию основной технологии автоматического перевода, и это неудивительно: он выдаёт связные и естественные тексты, чего раньше не могли обеспечить ни статистические, ни тем более правила-ориентированные системы. Однако было бы ошибкой полагать, что НМП свободен от лингвистических проблем. Многозначность, культурно-обусловленные выражения, расхождения в синтаксическом строе языков, не говоря уже о прагматических нюансах, – всё это по-прежнему создаёт серьёзные трудности для нейросетей. Именно здесь на сцену выходит промпт-инженеринг, который можно рассматривать как дисциплину, сосредоточенную на том, как грамотно сконструировать входной запрос, чтобы направить модель в нужное

русло. Пока что этот инструмент остаётся скорее недооценённым, хотя потенциал у него немалый. Задача, которую я ставлю в данной работе, – разобраться, какие именно лингвистические параметры влияют на эффективность промптов, и предложить ориентиры для их настройки, особенно в ситуациях, когда речь идёт о международной деловой переписке или документации. Методологически статья опирается на обзор и критическое осмысление имеющейся литературы по теме, а также на попытку систематизировать языковые факторы по уровням (синтаксис, семантика, прагматика). Кроме того, я использую элементы сопоставительного анализа английского и русского языка, поскольку их структурные различия часто становятся камнем преткновения при автоматическом переводе.

Что касается устройства самого НМП, то здесь стоит напомнить: это технология, основанная на глубоких нейросетях. В своём развитии она прошла три стадии – от систем с жёстко прописанными правилами через статистические методы к современным архитектурам, которые учатся на огромных массивах параллельных текстов [1]. Каждый этап давал прирост качества, но именно нейронные сети обеспечили тот самый скачок, когда перевод перестал звучать как машинный. Впрочем, вместе с новыми возможностями появилась и потребность в более тонком управлении результатом, что в итоге и привело к оформлению промпт-инжиниринга как отдельной области [4]. Если заглянуть внутрь типичной архитектуры, то мы увидим кодировщик и декодировщик: первый превращает исходный текст в некоторое векторное представление, второй на его основе шаг за шагом строит перевод [2]. Среди технических решений особо выделился трансформер с его механизмами внимания – они позволяют модели по-разному «взвешивать» слова исходного предложения при генерации каждого следующего слова [7]. Веса нейронных связей настраиваются в процессе обучения на больших корпусах, и сеть учится обобщать закономерности, но вот парадокс: чем мощнее модель, тем выше ожидания пользователей, а значит, и тем острее проблема контроля. Поэтому, чтобы промпт-инжиниринг был не гаданием, а осмысленной практикой, полезно понимать, как именно модель воспринимает контекст и

инструкции, – это напрямую связано с внутренней логикой работы механизмов внимания.

Сам промпт-инжиниринг определяют как деятельность по генерации и оптимизации входных запросов для нейросетевых моделей – в том числе и для переводческих [4]. В контексте НМП он играет роль мостика между тем, что хочет получить человек, и тем, как алгоритм интерпретирует входные данные. Если сформулировать запрос небрежно, модель может выдать текст с неверным стилем, не тем тоном, а то и вовсе с искажённым смыслом. Грамотно же составленный промпт способен решить сразу несколько задач: задать нужный регистр и формат, снизить неоднозначность за счёт чётких указаний, а также повысить релевантность перевода – особенно это касается деловой и узкоспециализированной лексики, где цена ошибки высока. И здесь нельзя не отметить, что в международном деловом общении цена такой ошибки становится ещё более осязаемой. Контракты, финансовые отчёты, официальная переписка – все эти жанры жестко стандартизированы и крайне чувствительны к культурным различиям. Малейшая неточность в формулировке промпта может обернуться не просто искажением фразы, а реальными юридическими или репутационными проблемами. Примеры таких случаев, к сожалению, не редкость.

Теперь о факторах, которые реально определяют качество перевода. Начну с синтаксиса, поскольку он задаёт каркас любого высказывания. От того, насколько сложно устроено предложение, зависит, сможет ли модель корректно вычленить смысловые блоки и установить между ними связи [3]. Длинные периоды с несколькими уровнями придаточных, частые инверсии и нестандартный порядок слов – для нейросети это серьёзный вызов. Особенно остро проблема стоит при переводе с английского на русский: их синтаксические системы устроены принципиально по-разному, и то, что естественно для одного языка, может внести хаос в другой. Поэтому при конструировании промпта я бы рекомендовала не пренебрегать простотой: дробить длинные предложения на более короткие, избегать лишних вложенных конструкций и, если нужно, прямо указывать

модели, какой порядок слов и какую структуру придаточных следует предпочесть в целевом тексте.

Перехожу к семантике. Здесь главный враг – многозначность. Нейросети, как бы хорошо они ни были обучены, довольно часто спотыкаются о слова, у которых есть несколько значений, особенно в деловой. Омонимы и полисемия способны порождать прямо противоположные толкования. Вот, скажем, английское *claim* – в зависимости от контекста это может быть и «заявление», и «иск», и «претензия», и даже «требование по гарантии». Выбор неверного варианта, особенно в юридическом документе, – это уже не просто лексическая ошибка. Выход здесь видится в том, чтобы в самом промпте проставлять смысловые маркеры или даже давать краткий глоссарий на целевом языке, сужая поле интерпретации. Отдельного внимания заслуживает практика контролируемого языка, когда используется ограниченный лексикон и строго очерченные грамматические конструкции; такой подход заметно снижает риски [5]. И, пожалуй, стоит добавить, что для деловых текстов эта стратегия почти обязательна, поскольку цена семантической ошибки здесь – прямые финансовые потери или судебные иски.

Наконец, прагматика. Этот уровень часто остаётся в тени, хотя именно он отвечает за то, как высказывание будет воспринято адресатом. Идиомы, метафоры, культурно-специфические реалии – всё это для машинного перевода представляет собой едва ли не самую трудную зону. Буквальный перевод в таких случаях не просто искажает смысл, а делает текст неуместным, а иногда и оскорбительным [6]. Деловая коммуникация – как раз тот случай, где культурные нормы диктуют степень формальности, допустимые способы выражения просьбы или тональность обращения. И если промпт не содержит соответствующих указаний, модель с высокой вероятностью выдаст перевод, который формально правилен, но практически негоден. Поэтому я считаю, что в запросе обязательно нужно задавать уровень вежливости, предпочтительные стилистические клише и обязательно предупреждать о тех элементах, которые требуют особой культурной адаптации. В итоге, синтаксис, семантика и прагматика – это не три изолированных блока, а взаимосвязанная система: изменения в одном немедленно

отзываются в двух других, и при разработке промпта нужно держать в уме все три одновременно.

Стоит, однако, честно признать, что промпт-инжиниринг имеет вполне конкретные ограничения. Он не переучивает модель и не меняет её внутренних весов; он лишь направляет её внимание, а потому, если модель изначально плохо знакома с предметной областью, чудес не произойдёт. Кроме того, разные архитектуры по-разному реагируют на инструкции, и универсального рецепта просто не существует. Именно поэтому разумно рассматривать промпт-инжиниринг как дополнительный, а не основной механизм обеспечения качества. Без пост-редактирования и хотя бы беглой экспертной проверки в серьёзных проектах обходиться пока что не получается.

Что касается дальнейших исследований, то здесь можно выделить несколько направлений. Одно из них – эмпирическая проверка того, как разные типы промптов ведут себя на разных жанрах деловой документации: контрактах, письмах, технической документации. Другое – попытка разработать более или менее формализованные метрики, которые позволили бы оценивать качество самого промпта как лингвистического продукта. И, наконец, интересно было бы понять, влияет ли порядок инструкций внутри промпта на итоговый результат и если да, то в какой степени. Тем не менее, при всех оговорках, промпт-инжиниринг остаётся на сегодня самым быстрым и доступным способом повысить качество перевода, не прибегая к дорогостоящему дообучению модели, – и это его важнейшее практическое преимущество.

Список литературы

1. Линцов Л.А. Нейронный перевод: краткая история и перспективы / Л.А. Линцов, Е.Ю. Бабанина, Т.И. Кузнецова // Успехи в химии и химической технологии. – URL: <https://cyberleninka.ru/article/n/neyronnyy-perevod-kratkaya-istoriya-i-perspektivy>.
2. Мифтахова Р.Г. Машинный перевод / Р.Г. Мифтахова, Е.А. Морозкина // Вестник Башкирского университета. – 2019. – №2. – С. 497–502. – EDN VWVCPH

3. Новожилова А.А. Машинные системы перевода: качество и возможности использования / А.А. Новожилова // Вестник Волгоградского государственного университета. – URL: <https://cyberleninka.ru/article/n/mashinnye-sistemy-perevoda-kachestvo-i-vozmozhnosti-ispolzovaniya>.

4. Раренко М.Б. Машинный перевод: от перевода «по правилам» к нейронному переводу / М.Б. Раренко // Социальные и гуманитарные науки. Отечественная и зарубежная литература. – URL: <https://cyberleninka.ru/article/n/mashinnyy-perevod-ot-perevoda-po-pravilam-k-neyronnomu-perevodu>.

5. Родина С.В. Лингвистические проблемы машинного перевода (на материале русского и английского языков) / С.В. Родина, Е.В. Лакиза // Universum: филология и искусствоведение. – URL: <https://cyberleninka.ru/article/n/lingvisticheskie-problemy-mashinnogo-perevoda-na-materiale-russkogo-i-angliyskogo-yazykov>.

6. Трудности машинного перевода: контекстуальная языковая неопределённость / А.В. Суховерхов, Д. Дьюитт, И.И. Манасиди [и др.] // Вестник Волгоградского государственного университета. Серия 2: Лингвистика. – 2019. – №4. – С. 129–143.

7. Шевченко О.Н. Analyzing neural machine translation / О.Н. Шевченко // Innovative Science. – 2021. – No. 6. – С. 142–144. – EDN APXMEF