



This work is licensed under a Creative Commons Attribution 4.0 International License
<https://creativecommons.org/licenses/by/4.0/>

<https://doi.org/10.24833/2687-0126-2025-7-1-89-107>

ПРИНЦИПЫ РАЗРАБОТКИ КИТАЙСКО-РУССКОГО СЛОВАРЯ МНОГОЗНАЧНОЙ ЛЕКСИКИ В КАЧЕСТВЕ ОБУЧАЮЩЕГО КОРПУСА ДЛЯ НЕЙРОСЕТЕВОГО МАШИННОГО ПЕРЕВОДА

Е.В. Чистова

*Университет мировых цивилизаций им. В.В. Жириновского (Москва, Россия)
kovelena82@mail.ru*

Аннотация: Статья посвящена выявлению качественных характеристик детализации словаря многозначной лексики с целью повышения интерпретативности нейросетевых машинных переводчиков. Исследование выполнено в рамках частной теории перевода, а именно на материале китайско-русской языковой пары как одной из наиболее актуальных в современном переводческом контексте и одновременно проблематичных ввиду значительной межъязыковой и межкультурной асимметрии. Новизна исследования заключается в отсутствии методик и алгоритмов по разработке конечных словарей, позволяющих осуществлять нейросетевой машинный перевод наиболее адекватно современным условиям. Определяя многозначность как одну из причин некачественного автоматического генерирования текста, автор предлагает теоретическое обоснование принципов составления двуязычных словарей на основе ручной обработки параллельных текстов с целью заполнения определенных параметров лингвопереводческих схем для детализации употребления лексических единиц, сложно интерпретируемых нейросетевыми переводчиками. Проблема решается с помощью метода экокognитивного моделирования профессиональной коммуникации переводчиков, позволяющего провести комплексное исследование человеко-машинной коммуникации, на основе которой выделяются многозначные единицы перевода, требующие ручной предобработки. Автор определяет два сценария снятия многозначности у сложно интерпретируемых лексических единиц. По разработанному авторскому алгоритму в статье показаны детали анализа ручной обработки параллельных текстов и заполнение схем оформления словарных статей под нейросетевой переводчик. Обсуждается исследовательская логика по принятию тех или иных решений относительно категоризации и/или параметризации рассматриваемых лексических единиц и их контекстуальных особенностей. В анализе объясняются причины включения рассматриваемых единиц в корпусные данные для обучения нейросетей, иллюстрируется их лексико-грамматическая детализация и сочетаемость, а также приводится переводческий анализ для снятия контекстуальной вариативности за счет категоризации понятий и/или дискурсивной принадлежности. В результате исследования качественными характеристиками детализации параллельного обучающего китайско-русского корпуса выступают: лингвистические и дефиниционные параметры, словарная представленность, переводческая вариативность в зависимости от лексико-грамматической сочетаемости, дискурсивно-жанровой принадлежности и концептуально-категориальной таксономии.

Ключевые слова: обработка естественного языка, машинный перевод, нейронный перевод, полисемия, китайский язык, лексикология, терминология.

Для цитирования: Чистова, Е.В. (2025). Принципы разработки китайско-русского словаря многозначной лексики в качестве обучающего корпуса для нейросетевого машинного перевода *Дискурс профессиональной коммуникации*, 7(1), 89–107. <https://doi.org/10.24833/2687-0126-2025-7-1-89-107>

1. ВВЕДЕНИЕ

Нейросетевой машинный перевод – это технология, основанная на подходе к машинному переводу, в котором используется большая искусственная нейронная сеть. Данная сеть стала первым видом искусственного интеллекта, имитирующим биологические способности человека. Био-вдохновленные сети (Kurakin, 2024) используются в переводческих целях с 2015 года и во многом превзошли ранее существовавшие виды машинного перевода: *Rule-Based MT* – машинный перевод, основанный на правилах, которые описывают языковые структуры и их преобразования (развивается с 1950 гг.), а также *Statistical MT* – статистический перевод, базирующийся на поиске наиболее вероятного перевода предложения с использованием данных, полученных из двуязычной совокупности текстов (развивается с 2000 гг.) (Technologies, 2024).

Принцип работы нейросетевого машинного перевода представляет «механизм двунаправленных рекуррентных нейронных сетей» (*Bidirectional Recurrent Neural Networks*) (Там же), как правило, пользующихся данными параллельных корпусов, т.е. сегментами исходных и переводных текстов, выполненных человеком. Этот механизм построен на «матричных вычислениях и позволяет создавать сложные вероятностные модели» (Там же).

Пользователи нейросетевого машинного перевода отмечают его высокую эффективность в обработке текстов больших объемов, точное соблюдение терминологии, способность адаптировать перевод под конкретный запрос и др. (Машинный перевод. Инновации и влияние на переводческие услуги, n.d.). Однако одной из основных проблем по-прежнему остается перевод многозначной лексики (Wang, 2022). Межязыковая асимметрия различных языковых пар, потребность в контекстуальной верификации, недостаток фоновых знаний нейронных сетей или невозможность их распознавания в заданных ситуациях препятствует достижению требуемого уровня эквивалентности и адекватности.

Решение перечисленных проблем связано не только с техническим усовершенствованием математических алгоритмов, но и с разработкой релевантных параллельных корпусов для их обучения. Иначе говоря, для улучшения качества нейросетевого машинного перевода в аспекте обработки многозначной лексики и национально-детерминированных реалий необходим конечный словарь. Таким образом, цель данной работы – определить качественные характеристики детализации словаря как обучающего корпуса для нейросетей с целью снижения интерпретативности при переводе многозначной лексики. Отметим, что в настоящем исследовании мы используем примеры китайско-русского перевода как одного из наиболее проблематичных.

2. ТЕОРЕТИЧЕСКИЙ ОБЗОР

Проблема снятия многозначности в переводе интересует ученых уже не одно десятилетие. С позиции ручного перевода это решается в рамках интерпретативной теории перевода посредством контекстуального распознавания исходного смысла и принятия творческих решений (Велла, 2013; Убоженко, 2016). Однако такой способ не представляется эффективным на данном этапе развития современного нейросетевого машинного перевода. Для разрешения проблемы некоторые ученые предлагают разрабатывать приемы редактирования текста для того, чтобы вручную нивелировать возможную лексическую неоднозначность (Kokanova et al., 2022; Коканова, Пак, 2024). Другая группа исследователей пытается решить обозначенную проблему с помощью усовершенствования языковых моделей, в частности применения таких архитектур, как трансформеры и *BERT*, ставшие прорывом в проработке многозначности, синонимии и антонимии слов, а также в построении тематических моделей, однако далеко не во всех языковых парах (Ершова, Санникова, 2024). Третьим способом повышения интерпретативной способности нейросетей выступает создание дополнительных словарей или пополнение корпусных данных уже имеющихся словарей (Мифтахова, 2017).

Параллельные данные могут представлять корпуса отдельных символов, подслов, слов, фраз, предложений и даже концептов. Уровень детализации словаря/гlossария/корпуса зависит от поставленной лингвопереводческой задачи. Например, детализация на уровне символов (иероглифов) больше подходит для создания двуязычных словарей между этническими языками. Нейронные сети при этом используют для извлечения шаблонов преобразования языковых структур и правил функционирования (Resiandi et al., 2023). Эксперименты показывают (Casas et al., 2018), что предлагаемые стратегии определения словарного запаса на основе морфологии обеспечивают улучшение или сохранение сопоставимого качества при переводе текстов, не относящихся к предметной области, для языков с богатой морфологией, таких как немецкий и баскский. Без существенной потери качества перевода фиксируются также наблюдения за морфологически бедными языками, такими как английский язык.

Подобные морфологически-сконструированные пословные параллельные корпуса используются для расширения словарного запаса вымышленных языков. Алгоритм реализации подобной лингвистической задачи выглядит следующим образом: готовится ограниченный словарь из нескольких сотен слов и их переводов, на его основе нейронные сети обучают экстраполировать словарный запас языка, сохраняя при этом стиль создателя. В заданном алгоритме нейросети способны на естественном языке генерировать неологизмы, которые соответствуют уже существующим словам. Несмотря на то, что подобная работа сосредоточена на расширении словарей вымышленных языков, данный метод может быть использован для помощи малоресурсным и находящимся под угрозой исчезновения естественным языкам, словари которых со временем сокращаются (Zacharias et al., 2022).

Проблема перевода с малоресурсных языков (например, с египетского диалекта на современный нормативный арабский язык) преодолевается с помощью методологии глубокого обучения на основе экспериментов (Faheem, 2024) с тремя его различными подходами: контролируемый, неконтролируемый и полуконтролируемый методы. Контролируемый метод означает обучение большой языковой модели посредством набора данных, состоящих

из параллельных пар предложений на обоих языках. Полуконтролируемый метод сосредоточен на обучении модели с параллельными корпусами изначально, а затем на дальнейшем обучении с использованием одноязычных корпусов. Неконтролируемый метод включает обучение модели с использованием одноязычных предложений на обоих языках. В результате экспериментов, полуконтролируемый метод объединил сильные стороны как контролируемого, так и неконтролируемого обучения, начав с обучения на параллельных корпусах, а затем перейдя к обучению на одноязычных корпусах. Это демонстрирует эффективность объединения как маркированных (параллельные корпуса), так и немаркированных данных (одноязычные корпуса) в процессе обучения.

Подобный гибридный подход к машинному переводу используется в технологии *Prompt Neural*. Алгоритмы сначала анализируют текст и далее к разным его фрагментам выбирают наиболее релевантный метод переводческой обработки: *Rule-Based MT* – машинный перевод, основанный на лингвистической информации, описывающей языковые структуры и их преобразования исходного языка и языка перевода, или нейросетевой подход (Technologies, n.d.). Гибридную систему выбирают и разработчики «Яндекс.Переводчик», однако наиболее продуктивной они считают комбинацию статистического и нейросетевого подходов. Входной текст сначала обрабатывается двумя способами, а затем алгоритм, основанный на методе обучения *CatBoost*, оценивает полученные варианты переводов и выбирает лучший. Согласно алгоритму оценивания во внимание принимаются многие параметры (длина предложения, синтаксическая структура, контекстуальное окружение и др.). Пользователю предлагается наиболее высокорейтинговый вариант перевода (Компания Яндекс – Технологии – Машинный перевод, n.d.).

Нейросетевой машинный перевод в чистом виде используется в системе *Google Translate*. При обработке текста предложения делятся не на слова, а на сегменты, далее определяется «вес» каждого сегмента в предложении, к которому подбираются наиболее вероятные варианты перевода. На заключительном этапе происходит объединение переведенных сегментов с учетом грамматических и стилистических правил. Несмотря на то, что точность *Google Translate* в самых популярных языковых парах (французский – английский, испанский – английский) достигает 85-87%, нейросетям еще требуется активное пополнение корпусных данных (Как работает нейросеть Google Translate, 2018).

Для китайско-русского нейронного перевода особенно актуальна разработка словаря многозначной лексики. Как показывают исследования (Ко, 2024), даже в специальных текстах за счет значительного присутствия культурного компонента в терминологии, требуется снятие интерпретативности. Наиболее ярко это демонстрируется на примерах машинного перевода текстов по тематике традиционной китайской медицины, где особенно сильно размыты границы между общеупотребительными словами и терминами. Например, нейросеть не идентифицирует в качестве многокомпонентных терминов такие метафоричные словосочетания, как 犛鼻 («телячий нос» – точка иглоукалывания под коленной чашечкой) или 足少阳 («нога, ступня» + «основные сосуды туловища (желудка, желчного и мочевого пузыря)»; «сосуды сердца и почек»; «меридиан почек» (Zhonga. Китайский словарь и переводчик онлайн – Чжунга, n.d.) – «точки меридиана стопа-шаоинь») (Ко, 2024).

По мнению ученых, для снятия интерпретативности национально-маркированных реалий в более продвинутых переводческих целях преимуществом обладают словари на уровне морфологически сегментированных слов (Casas et al., 2018). Для обработки естественного языка нейронные сети обычно разделяют исходную строку символов на последовательность подстрок и представляют каждую подстроку как отдельный токен. Детализация словарей

осуществляется на уровне символов, подслов, слов, кросс-языковых вложений слов и др. В ходе изучения уровней детализации словарей в качестве обучающих корпусов ученые отмечают наиболее высокий потенциал лингвистически управляемой сегментации словарного запаса, которая позволяет морфологически осознавать связь токена со словом. Каждый раз, при добавлении новой лингвистической информации в нейронные системы, они качественно прогрессируют в интерпретируемости многозначной лексики, грамматических конструкций и др.

Для устранения полисемии в научно-технических текстах внедряется метод построения предметной области на основе терминологических словарей. Предметные области конструируются благодаря онтологическим моделям, являющимся по существу графическим отображением концептуальных словарей различных естественных наук (биологии, химии, физики и др.). Для представления онтологии в системах машинного перевода дополнительно создаются: 1) словарь концептов, содержащий экземпляры концептов, атрибуты экземпляров, синонимы и акронимы концепта; 2) таблица бинарных отношений с фиксацией имени концепта-источника и целевого концепта, инверсных отношений и др.; 3) таблица атрибутов экземпляра, содержащая имя атрибута, тип значения, значение «по умолчанию», формулы и правила для вывода и др.; 4) таблица атрибутов класса; 5) таблица экземпляров для каждого входа в словарь концептов (Моренцова, 2019). Однако данный метод эффективен в случае полноценного отображения имеющихся научно-технических концептов в терминологических словарях. На практике оказывается, что многие актуальные языковые единицы не зафиксированы в тезаурусах или описаны в их устаревших/недостовверных/неактуальных значениях. Еще одной проблемой многозначности выступает процесс (де)терминологизации лексических единиц, что также скудно представлено в современных словарях. Таким образом, разработка параллельных корпусных данных на базе ручной обработки естественного языка для обучения нейросетевых машинных переводчиков по-прежнему остается актуальной.

3. МЕТОДОЛОГИЯ И МАТЕРИАЛ ИССЛЕДОВАНИЯ

В исследовании используется метод экокognитивного моделирования профессиональной коммуникации переводчиков, подробно описанный в нашем диссертационном исследовании (Чистова, 2022, с. 185-186). Технология метода предполагает несколько этапов. На первом определяются участники переводческого процесса как взаимодействующие субъекты и объекты материального мира, формирующие единую надындивидуальную когнитивную систему. В данном случае мы выделяем переводчика как субъекта познания, *Google*-переводчик как объект материального мира, способный взять на себя часть когнитивной нагрузки человека, а также цифровую среду как канал человеко-машинной коммуникации.

На втором этапе осуществляется сбор данных: к массиву текстов, предоставленных для исследования переводческим агентством ООО «Диалог», применяется сплошная выборка на основе неадекватных переводческих решений, произведенных нейросетевым переводчиком, по отношению к многозначной лексике. Иными словами, если *Google*-переводчик не справляется с интерпретацией той или иной лексической единицы в соответствии с заданным контекстом, то такая единица становится частью корпусных данных. Стоит отметить, что при отборе важен еще один критерий – неспособность «машин» снять многозначность, т.е. переводчик-исследователь на данном этапе прогнозирует

степень сложности интерпретации того или иного слова/термина человеком и нейросетевым переводчиком, для разных субъектов перевода степень сложности значительно отличается. Например, рассмотрим исходный текст: 东北红烧肘子大拼盘儿, 制作全程流口水, 软软糯糯, 香迷糊了.¹ Google-переводчик предлагает следующий вариант: *Большое блюдо тушеных свиных локтей по-северо-восточному. В течение всего процесса приготовления у меня текли слюнки. Оно было мягким и клейким, и меня смутил аромат. Яндекс-переводчик также дает вариант: «...а его аромат вызывает недоумение».* Помимо других имеющихся недочетов сосредоточимся на переводе слова 香. В электронной версии Большого китайско-русского словаря перечисляются такие варианты перевода, как «приятный запах, ароматный, благоухать» и т.п. (Zhonga. Китайский словарь и переводчик онлайн – Чжунга, n.d.). Однако из контекста мы понимаем, что запах как раз рассказчику не понравился, соответственно, его нельзя назвать ароматом. Человек-переводчик сможет уловить этот нюанс неосознанно и перевести как «но меня смутил запах», в то время как «машине» необходимо уточнить эту разницу за счет контекстуального окружения. Очевидно, что в переводе подобных лексических единиц необходимы дополнительные пояснения и объяснения, которые помогли бы нейронным сетям лучше «чувствовать» контекст.

На третьем этапе устанавливаются «нечувствительные» для «машины» контексты, требующие дополнительных описаний. Это выполняется на основе наблюдений переводчика-исследователя, его способности к прогнозированию, а иногда, для избегания субъективности, становятся необходимы сравнительно-сопоставительные исследования ручных переводов, контекстологический анализ, дефиниционный анализ, лексикографический и терминографический виды анализа, а также опросы и эксперименты с людьми-переводчиками.

На заключительном этапе отобранные многозначные лексические единицы систематизируются, классифицируются и подвергаются подробному описанию по заранее разработанной схеме.

Материалом исследования на втором этапе послужили исходные тексты на китайском языке, предназначенные для реальной переводческой деятельности. Тексты имеют разную жанровую принадлежность: от официально-деловых документов до развлекательного контента на веб-сайтах. В качестве материала исследования на третьем этапе привлекались интернет-статьи, поликодовый и мультимодальный веб-контент, словари и специализированные справочники в заданном виде дискурса. Это помогло обогатить палитру примеров функционирования рассматриваемых лексических единиц в контексте, а также выделить особенности их употребления для последующей категоризации под нейросетевой переводчик.

В качестве тестируемых нейросетевых переводчиков используются Google-переводчик и Яндекс-переводчик. Для демонстрации примеров выбирается наиболее «удачный» вариант одного из автоматических переводчиков, при этом имеющий слабый интерпретативный потенциал в рассматриваемой лексической единице. Проиллюстрируем на том же исходном тексте: 东北红烧肘子大拼盘儿, 制作全程流口水, 软软糯糯....² Яндекс-переводчик

¹ 王冬怎么说[Wang Dong Zenme Shuo]. (2024, March 14). 东北红烧肘子大拼盘儿, 制作全程流口水, 软软糯糯, 香迷糊了 [Northeast red-braised pork knuckle platter, the whole process of making it mouthwatering, soft and sticky, incredibly fragrant] [Video]. YouTube. <https://www.youtube.com/watch?v=9N-WALXaRP8>

² 王冬怎么说[Wang Dong Zenme Shuo]. (2024, March 14). 东北红烧肘子大拼盘儿, 制作全程流口水, 软软糯糯, 香迷糊了 [Northeast red-braised pork knuckle platter, the whole process of making it mouthwatering, soft and sticky, incredibly fragrant] [Video]. YouTube. <https://www.youtube.com/watch?v=9N-WALXaRP8>

предлагает такой вариант: *Рулет из тушеного мяса на северо-востоке в процессе приготовления получается мягким и воскообразным...* Помимо других неточностей и странностей обратим внимание на перевод лексической единицы 拼盘儿. Имеющиеся видеоряд и фотографии, размещенные на сайте, доказывают, что описываемое китайское лакомство никак не может называться рулетом, поскольку на большой тарелке отчетливо просматриваются отдельно расположенные мясные фрагменты. Google-переводчик дает свой вариант: *Большое блюдо тушеных свиных локтей по-северо-восточному. В течение всего процесса приготовления у меня текли слюнки. Оно было мягким и клейким...* Поскольку перевод 拼盘儿 как «большое блюдо» представляется более адекватным в заданном контексте, то в примере будет обсуждаться только вариант, предлагаемый Google-переводчиком.

Таким образом, алгоритм исследования заключается в следующем: 1) выявить неадекватные переводческие решения, выполненные нейросетевыми переводчиками; 2) определить, является ли многозначность причиной неадекватного варианта машинного перевода; 3) установить причину низкого интерпретативного потенциала по отношению к рассматриваемой лексической единице; 4) описать возможные способы снятия многозначности, прогнозируя логику нейросетевого переводчика; 5) внести тестируемую лексическую единицу в словарь; 6) разработать схему оформления словарной статьи под нейросетевой переводчик; 7) в зависимости от классификации отобранных лексических единиц усовершенствовать схему оформления или разработать несколько видов схем под запросы каждого класса.

Представленный алгоритм действий дает комплексное понимание процесса исследования человеко-машинной коммуникации и способствует экспериментальной воспроизводимости. С точки зрения проблематики, выбранные методология и материал исследования представляются релевантными и продуктивными как с позиции теоретического фундамента, так и с позиции получения практической значимости.

4. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ДИСКУССИОННЫЕ АСПЕКТЫ

Сбор корпусных данных позволяет выделить два основных вида многозначной с точки зрения нейросетевого перевода лексики: 1) контекстно-обусловленный и 2) фразеологически-связанный.

Многозначность первого вида может проявляться на уровне смысла – в определенной коммуникативной ситуации, и на формальном уровне – при особом лексико-грамматическом окружении, поддающемся машинной обработке. В качестве одного из вариантов снятия такого рода полисемии ученые предлагают использовать интерактивный подход к процессу нейросетевого перевода, когда при выявлении случаев неоднозначности «машина» предлагает переводчику выбрать наиболее уместный вариант из фильтра значений (Мукабенов, Ахмадуллина, 2023, С. 178). Это действительно представляется эффективным решением проблемы, при условии, что нейросетевой переводчик качественно справляется с распознаванием многозначных слов, количество которых адекватно с точки зрения трудо- и времязатрат человека-переводчика при их корректировке. Помимо встроенных фильтров значений в виде «всплывающего облака», мы считаем, что для китайско-русской языковой пары следует разрабатывать большие массивы описательных схем, предназначенных для обучения нейросетей. В противном случае корректировка при ручном выборе контекстуально-обусловленных значений будет занимать слишком много времени ввиду обилия многозначных слов, не зафиксированных на сегодняшний день в словарях.

Рассмотрим алгоритм исследования на примере лексической единицы 拼盘儿, довольно часто встречающейся в рекламном, развлекательном и коммерческом видах дискурса. Согласно словарю БКРС, 拼盘(儿) – pīnpán (pīnpánr) – это «ассорти из холодных закусок, холодные закуски (еда)» (Zhonga. Китайский словарь и переводчик онлайн – Чжунга, n.d.).

Изученные в Интернет-сети контексты дают понять, что под данной словарной единицей далеко не всегда подразумеваются холодные закуски, например: #好吃到停不下来 #火鸡面套餐拼盘儿.³ Google-переводчик: # Также вкусно, невозможно остановиться # Блюдо с лапшой из индейки (Авторский перевод (далее – АП): # Так вкусно, что невозможно остановиться # Ассорти комплексных обедов из лапши с идейкой).

Более того, предложенные в Словаре варианты перевода выглядят громоздкими, они обладают низким словообразовательным потенциалом, что затрудняет их употребление при встраивании в контекст. В ручном переводе человек чувствует эти стилистические ограничения и выбирает отличные от словарных варианты перевода. Интересно, что нейросетевые переводчики действуют по тому же сценарию.

Во многих примерах 拼盘儿 нейросеть переводит как «блюдо», что в ряде случаев довольно уместно, например: 东北红烧肘子大拼盘儿, 制作全程流口水, 软软糯糯, 香迷糊了.⁴ Google-переводчик: Большое блюдо тушеных свиных локтей по-северо-восточному. В течение всего процесса приготовления у меня текли слюнки. Оно было мягким и клейким, и меня смутил аромат (АП: Локти получились мягкими и клейкими, но меня смутил запах).

В данном предложении «блюдо» представляется наиболее удачным переводом, так как «большое ассорти» в основном употребляется отдельно с последующим описанием его наполнения, например: «Серия подарочных наборов “Большое ассорти”. Состоит из...».

Однако в другом примере перевод 拼盘儿 как «блюдо» представляется не совсем адекватным в связи с несочетаемостью слов:

找五谷杂粮拼盘, 上阿里巴巴, 厂家直销, 源头厂货!⁵ Google-переводчик: Если вы ищете блюдо из цельнозерновых продуктов, зайдите на Alibaba. Прямые продажи с фабрики, исходные фабричные товары! (АП: Если вы ищете ассорти из цельнозерновых продуктов, зайдите на Alibaba. Прямые продажи с фабрики, оригинальная фабричная продукция!). Сами того не осознавая, мы используем лексему «блюдо» только с готовой едой. Если, как в данном примере, в наборе имеются пищевые продукты, требующие обработки, т.е. варки перед употреблением, то перевод «блюдо» звучит неуместно.

В качестве вариантного соответствия 拼盘儿 также иногда встречается «тарелка», например: 酱卤拼盘儿.⁶ Google-переводчик: Тарелка с тушеным соусом. Очевидно, что перевод «тарелка» в данном контексте тоже не совсем адекватно звучит, так как соус у нас подают в отдельной специальной посуде и как сопутствующий компонент к определенному главному блюду, которое в предложенном варианте перевода не было упомянуто. Читая

³ 抖音开放平台 [Douyin Open Platform]. <https://www.douyin.com/video/7405410421840071987>

⁴ 王冬怎么说[Wang Dong Zenme Shuo]. (2024, March 14). 东北红烧肘子大拼盘儿, 制作全程流口水, 软软糯糯, 香迷糊了 [Northeast red-braised pork knuckle platter, the whole process of making it mouthwatering, soft and sticky, incredibly fragrant] [Video]. YouTube. <https://www.youtube.com/watch?v=9N-WALXaRP8>

⁵ 自制卤味拼盘-百度经验 [Homemade Braised Food Platter - Baidu Jingyan]. (2024, March 5). 自制卤味拼盘 [Homemade braised food platter]. Baidu Jingyan. <https://jingyan.baidu.com/article/54b6b9c05eaf56c583b47a7.html>

⁶ Sauced Braised Food Platter - Mingchu. (n.d.). 酱卤拼盘儿 [Sauced braised food platter]. Mingchu. <https://m.mingchu.co/foodview?id=143040>

версию «Тарелка с тушеным соусом», можно понять, что при заказе мы получим несколько видов соуса, в то время как на самом деле речь идет о блюде из тушеной свинины, которая подается с пикантным соусом. Так, нейросетевой *Google*-перевод вводит пользователя в заблуждение. Стоит отметить, что Яндекс-переводчик предлагает более адекватный вариант: *Блюдо с тушеным мясом в соусе*.

Помимо пищевой тематики 拼盘儿 также используется и в других контекстах, например: 拼盘儿的第1本书.⁷ *Google*-переводчик: *книга 1 блюда*. Очевидно, что нейросеть не обучена отличным от еды тематикам, в частности литературной, что доказывает ее неадекватный перевод как данного слова, так и словосочетания в целом. Человек понимает, что по аналогии с «ассорти/ассортиментом» на обложке книги 拼盘儿 может переводиться как «собрание сочинений / сборник / трилогия» и т.п. Однако нейросеть не способна провести такие параллели. Соответственно, подобные контексты и варианты перевода необходимо для нее «прописать».

Итак, в результате первых этапов алгоритма исследования установлено, что 拼盘儿 является лексической единицей, вызывающей трудности при нейросетевом машинном переводе. Причиной затруднений является, во-первых, ее многозначность в различных дискурсах, а во-вторых, высокая смысловая и стилистическая вариативность при переводе на русский язык. Перечисленные причины приводят к неадекватным переводческим решениям. Одним из способов повышения интерпретативного потенциала нейросетевого переводчика является подробное описание контекстуального окружения, указывающего на нюансы значений при выборе наиболее подходящего варианта перевода. Для реализации данной задачи разработана схема оформления словарной статьи под нейросетевой переводчик (см. Табл. 1).

Основным принципом разработки схемы является ориентирование на логику нейросетевого переводчика, т.е. акцент ставится на тех аспектах, с которыми при интерпретации лексики «машина» не справляется. Для этого выделяется тема (основные виды дискурса, в которых была зафиксирована рассматриваемая лексема), подтема (основные виды коммуникативных ситуаций и жанры текстов), лексическая единица на исходном языке с возможными вариантами ее сокращений и представленность ее переводческих эквивалентов в двуязычных словарях, грамматические характеристики и определение в исходном и переводном языках, сочетаемость лексической единицы (формат схемы позволяет увидеть межъязыковую асимметрию на грамматическом уровне, влияющую на адекватность вариантов перевода), варианты перевода в зависимости от дискурса, варианты перевода в зависимости от категоризации понятия (описание возможных комбинаций контекстуального окружения, влияющего на перевод, в формате схем лексико-грамматической сочетаемости и категоризации понятия).

⁷ 拼盘. (2020, July 29). 拼盘儿的第1本书[Pan'er's First Book] 拼盘儿的第1本书小说_番茄小说网. 拼盘儿的第1本书. <https://fanqienovel.com/page/6851793729183812619>

Таблица 1. Схема оформления словарной статьи лексической единицы 拼盘儿
под нейросетевой переводчик

Язык	Китайский	Русский
Тема	Гастрономический дискурс, коммерческий дискурс, рекламный дискурс, развлекательный дискурс, издательский дискурс	
Подтема	Заказ еды по интернету, оформление меню ресторана, видео-рецепты, шуточные или омерзительные видео-ролики для продвижения личного бренда в социальных сетях, обложки книжных изданий	
Лексическая единица	拼盘儿 - pīnpán	Переводы, зафиксированные в словарях: • ассорти из холодных закусок, холодные закуски (еда) (Zhonga. Китайский словарь и переводчик онлайн – Чжунга, n.d.); • общ. ассорти из холодных закусок (Китайско-русский словарь Мультигран, n.d.)
Возможные сокращения	拼盘 - pīnpán	-
Грамматическая характеристика	名词 (существительное)	В зависимости от приема перевода: • существительное (тарелка, блюдо и др.); • словосочетание (ассорти из...; холодные закуски и др.)
Определение	就是各种新鲜 (熟制, 油炸, 炖, 等等) 食品混合在一起, 做成的饭后甜点品, 可以任意组合各种食品。	• Ассорти – «специально подобранная смесь чего-н., набор» (Толковый словарь Ожегова онлайн, n.d.); • Тарелка и блюдо в заданном значении не зафиксированы в современных словарях
Лексико-грамматическая сочетаемость	名词/形容词 (сущ. или прилаг.) + 拼盘儿	• Прилагательное + ассорти + из/с + существительное(ые); • Прилагательное + тарелка + с + существительное (существительные); • Прилагательное + блюдо + «название»; • Холодные закуски (употребляется самостоятельно)
Варианты перевода в зависимости от дискурса	<ul style="list-style-type: none"> • гастрономический дискурс: 拼盘儿 – тарелка, блюдо, ассорти, холодные закуски, ассорти из холодных закусок; • коммерческий дискурс: 拼盘儿 – ассорти; • рекламный дискурс: 拼盘儿 – ассорти; • развлекательный дискурс: 拼盘儿 – блюдо, тарелка, ассорти; • издательский дискурс: 拼盘儿 – собрание сочинений, сборник сочинений, коллекция, трилогия. 	
Варианты перевода в зависимости от категоризации понятия	<ol style="list-style-type: none"> 1. Категория продуктов питания, от которых образуются благозвучные прилагательные (овощи, фрукты, сыры, мясо и т.п.), + 拼盘儿 переводится как «тарелка», например: 水果拼盘 – фруктовая тарелка, 素拼盘 – вегетарианская тарелка, 米饭拼盘 – рисовая тарелка. 2. Категория продуктов питания, от которых не образуются прилагательные (морепродукты, бекон и т.п.), + 拼盘儿 переводится как «ассорти из», например: 咸菜拼盘 – ассорти из маринованных огурцов, 熏肉拼盘 – ассорти из бекона, 寿司拼盘 – ассорти из суши, 卤味拼盘 – ассорти из тушеной свинины, вареных яиц и тофу, 熟食拼盘 – ассорти из мясных закусок, 海鲜拼盘 – ассорти из морепродуктов. 3. Известное национальное блюдо + 拼盘儿 переводится как «блюдо», например: 寿大吉拼盘 – блюдо «шюдацзи». 4. Несколько категорий продуктов питания, ингредиентов или способы приготовления блюд + 拼盘儿 переводится как «блюдо», например: 熏酱拼盘 – блюдо с соусом с ароматом дыма, 酱卤拼盘儿 – блюдо из тушеной свинины с соусом. 5. Качественная характеристика + 拼盘儿 переводится как «блюдо», например: 美食拼盘 – изысканное блюдо (для гурманов). 	

Многозначность второго вида актуализируется в устойчивом словосочетании, ранее не зафиксированном в словарях. Для устранения подобной проблемы схемы с подробным описанием всех характеристик лексической единицы избыточны, достаточно выявить такие лексико-фразеологические или терминологические словосочетания и оформить в словарную статью.

Рассмотрим алгоритм исследования с целью упрощения схемы оформления на примере терминологического словосочетания 成交方式. Яндекс-переводчик предлагает вариант: «способ совершения транзакции». Google-переводчик выдает: «метод транзакции».

Ни одно, ни другое словосочетание не встречается в интернет-текстах, согласно показателю частотности употребления слов. В деловой русскоязычной среде термин «транзакция» также редко используется ввиду своей многозначности: 1) «банковская операция, состоящая в переводе денежных средств с одного счета на другой» (Финансовый словарь, n.d.); 2) «процесс обмена денежными средствами или продажи активов, товаров или услуг между двумя или более участниками» (Руднева, 2023); 3) «любая оплата банковской картой» (Солдаткин, 2023); 4) «операция по перемещению денежных средств, совершение сделки купли-продажи» (Транзакция, n.d.) и т.п. Очевидно, что под транзакцией одновременно подразумевается и платеж, и процесс сделки, и договор, и даже транспортировка, но в большинстве случаев платеж как банковская операция. Это подтверждается наличием в интернет-текстах таких словосочетаний как «банковская/банкоматная транзакция» или «онлайн-транзакция», «оффлайн-транзакция».

На китайском языке все перечисленные «синонимы» имеют отдельные переводческие эквиваленты. Более того, в русскоязычных словарях и учебниках по коммерческому переводу давно зафиксирован вариант перевода 成交 как «сделка» (Zhonga. Китайский словарь и переводчик онлайн – Чжунга, n.d.; Дашевская, Кондрашевский, 2003, с. 33). Очевидно, что нейросети не обучены данному переводческому соответствию, при том, что сделка – более широкое по семантическому объему понятие, включающее платеж как один из финальных этапов длительного (как часто бывает между российскими и китайскими предпринимателями) процесса взаимодействия двух заинтересованных сторон.

Таким образом, для снятия многозначности в случае устойчивого словосочетания можно предложить упрощенный вариант схемы (см. Табл. 2):

Таблица 2. Схема оформления словарной статьи терминологической единицы 成交方式 под нейросетевой переводчик

Язык	Китайский	Русский
Тема	Коммерческий дискурс, деловой дискурс, банковский дискурс	
Подтема	Деловые переговоры, обсуждение условий сделки, официально-деловые документы	
Лексическая единица	成交方式 - chéngjiāo fāngshì	Переводы, зафиксированные в словарях: <ul style="list-style-type: none"> • 成交 chéngjiāo – совершить торговую сделку; произвести обмен (куплю-продажу) • 方式 fāngshì – 1) система; режим; образ; способ; вид; система; 2) образец; образ; модель; 3) метод, способ (Zhonga. Китайский словарь и переводчик онлайн – Чжунга, n.d.) • 成交方式 – способ совершения сделки (Китайско-русский словарь Мультитран, n.d.)
Определение	成交方式是根据签订出口合同的贸易术语填报的,但是和贸易术语并非完全一致,报关单的成交方式一共有7种,实务中一般能见到CIF、C&F、FOB、EXW这四种。	Сделка – это юридически значимое соглашение, договор или действия между двумя или более сторонами, которые устанавливают их права и обязанности в отношении определенных юридических или финансовых вопросов (Гражданский кодекс Российской Федерации. Статья 153, n.d.).
Лексико-грамматическая сочетаемость	名词 (сущ.) + 成交方式 成交方式+根据...选择	<ul style="list-style-type: none"> • Вид сделки • Вид сделки + FOB/CIF/EXW и др. • Вид сделки определяется в зависимости от...
Варианты перевода в зависимости от дискурса	<ul style="list-style-type: none"> • деловой дискурс: 成交方式 – вид сделки; • коммерческий дискурс: 成交方式 – вид сделки; • банковский дискурс: 成交方式 не используется, вместо данной терминологической единицы встречаются другие словосочетания: 付款条件 – условия платежа и 付款方式 – способ оплаты. 	

Обработанные корпусные данные позволяют выявить несколько рубрик словаря, в который включаются:

1. Слова и термины, не зафиксированные в словарях и требующие лексико- или терминографического описания, например, современные социальные концепты:

а) 三无青年 – понятие, применяемое к молодым людям, у которых нет квартиры, нет машины и нет денег; ориентировочные варианты перевода: «три молодежных “нет”», «три неимения молодости», «молодежь с тремя “нет”» и т.п.;

б) 留吃不留住 – понятие, относящееся к пожилым людям, которые не хотят жить со своими детьми, но готовы принимать заботу в виде семейного ужина в их доме (букв. «оставаться поесть, но не оставаться жить»); ориентировочные варианты перевода: «пожилые люди, стремящиеся сохранить свою автономию от детей», «родители пожилого возраста, отказывающиеся жить с детьми», «вместе есть – да, вместе жить – нет!», «пожилые сингелътоны» и т.п.).

К данной рубрике также могут относиться наименования официально-деловых документов:

с) 商务推销信 – вид документа, содержащего краткую информацию о компании с описанием технических характеристик разрабатываемой ею продукции, с перечислением преимуществ от сотрудничества с данной компанией в формате экономических, производственных и финансовых выгод, а также с представлением имеющихся партнеров и заказчиков. Наименование документа имеет фиксацию в словаре в виде отдельных слов и словосочетаний «коммерческое дело» + «реализовывать» + «письмо» (Zhonga. Китайский словарь и переводчик онлайн – Чжунга, n.d.). В противовес вариантам, предложенным Яндекс-переводчиком «письмо о продвижении бизнеса» и Google-переводчиком «коммерческое письмо о продаже» целесообразно зафиксировать устоявшийся в профессиональной среде вариант – «коммерческое предложение», наиболее полно отражающий суть рассматриваемого документа.

2. Слова и термины, зафиксированные в словарях, но требующие верификации значения и/или выявления нового значения для актуализации лексико- или терминографического описания, например: (阳极)钢爪 – способ изготовления держателя анода в процессе алюминиевого электролиза. В словаре встречаются переводы в виде отдельных слов и словосочетаний «анод» + «сталь» + «коготь, (тех.) лапа, ножка» (Zhonga. Китайский словарь и переводчик онлайн – Чжунга, n.d.). В противовес вариантам, предложенным автоматическими генераторами текста, в том числе через английский язык, – «стальной коготь», «стальная пластина», «захват из анодированной стали» и др., целесообразно зафиксировать устоявшийся в профессиональной среде вариант – «стальной зажим анода», наиболее подходящий в рассматриваемом профессиональном контексте.

3. Слова и термины, зафиксированные в словарях, но требующие контекстологического анализа для определения точных переводческих эквивалентов в зависимости от дискурсивной направленности и жанровой принадлежности, например: 住房 – термин, обозначающий жилое помещение и используемый в официально-деловой документации; варианты перевода в зависимости от дискурса: юридический дискурс: 住房 – жилье, жилое помещение; деловой дискурс: 住房 – жилое здание, жилье; в данном дискурсе также встречается 房间 в значении «гостиничный номер»; бытовой дискурс: 住房 – жилье, квартира, но, как правило, используются общеупотребительные слова: 别墅 – загородный дом; 房子 – дом, квартира; 屋子 – квартира, дом; 房间 – комната.

4. Слова и термины, зафиксированные в словарях, но требующие контекстологического анализа для определения точных переводческих эквивалентов в зависимости от лексико-грамматической сочетаемости;

а) 委屈 – помимо словарных (Zhonga. Китайский словарь и переводчик онлайн – Чжунга, n.d.) существительных «обида, недовольство» и глаголов «обижаться», «маневрировать, изворачиваться, приспособливаться (к кому-л.)», целесообразно добавить такие варианты перевода в форме прилагательного, как «обиженный, униженный», наречия – «обидно, обиженно» и глаголов – «поступать несправедливо», а также «довольствоваться, мириться, терпеть»;

б) 老实 – помимо словарных вариантов «скромный, простой; уживчивый, покладистый»; «честный, правдивый, прямой, искренний; верный», «быть благональным (воспитанным)», «успокоиться, притихнуть» следует добавить «быть благоразумным», «вести себя благоразумно», «соблюдать (правила, дисциплину и т.п.)». Так, в противовес Яндекс-переводчику, предлагающему на китайскую реплику полицейского (警察说：你老实点儿) калькированный вариант перевода «Будьте честны!», в результате дообучения может быть сгенерирован более адекватный вариант: «Ведите себя благоразумно!», т.е. не делайте того, чего не должны делать.

5. Слова и термины, зафиксированные в словарях, но требующие контекстологического анализа для определения точных переводческих эквивалентов в зависимости от концептуально-категориальной таксономии, например:

а) в переводе слова «гостиница» на русский язык иногда стоит давать пояснения в зависимости от имеющихся в китайском языке значений: 酒店 – категория крупных гостиниц, таких как пятизвездочные отели; 旅馆 – категория небольших гостиниц, в которых, как правило, есть только размещение, без завтрака, тренажерного зала и других удобств; 宾馆 – категория любых гостиниц, отелей, постоялых дворов, в которых можно на время остановиться; 饭店 – категория наименования, которая иногда встречается в названиях крупных гостиниц (нейросеть, как правило, интерпретирует данную лексическую единицу только как «ресторан»);

б) «ресторан» также требует пояснений в зависимости от китайских значений: 餐厅 – категория ресторанов, относящихся к высококлассным, с изысканным оформлением и высокими ценами; 餐馆 – категория небольших ресторанов; 饭店 – категория ресторанов, в которых можно только поесть; 饭馆 – категория маленьких ресторанчиков; как правило, используется в устной речи; является наиболее частотным понятием;

в) наименования руководителей также часто вносят путаницу, поэтому стоит уточнять перевод в зависимости от категориальных различий: 主席 – категория лиц, возглавляющих организации/структуры (董事会主席 – Председатель Совета директоров, 大会主席 – Председатель Генеральной ассамблеи); 总裁 – категория руководителей крупных компаний (集团总裁 – президент группы компаний/корпорации/холдинга); 经理 – категория руководителей компаний среднего и малого бизнеса (公司总经理 – генеральный директор компании); 主任 – категория руководителей среднего звена (办公室主任 – начальник офиса, 销售部主任 – начальник отдела продаж, 学校的教导主任 – зам. директора по учебной работе); 老板 – категория лиц, возглавляющих небольшую компанию, однако в разговорной речи может использоваться для обозначения руководителей компании любого масштаба (我们公司的老板 – наш босс); 领导 – категория, номинирующая и президента, и генерального директора, и босса, однако в устной речи обозначающая непосредственно руководителя

(我是一个教师，我的领导是主任。Я – преподаватель, моим руководителем является зав. кафедрой. 我是一个主任，我的领导是校长。Я – зав. кафедрой, моим руководителем является ректор/директор Института).

5. ЗАКЛЮЧЕНИЕ

Исследование показывает, что для выявления неадекватных переводческих решений требуется личный опыт и профессионализм практикующих переводчиков, а также качественно проведенные контекстологический, дефиниционный, лексикографический или терминографический виды анализа. На текущий момент нейросетевые переводчики не обладают подобными исследовательскими компетенциями, фоновыми знаниями и опытом, соответственно, не справляются с идентификацией явления лексической многозначности. Решением данной проблемы является разработка двуязычных словарей на основе ручной обработки параллельных исходных и переводных текстов, предназначенных для реального перевода.

Качественными характеристиками детализации подобных словарей, предназначенных для обучения нейросетевых машинных переводчиков, выступают: идентификация сферы употребления лексической единицы, лингвистические и дефиниционные маркеры, словарная представленность и переводческая вариативность в зависимости от категории понятия, лексико-грамматической сочетаемости и дискурсивной направленности.

Таким образом, при соблюдении установленных принципов разработки китайско-русский словарь многозначной лексики представляет эффективное решение по устранению полисемии и повышению интерпретативности нейросетевых машинных переводчиков. Это вносит дополнительный вклад в развитие теории дискурса и профессиональной коммуникации, способствуя повышению качества машинного перевода, высоко востребованного специалистами разных стран для целей обмена опытом и выстраивания продуктивного взаимодействия между ними.

Конфликт интересов:

Автор заявляет об отсутствии конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Велла Т.М. Переводческие константы интерпретативной теории перевода // Известия Воронежского государственного педагогического университета. 2013. № 2 (261). С. 204-206.
2. Гражданский кодекс Российской Федерации. Статья 153 [Электронный ресурс] // Гардиум. URL: <https://base.garant.ru/10164072/172a6d689833ce3e42dc0a8a7b3cddf9/> (дата обращения: 13.12.2024).
3. Дашевская Г.Я., Кондрашевский А.Ф. Китайский язык для делового общения. 3-е изд., испр. М.: Муравей, 2003.
4. Ершова Ю.Н., Санникова Ю.А. Нейросетевые архитектуры для решения задач лексикографии // Наукосфера. 2024. № 10-2. С. 236-242. doi:10.5281/zenodo.13969346
5. Как работает нейросеть Google Translate [Электронный ресурс] // Cossa. 28 февраля 2018. URL: <https://www.cossa.ru/trends/196086/> (дата обращения: 12.05.2024).

6. Китайско-русский словарь Мультитран [Электронный ресурс]. URL: <https://www.multitrans.com/m.exe?l1=17&l2=2> (дата обращения: 17.12.2024).
7. Ко Ч. Машинный перевод текстов в области традиционной китайской медицины // От машинного перевода к машинному обучению: Сборник научных статей (27–28 октября 2023 г.). Выпуск 3 / под ред. Е.С. Кокановой. Архангельск, 2024. С. 70-73.
8. Коканова Е.С., Пак Н.С. «Рабочие» приемы предредактирования текста английской медицинской инструкции для машинного перевода на русский и белорусский языки // Англистика в третьем тысячелетии: новые подходы и пути развития: Тезисы докладов Международной научной конференции. Минск, 2024. С. 69-70.
9. Компания Яндекс – Технологии – Машинный перевод [Электронный ресурс] // Компания Яндекс. URL: <https://yandex.ru/company/technologies/translation> (дата обращения: 19.03.2024).
10. Машинный перевод. Инновации и влияние на переводческие услуги [Электронный ресурс] // Apriori. Лингвистические услуги. URL: <https://apriori-ltd.ru/apriori-news-blogs-and-articles/tpost/2d59h4s0i1-mashinnii-perevod-innovatsii-i-vliyanie> (дата обращения: 20.06.2024).
11. Мифтахова Р.Г. Методы пополнения корпусных данных в статистическом машинном переводе // Доклады Башкирского университета. 2017. Т. 2. № 1. С. 97-103.
12. Моренцова А.В. Устранение лексической многозначности при машинном переводе: от терминологических словарей к онтологии предметной области // Актуальные научные исследования в современном мире. 2019. № 3-5 (47). С. 69-73.
13. Мукабенов К.И., Ахмадуллина Е.Н. Основные проблемы машинного перевода и пути их решения // Проблемы языка и перевода в трудах молодых ученых. 2023. № 22. С. 176-181.
14. Руднева Р. Транзакция – это [Электронный ресурс] // Банки.ру. 3 июня 2023. URL: <https://www.banki.ru/wikibank/tranzaktsiya/> (дата обращения: 13.03.2024).
15. Солдаткин Д. Транзакция [Электронный ресурс] // Бизнес-секреты. 19 июля 2023. URL: https://secrets.tinkoff.ru/glossarij/tranzaktsiya/?internal_source=copypaste (дата обращения: 13.03.2024).
16. Толковый словарь Ожегова онлайн [Электронный ресурс]. URL: <https://slovarozhegova.ru/word.php?wordid=772> (дата обращения: 13.03.2024).
17. Транзакция [Электронный ресурс] // MyFin. Словарь банковских терминов. 22 января 2020. URL: <https://myfin.by/wiki/term/tranzaktsiya> (дата обращения: 13.03.2024).
18. Убоженко И.В. О когнитивном моделировании интуиции и творчества в переводе: интерпретативно-семиотический подход // Вестник Санкт-Петербургского университета. Серия 9. Филология. Востоковедение. Журналистика. 2016. № 4. С. 122-141. doi:10.21638/11701/spbu09.2016.410
19. Финансовый словарь [Электронный ресурс]. URL: https://dic.academic.ru/dic.nsf/fin_enc/30557 (дата обращения: 13.03.2024).
20. Чистова Е.В. Экокогнитивная модель профессиональной мультимодальной коммуникации (на примере кейса синхронных переводчиков) (Дис.... д-ра н.). Красноярск, 2022.
21. Casas N., Costa-juss`a M.R., Fonollosa J.A.R., Alonso J.A., Fanlo R. Linguistic knowledge-based vocabularies for Neural Machine Translation // Natural Language Engineering. 2018. № 27(4). P. 1-22. doi:10.1017/S1351324920000364
22. Faheem M.A., Wassif K.T., Bayomi H., Abdou Sh.M. Improving neural machine translation for low resource languages through non-parallel corpora: a case study of Egyptian dialect to modern standard Arabic translation // Scientific Reports. 2024. № 14(1). P. 2265. <https://doi.org/10.1038/s41598-023-51090-4>

23. Kokanova E.S., Berendyaev M.V., Kulikov N.Yu. Pre-editing English news texts for machine translation into Russian // *Language Studies and Modern Humanities*. 2022. № 4(1). P. 25-30. <https://www.doi.org/10.33910/2686-830X-2022-4-1-25-30>
24. Kurakin G. How AI originates from biology – and how it returns to it // *The Biochemist*. 2024. № 46(2). P. 3–6. https://doi.org/10.1042/bio_2024_120.
25. Resiandi K., Murakami Y., Nasution A.H. (2023). Neural Network-Based Bilingual Lexicon Induction for Indonesian Ethnic Languages // *Applied Sciences*. 2023. № 13(15). P. 8666. <https://doi.org/10.3390/app13158666>
26. Technologies [Электронный ресурс] // Promt. Глоссарий. URL: <https://www.promt.ru/company/technology/glossary> (дата обращения: 13.03.2024).
27. Wang J. Research on Cultural Translation Based on Neural Network [Электронный ресурс] // *Mathematical Problems in Engineering*. URL: <https://onlinelibrary.wiley.com/doi/10.1155/2022/6330814> (дата обращения: 19.10.2024). <https://doi.org/10.1155/2022/6330814>
28. Zacharias T., Taklikar A., Giryas R. Extending the Vocabulary of Fictional Languages using Neural Networks [Электронный ресурс] // *Workshop Machine Learning for Creativity and Design*. URL: <https://www.researchgate.net/publication/357953278> (дата обращения: 16.12.2024). doi:10.48550/arXiv.2201.07288
29. Zhonga. Китайский словарь и переводчик онлайн – Чжунга [Электронный ресурс]. URL: <https://www.zhonga.ru/> (дата обращения: 13.03.2024).

REFERENCES

1. Casas, N., Costa-juss`a, M.R., Fonolossa, J.A.R., Alonso, J.A., & Fanlo, R. (2018). Linguistic knowledge-based vocabularies for Neural Machine Translation. *Natural Language Engineering*, 27(4), 1-22. doi:10.1017/S1351324920000364
2. Chistova, E.V. (2022). *Ekokognitivnaya model' professional'noj mul'timodal'noj kommunikacii (na primere kejsa sinhronnyh perevodchikov)* [Eco-cognitive model of professional multimodal communication (using the case of simultaneous interpreters as an example)] [Doctoral dissertation, Siberian Federal University]. Krasnoyarsk.
3. Cossa. (2018, February 28). *Kak rabotaet neiroset' Google Translate* [How the Google Translate neural network works]. <https://www.cossa.ru/trends/196086/> (in Russian).
4. Dashevskaya, G.Ya., & Kondrashevskij, A.F. (2003). *Kitajskij yazyk dlya delovogo obshcheniya* [Chinese for Business Communication]. Moscow: Muravey (in Russian).
5. Ershova, Yu.N., & Sannikova, Yu.A. (2024). Nejrosetevye arhitektury dlya resheniya zadach leksikografii [Neural network architectures for solving lexicography tasks]. *Naukosfera*, 10(2), 236-242 (in Russian). doi:10.5281/zenodo.13969346
6. Faheem, M.A., Wassif, K.T., Bayomi, H., & Abdou, Sh.M. (2024). Improving neural machine translation for low resource languages through non-parallel corpora: a case study of Egyptian dialect to modern standard Arabic translation. *Scientific Reports*, 14(1), 2265. <https://doi.org/10.1038/s41598-023-51090-4>
7. Finansovyi slovar' [Financial dictionary]. (n.d.). Retrieved March 13, 2024, from https://dic.academic.ru/dic.nsf/fin_enc/30557 (in Russian).
8. Grazhdanskii kodeks Rossiiskoi Federatsii. Stat'ya 153 [Code of civil laws of the Russian Federation]. (n.d.). *Gardium*. Retrieved December 13, 2024, from <https://base.garant.ru/10164072/172a6d689833ce3e42dc0a8a7b3cddf9/> (in Russian).

9. *Kitaisko-russkii slovar' Mul'titran* [Multitran. Chinese-Russian dictionary]. (n.d.). Retrieved December 17, 2024, from <https://www.multitran.com/m.exe?l1=17&l2=2>
10. Kokanova, E.S., & Pak, N.S. (2024). "Rabochie" priemy predredaktirovaniya teksta anglijskoj medicinskoj instrukcii dlya mashinnogo perevoda na russkij i belorusskij yazyki ["Working" techniques for pre-editing the text of English medical instructions for machine translation into Russian and Belarusian]. *Anglistika v tret'em tysyacheletii: novye podhody i puti razvitiya. tezisy dokladov Mezhdunarodnoj nauchnoj konferencii* [English language studies in the third millennium: new approaches and ways of development: proceedings of the International scientific conference] (pp. 69-70). Minsk.
11. Kokanova, E.S., Berendyaev, M.V., & Kulikov, N.Yu. (2022). Pre-editing English news texts for machine translation into Russian. *Language Studies and Modern Humanities*, 4(1), 25-30. <https://www.doi.org/10.33910/2686-830X-2022-4-1-25-30>
12. *Kompaniya Yandeks — Tekhnologii — Mashinnyi perevod* [Yandex Company – Technologies – Machine Translation] (n.d.). Retrieved March 13, 2024, from <https://yandex.ru/company/technologies/translation> (in Russian).
13. Kuo, Ch. (2024). Mashinnyj perevod tekstov v oblasti tradicionnoj kitajskoj mediciny [Machine translation of texts in the field of traditional Chinese medicine]. In E.S. Kokanova (Ed.), *Ot mashinnogo perevoda k mashinnomu obucheniyu* [From machine translation to machine teaching]: a collection of scientific articles (pp. 70-73). Archangelsk (in Russian).
14. Kurakin, G. (2024). How AI originates from biology – and how it returns to it. *The Biochemist*, 46(2), 3–6. https://doi.org/10.1042/bio_2024_120
15. *Mashinnyi perevod. Innovatsii i vliyanie na perevodcheskie uslugi* [Machine translation. Innovations and their influence on translation services]. (n.d.). Apriori linguistic services. Retrieved June 20, 2024, from <https://apriori-ltd.ru/apriori-news-blogs-and-articles/tpost/2d59h-4s0i1-mashinnii-perevod-innovatsii-i-vliyanie> (in Russian).
16. Miftakhova, R.G. (2017). Metody popolneniya korpusnyh dannyh v statisticheskom mashinnom perevode [New approaches to resolving the problem of corpora data shortage]. *Doklady Bashkirskogo Universiteta*, 2(1), 97-103 (in Russian).
17. Morentsova, A.V. (2019). Ustranenie leksicheskoy mnogoznachnosti pri mashinnom perevode: ot terminologicheskikh slovarej k ontologii predmetnoj oblasti [Elimination of Lexical Ambiguity in Machine Translation: From Terminological Dictionaries to the Domain Ontology]. *Aktual'nye nauchnye issledovaniya v sovremennom mire* [Relevant Research in the Modern World], 3-5(47), 69-73 (in Ukrainian).
18. Mukabenov, K.I., & Akhmadullina, E.N. (2023). Osnovnye problemy mashinnogo perevoda i puti ih resheniya [Main challenges of machine translation: the ways to address]. *Problemy yazyka i perevoda v trudah molodyh uchenykh* [Language and translation issues in the works of young researchers], 22, 176-181 (in Russian).
19. Prompt. Glossary. (n.d.). Technologies. In *Prompt*. Retrieved March 13, 2024, from <https://www.prompt.ru/company/technology/glossary>
20. Resiandi, K., Murakami, Y., & Nasution, A.H. (2023). Neural Network-Based Bilingual Lexicon Induction for Indonesian Ethnic Languages. *Applied Sciences*, 13(15), 8666. <https://doi.org/10.3390/app13158666>
21. Rudneva, R. (2023, June 3). Tranzaktsiya – chto [Transaction is]. *Banki.ru*. <https://www.banki.ru/wikibank/tranzaktsiya/> (in Russian).
22. Soldatkin, D. (2023, July 19). Tranzaktsiya [Transaction]. *Biznes-sekrety* [Business secrets]. https://secrets.tinkoff.ru/glossarij/tranzaktsiya/?internal_source=copypaste (in Russian).

23. Tolkovyi slovar' Ozhegova onlain [Ozhegov's defining dictionary online]. (n.d.). Retrieved March 13, 2024, from <https://slovarozhegova.ru/word.php?wordid=772> (in Russian).
24. Tranzaktsiya [Transaction]. (2020, January 22). *MyFin. Slovar' bankovskikh terminov* [MyFin. Dictionary of banking terms]. Retrieved March 13, 2024, from <https://myfin.by/wiki/term/tranzakciya> (in Russian).
25. Ubozhenko, I.V. (2016). O kognitivnom modelirovanii intuicii i tvorchestva v perevode: interpretativno-semioticheskij podhod [On cognitive modeling of intuition and creativity in translation: interpretive and semiotic approaches]. *Vestnik Sankt-Peterburgskogo universiteta. Seriya 9. Filologiya. Vostokovedenie. Zhurnalistika* [Vestnik of Saint Petersburg University. Language and Literature], 4, 122-141 (in Russian). doi:10.21638/11701/spbu09.2016.410
26. Vella, T.M. (2013). Perevodcheskie konstanty interpretativnoj teorii perevoda [Translation constants of the interpretative translation theory]. *Izvestiya Voronezhskogo gosudarstvennogo pedagogicheskogo universiteta* [Izvestia. Voronezh State Pedagogical University], 2(261), 204-206 (in Russian).
27. Wang, J. (2022). Research on Cultural Translation Based on Neural Network. In N. Jan (Ed.), *Mathematical Problems in Engineering*. <https://doi.org/10.1155/2022/6330814>
28. Zacharias, T., Taklikar, A., & Giryes, R. (2022). Extending the Vocabulary of Fictional Languages using Neural Networks. Workshop Machine Learning for Creativity and Design. doi:10.48550/arXiv.2201.07288.
29. Zhonga. Kitaiskii slovar' i perevodchik onlain – Chzhunga [Zhonga. Online Chinese Dictionary]. (n.d.). Retrieved March 13, 2024, from <https://www.zhonga.ru/>.

Информация об авторе:

Елена Викторовна Чистова – доктор филологических наук, доцент, профессор кафедры теории, практики и дидактики перевода АНО ВО «Университет мировых цивилизаций имени В.В. Жириновского» (Москва, Россия). ORCID: 0000-0002-4016-7935.

Получено: 11 ноября 2024 г.

Принято к публикации: 10 января 2025 г.

PRINCIPLES OF DEVELOPING A CHINESE-RUSSIAN POLYSEMANTIC DICTIONARY AS A MEANS OF IMPROVING INTERPRETABILITY OF NEURAL MACHINE TRANSLATORS

Elena V. Chistova

Zhirinovskiy University of World Civilizations

kovelena82@mail.ru

Abstract. This research addresses the challenge of polysemy in neural machine translation (NMT), particularly for the Chinese-Russian language pair, known for its significant interlingual and intercultural asymmetry. Despite considerable advancements in NMT, the accurate translation of polysemous words remains a key obstacle to achieving high-quality automated text generation, often leading to misinterpretations and hindering effective communication. Currently, methodologies for developing specialized dictionaries that can effectively address this issue for NMT systems are lacking. This article aims to define the qualitative characteristics for detailed polysemantic dictionaries designed to enhance the interpretability of NMT, specifically for Chinese-Russian translation. The study employs eco-cognitive modeling of professional translator communication to investigate human-machine interaction in handling lexical ambiguity, focusing on the cognitive processes involved in disambiguation. Parallel Chinese-Russian texts serve as the material, subjected to manual processing to identify polysemous units challenging for NMT. The article proposes a theoretical framework for bilingual dictionary compilation based on this manual analysis, outlining principles for structuring dictionary entries to capture subtleties of lexical usage. The developed algorithm details the manual processing of parallel texts and the design of dictionary entry schemes tailored for NMT. The research identifies key qualitative characteristics for detailed Chinese-Russian parallel training corpora. These include linguistic and definitional parameters, comprehensive dictionary representation, and translation variability informed by lexico-grammatical compatibility, discourse-genre affiliation, and conceptual-categorical taxonomy. This study contributes to translation theory by offering a practical approach to enhance NMT interpretability through targeted dictionary development. The findings are relevant for improving machine translation quality, particularly for complex language pairs, ultimately facilitating more effective cross-lingual communication and knowledge exchange in different spheres, including business and academic research.

Keywords: Natural language processing, machine translation, neural translation, polysemy, the Chinese language, lexicology, terminology.

How to cite this article: Chistova, E.V. (2025). Principles of Developing a Chinese-Russian Polysemantic Dictionary as a Means of Improving Interpretability of Neural Machine Translators. *Professional Discourse & Communication*, 7(1), 89–107 (in Russian). <https://doi.org/10.24833/2687-0126-2025-7-1-89-107>

About the author:

Elena V. Chistova, Dr. Sci. (Philology), is a Professor in the Department of Theory, Practice, and Didactics of Translation at Zhirinovskiy University of World Civilizations (Moscow, Russia). ORCID: 0000-0002-4016-7935.

Received: November 11, 2024.

Accepted: January 10, 2025.