**Применение теории Московской семантической школы в русско-китайском машинном переводе и параллельном корпусе**

*Чэнь Лань*

*Аспирант, Российского университета дружбы народов им. Патриса Лумумбы, Москва Россия*

*Email:1014129604@qq.com*

Машинный перевод — это новая дисциплина, которая использует электронные компьютеры для перевода с одного языка (исходного языка) на другой язык (целевой язык). Это также междисциплинарный предмет, объединяющий лингвистику, информатику, математику и другие дисциплины. Будучи новой технологией, машинный перевод относится к области компьютерной лингвистики, а в математике машинный перевод относится к области исследований математической логики и формальных методов. Естественный язык гораздо сложнее, чем мы представляем.

С 1950-х годов машинный перевод пережил несколько взлетов и падений, но у него все еще много недостатков. Машинный перевод часто не передает смысл или даже не имеет логики выражения, что не является полностью удовлетворительным. Точность машинного перевода не улучшилась из-за обновлений программного обеспечения. Мы знаем, что естественный язык — чрезвычайно сложное явление. Если мы хотим, чтобы машины переводили естественный язык, нам нужны не только точные теории, связанные с компьютерами, но и обширная лингвистическая теория, особенно семантика, для ее поддержки.

Будучи важной школой русской лингвистики, Московская семантическая школа (МСШ) оказывает глубокое влияние в области обработки естественного языка благодаря своей уникальной «Теории смысла ⇔ текста» (ТСТ) и системе семантического метаязыка. В данной статье в качестве объекта исследования рассматривается русско-китайский машинный перевод и параллельное построение корпуса, а также изучается практическая ценность теоретической базы Московской семантической школы в межъязыковом семантическом анализе, оптимизации модели перевода и разработке корпусных ресурсов. Исследование объединяет методы корпусной лингвистики и использует эмпирический анализ для проверки эффекта применения теории Московской школы семантики при решении таких проблем, как различия в русско-китайских языковых структурах, конфликты семантической валентности и культурно нагруженные преобразования слов, предоставляя новые идеи для решения проблем русско-китайского машинного перевода.

Во-первых, в данной статье систематически разбирается основная теоретическая система Московской школы семантики, включая глубокое структурное представление семантики, теорию лексических функций и модель семантической валентности. По сравнению с ограничениями традиционного машинного перевода, который опирается на поверхностное синтаксическое выравнивание, МСШ подчеркивает динамическое отображение семантических ролей и формальное описание правил комбинирования лексики. Он особенно подходит для языковых пар, таких как русский и китайский, со сложной морфологией и высокой свободой порядка слов. На примере категории вида русского глагола и структуры дополнения китайского языка исследование продемонстрировало, как правила семантической трансформации, основанные на лексических функциях, могут эффективно решать ошибки перевода, вызванные асимметрией времени, вида и залога.

Во-вторых, мы занимаемся аннотацией и оптимизацией параллельных корпусов. Традиционные двуязычные корпуса в основном опираются на разметку частей речи и синтаксический анализ дерева, в то время как теория МСШ обеспечивает детальную парадигму семантической разметки для корпусной лингвистики. Благодаря созданию русско-китайского параллельного корпуса, включающего уровни семантической валентности, тематические роли и метки лексических функций, русско-китайский параллельный корпус в Национальном корпусе русского языка переходит от «формального соответствия» к «семантической эквивалентности». Например, в соответствии между русскими герундиальными оборотами (такими как «проведение исследования») и китайским «проводить исследование» семантическая метаязыковая аннотация может точно идентифицировать неявный субъект действия и дополнение, уменьшая логическую неоднозначность перевода.

В ответ на сложные морфологические изменения в русском языке (например, падежную систему) и появление китайских служебных слов модель генерирует переводы, которые соответствуют привычкам целевого языка посредством семантических ролей. Например, в русском предложении "Он дал книгу другу" (He gave a book to his friend) необходимо скорректировать порядок слов в дательном падеже существительного "другу" для достижения семантической эквивалентности в китайском языке. Модель автоматически выбирает структуру "给+косвенное дополнение" вместо буквального перевода "向朋友给书" [4].

Это исследование инновационно объединяет теорию Московской школы семантики с методами корпусной лингвистики, преодолевая техническое узкое место глубокого семантического структурного моделирования в русско-китайском машинном переводе. В будущей работе будет изучена интеграция мультимодальной семантической аннотации и предварительно обученных языковых моделей для дальнейшего повышения межъязыковой семантической рассудочной способности машинного перевода, а также будут предоставлены теоретические рекомендации и предложения по созданию современного двуязычного русско-китайского параллельного корпуса, в надежде на определенное улучшение качества русско-китайского машинного перевода.

**Литература**

[1] Городецкий Б.Ю. Компьютерная лингвистика: моделирование языкового общения. М ,1989.

[2] Жолковский А.К., Мельчук И.А. К построению действующей модели языка<<Смысл<=>текст>>//Машинный перевод иприкладная линвистика. Вып. 11 [M], 1969.

[3] Сунь Шуан. Об истории и современном состоянии российского и китайского машинного перевода // «Голоса молодых ученых». Вып. 22 [M], 2009.

[4] Национальный корпус русского языка: [Russian National Corpus](https://ruscorpora.ru/en) https://ruscorpora.ru/en