

Д. Возняк

НАУЧНО-ТЕХНИЧЕСКИЙ ТЕКСТ КАК ОБЪЕКТ АНАЛИЗА В СИСТЕМАХ АВТОМАТИЧЕСКОГО КИТАЙСКО-РУССКОГО ПЕРЕВОДА

В ходе создания систем автоматического перевода (АП) главной задачей является формализация текста, и, в конечном итоге, получение работающих алгоритмов, которые позволяют получать текст перевода. Эффективное осуществление формализации отмечается на морфологическом, лексическом и синтаксическом уровнях. Трудности для формализации представляет семантический уровень, что обусловлено наличием контекстуальных значений, а также связи со смыслом конкретного высказывания в определенной коммуникативной ситуации. Помимо этого, результат перевода предполагает передачу смысла текста в целом.

Текст представляет собой объект перевода, который связан с действительностью языком коммуникации, мыслью и мышлением автора. Проявление этой связи может отмечаться на предметном, логическом и лингвистическом уровнях. В первом случае это проявление представляет собой экстралингвистическую информацию, во втором – средства развертывания и организации авторской мысли, в третьем – конкретные языковые средства, используемые автором. Не менее важным элементом является адресат, так как он получает и воспринимает текст. Следовательно, именно на него осуществляется ориентация переводчика в процессе выбора того или иного варианта перевода.

Нами был проанализирован перевод с китайского языка на русский научно-технических текстов. Анализ текста в процессе перевода предполагает преобразования и трансформации средств выражения языка. Лингвистический анализ текста, который занимает центральное место, не является достаточным для осуществления качественного перевода. Помимо него проявляется необходимость в логическом и предметном анализе текста.

В 2016 г. Google представила систему нейронного машинного перевода, которая использует искусственную нейронную сеть для улучшения качества перевода. Нейронная сеть переводит предложение целиком, учитывая контекст. Система не запоминает сотни вариантов перевода фраз – она оперирует семантикой текста. При переводе предложение разбивается на словарные сегменты. Затем с помощью специальных декодеров система определяет «вес» каждого сегмента в тексте. Далее вычисляются максимально вероятные значения и перевод сегментов. Последний этап – соединить переведённые сегменты с учетом грамматики.

На данный момент перевод на основе правил и нейронный машинный перевод являются лидерами среди методов машинного перевода, однако потенциал систем перевода на основе правил рано или поздно будет исчерпан, в то время как качество перевода систем, основанных на нейронных сетях стремительно улучшается.