

ЛИНГВИСТИКА ТЕКСТА И МАШИННЫЙ ПЕРЕВОД

Бул макалада тексттерди автоматикалык турдо иштетуу формаларын машиналык котормону колдонуулуп концептуалдуу суроолор каралган. Илимий техникалык адабияттарды мамлекеттик тилге которуудагы эн негизги инструмент болуп машиналык котормо болуп эсептелет.

В данной статье рассматриваются концептуальные вопросы применения машинного перевода как формы автоматической обработки текста. Машинный перевод может являться одним из главных инструментов перевода научно-технической литературы на государственный язык.

In this article considered conceptual items of using machinery translation as form of automatic adaptation of the text. Machine translation may appear one of the main instrument of scientific – technical literature for state language.

К области компьютерной лингвистики в определенной степени могут быть отнесены работы в области создания гипертекстовых систем, рассматриваемых как особый способ организации текста и даже как принципиально новый вид текста, противопоставленный по многим своим свойствам обычному тексту. К компетенции компьютерной лингвистики относится и машинный перевод. Машинный перевод является высшей формой автоматической переработки текста. Особая значимость машинного перевода проявляется и в том, что он предполагает применение новых информационных технологий в комплексе с прикладной математикой и лингвистикой.

Если машинная атрибуция, аннотирование и реферирование, имеющие целью извлечь из текста только общий смысловой текст, ограничиваются простыми приемами распознавания смысла у отдельных элементов текста, то машинный перевод (МП), ориентированный на полную смысловую переработку текста, использует более сложные формы распознавания, охватывающие в идеале все единицы текста. Поэтому, с одной стороны, в алгоритмах МП оказываются представленными все приемы отбора, сортировки, сегментации, сравнения и отождествления ЛЕ, использующиеся в низших формах переработки текста. С другой стороны, успешное создание действующего МП в гораздо большей степени, чем построение простых форм переработки, зависит от выбора правильной лингвистической теории. Разработка систем МП может осуществляться либо на основе теории порождающих грамматик, либо опираясь на идеи лингвистики текста. Порождающая грамматика обещает перспективу полного формального распознавания и порождения любого правильного текста при условии, что будет определен словарь исходных единиц и глубинных структур, а также задан алгоритм порождения. Правильное распознавание и переработка на компьютерах актуальных значений ЛЕ входного текста даст возможность выдавать такой машинный результат, который не только воспринимался бы потребителем, но и сохранял основную смысловую информацию входа. Фундаментом всякой системы МП должен быть лексический перевод, опирающийся на автоматические словари, потом следует строить семантический перевод, который опирается на алгоритмы устранения многозначности, позволяющие автоматически распознавать актуальное значение ЛЕ в перерабатываемом тексте.

Термин «квантитативная лингвистика» характеризует междисциплинарное направление в прикладных исследованиях, в котором в качестве основного инструмента изучения языка и речи используются количественные или статистические методы анализа. Иногда квантитативная (или количественная) лингвистика противопоставляется комбинаторной

лингвистике. В последней доминирующую роль занимает «неколичественный» математический аппарат – теория множеств, математическая логика, теория алгоритмов и т.д. С теоретической точки зрения, использование статистических методов в языкознании позволяет дополнить структурную модель языка вероятностным компонентом, т.е. создать структурно-вероятностную модель, обладающую значительным объяснительным потенциалом. Эту сторону использования количественных методов следует считать приложением статистики в языкознании. Иными словами, задача построения структурно-вероятностной модели функционирования языка относится к теоретическим проблемам лингвистики и входит в компетенцию теории языка.

Наука о переводе (англ. translation studies; нем. Übersetzungswissenschaft или Translationswissenschaft) включает несколько направлений, среди которых выделяются теория перевода, анализ перевода, методика обучения переводу. Особое место занимает машинный перевод – научная и одновременно технологическая дисциплина, связанная и с наукой о переводе, и с компьютерной лингвистикой. Важнейшая лингвистическая проблема науки о переводе – адекватная передача так называемых «скрытых категорий» – категории неотчуждаемой принадлежности, определенности-неопределенности и пр. Будучи в одних языках грамматическими, в других языках эти аспекты значения могут выражаться не регулярно, а лишь при определенных условиях, что существенно осложняет работу переводчика.

Начало работ по машинному, или автоматическому переводу относится к концу 1940-х – началу 1950-х годов. Идея машинного перевода во многом, хотя и не во всем обязана своим происхождением практическим нуждам. В указанный период существенно возрос поток научно-технической информации, а отслеживание этого потока в условиях напряженного и дорогостоящего научно-технического соперничества стало весьма насущной задачей. В полной мере решить ее средствами традиционного, «человеческого» перевода не представлялось возможным. В 1949 г

американский специалист по дешифровке Уоррен Уивер составил меморандум, в котором теоретически обосновал принципиальную возможность создания систем машинного перевода. Он исходил из того, что структурное сходство между языками может быть формально описано, а это является необходимым условием разработки алгоритмов для ЭВМ. Первоначально предполагалось, что системы машинного перевода смогут осуществлять перевод научно-технических текстов любых типов, что позволит отказаться от дорогостоящего труда обычных переводчиков. Оказалось, однако, что создание систем машинного перевода требует существенной и нетривиальной информации о функционировании естественного языка, которая отсутствовала в традиционных грамматических описаниях. Кроме того, компьютерное моделирование способности человека к переводу с одного языка на другой требовало учета не только лингвистических, но и психологических, социальных и других факторов. Первые попытки разработки программ автоматического перевода оказались неудовлетворительными: результаты перевода ЭВМ требовали серьезного редактирования, а стоимость машинного перевода зачастую была выше оплаты труда обычного переводчика. В настоящее время развитие систем машинного перевода идет по нескольким направлениям. Во-первых, разрабатываются человеко-машинные системы, предусматривающие прямое участие человека в процессе перевода на различных этапах. Во-вторых, проблемная область системы ограничивается конкретным подязыком, например текстами химического машиностроения или ядерной физики. В-третьих, системы машинного перевода рассматриваются как вид систем искусственного интеллекта, что позволяет использовать компьютерные технологии, разработанные в компьютерном моделировании мышления человека.

Список литературы

1. Малаховский Л.В. Статистика речи и автоматический анализ текста /Институт языкознания НАН КР. – Бишкек: Илим, 1990.
2. Баранов А.Н. Автоматизация лингвистических исследований: корпус текстов как лингвистическая проблема //Русистика сегодня. – 1998. - № 12.
3. Хомский Н. Синтаксические структуры //Новое в лингвистике, вып. 2. - М., 1962.
4. Гладкий А.В., Мельчук И.А. Элементы математической лингвистики. - М., 1969.
5. Маркус С. Теоретико-множественные модели языков /Пер. с англ. - М., 1970.
6. Гладкий А.В. Формальные грамматики и языки.- М., 1973.