

К ПРОБЛЕМЕ ФОРМАЛИЗАЦИИ ТРАНСПОРТНЫХ ДОКУМЕНТОВ

Формализация текстов на естественном языке является одной из основных актуальных задач, которыми занимается прикладная лингвистика сегодня. Формализация узко профильных текстов вызывает особый интерес с точки зрения возможности дальнейшего многофункционального применения этих результатов для построения систем искусственного интеллекта.

В исследовании, посвящённом созданию системы автоматического создания транспортных документов, рассматривались тексты трёх видов транспортных договоров в области автомобильного сообщения: договоров на транспортно-экспедиционные услуги, договоров на перевозку грузов и договоров на оказание логистических услуг. Указанные договоры отличаются определённой шаблонностью и сходством в семантико-синтаксической организации.

Одним из этапов разработки системы создания автоматических транспортных документов, в частности транспортных договоров для автомобильного сообщения (на русском, немецком и английском языках), являлось определение семантико-синтаксической организации исследуемых текстов. В рамках этого этапа были построены алфавитно-частотные словари данных типов текстов, определено их логико-семантическое содержание, где автоматически были выделены главные опорные слова (ГОС) и второстепенные опорные слова (ВОС) для каждого раздела всех анализируемых текстов, на основе списков опорных слов обозначены микротемы и построены логико-семантические формулы для каждого раздела текста документа. Следующим шагом стало выделение семантических подклассов слов всех самостоятельных частей речи. Всем служебным частям речи изначально присваивался однозначный код. Семантическая классификация составлялась в рамках узких тем «Логистика», «Транспортно-экспедиционная деятельность» и «Перевозка грузов».

Структурированные лингвистические данные использовались для построения непосредственно семантико-синтаксических формул исследуемых текстов. Для реализации этого этапа исследования применялся формальный семантико-синтаксический язык СЕМСИНТ, который был разработан для адекватного отображения структуры и семантической индивидуальности моделируемого текста. Язык СЕМСИНТ опирается на семантико-синтаксические особенности абзаца в качестве основного семантического компонента текста и существенно отличается от многих семантических (информационных) языков тем, что позволяет

передать одну и ту же информацию различными способами (различными словами и предложениями).

Семантико-синтаксический язык включает следующие составляющие (Зубов А.В.):

1) алфавит языка СЕМСИНТ составляют знаки основного набора клавиатуры современных персональных компьютеров. Комбинации букв, цифр и специальных знаков алфавита позволяют описывать и все остальные составляющие русского, немецкого и английского языков;

2) систему средств для записи семантических отношений между словами предложения;

3) систему средств для записи синтаксических отношений между словами предложения;

4) систему элементов для фиксации логико-семантических отношений между предложениями отдельного абзаца и абзацами текста;

5) систему средств для записи ГОС и ВОС, задающих содержание будущего текста.

Для описания смысла предложения в языке СЕМСИНТ используется падежно-ролевой подход, предложенный Ч. Филлмором [Филлмор Ч., 2004], в соответствии с которым семантическая формула предложения должна отражать семантико-синтаксические роли конкретных участников реального события, отображаемого глаголом предложения.

Семантическую структуру предложения Ч. Филлмор называет глубинной структурой (ГС), а в качестве основы для ее репрезентации он использует предикативно-аргументный принцип символической логики (Филлмор Ч.).

Для выражения принадлежности слова к той или иной семантической группе нами были использованы специальные коды семантических функций, описанные в исследовании (Кунцевич С.Е.): AGA — активный одушевлённый инициатор события, OAN — одушевлённый объект события, DDA — одушевлённый адресат некоторого события, MNN — неодушевлённое средство реализации некоторого события типа «предмет», IMN — неодушевлённое орудие реализации некоторого события типа «предмет», TIM — время протекания события, FRQ — частота повторений некоторого события, LOC — место реализации некоторого события, RST — результат реализации некоторого события, PPS — цель реализации некоторого события, CND — условие для реализации некоторого, MOD — способ реализации события и т.д.

Семантические функции в языке СЕМСИНТ выражают отношения не между отдельными словами предложения, а между группами слов, называемыми аргументными группами (АГ), состоящими из цепочки кодов соответствующих ТГ, соединенных знаком (*), и заключаются в угловые скобки (< >), перед которыми ставится код соответствующей СФ.

Для обозначения в семантико-синтаксической структуре предложения места глагола-сказуемого используется код Rn, где n — цифра, отражающая семантическую валентность глагола. Семантическая валентность глагола

определяется количеством АГ, которые данный глагол объединяет в каждом конкретном случае.

В семантико-синтаксической формуле каждого абзаца текста главные и второстепенные опорные слова указываются вслед за кодом соответствующего семантического подкласса имени существительного, глагола и т.д. через знак (/) (Зубов А.В.).

Например, следующее предложение из текста транспортного договора будет иметь такую семантико-синтаксическую формулу:

Для оказания транспортно-экспедиционных услуг клиентом выдается заполненное и подписанное им поручение экспедитору.

PPS<G06*N10*A13*N10>+OAN<N03`/BOC3>+R4<V12>+ONN<X14*C01*X19*P01*N23>+DDA<N03``/BOC12>.

Как видно из описания и приведенного примера, язык СЕМСИНТ позволяет записывать содержание текста с точным указанием АГ, а также синтаксических отношений, зафиксированных автором между элементами ситуации, описываемой в тексте.

Этап определения семантико-синтаксической организации текстов транспортных договоров является ключевым в процессе формализации подобных текстов и предопределяет качество дальнейшей системы автоматического создания транспортных документов. Эти результаты позволят решать прикладные лингвистические задачи, такие как, автоматическое создание транспортного договора, подготовка эквивалентов аналогичного текста договора на других языках.