
Intelligent NL Processing

ФОРМАЛЬНОЕ ОПРЕДЕЛЕНИЕ СИТУАЦИИ ДЛЯ СЕМАНТИЧЕСКОГО АНАЛИЗА ТЕКСТОВ ЕСТЕСТВЕННОГО ЯЗЫКА

Леонид Святогор, Виктор Гладун

Аннотация: Предложен новый метод семантического анализа текстов естественного языка, основанный на совместном использовании знаний о Мире, знаний в Предметной области и знаний о ситуации. Введено понятие «конструкт ситуации», на основе которого построен формат Лексической структуры ситуации. Обсуждаются подходы к ситуационному мышлению и определению смысла. Дан пример семантического анализа и выделения смысла дискурса.

Ключевые слова: текст естественного языка, семантическая структура, ситуация, онтология, смысл текста

ACM Classification Keywords: 1.2.7 Natural Language Processing – Text analysis

Введение

Задача автоматического анализа и компьютерного понимания семантически-насыщенных текстов естественного языка (деловой прозы, литературных произведений, и т.д.) является одной из актуальных задач Искусственного интеллекта. Здесь сливаются два потока изучения: лингвистический анализ (*Text processing*) и представление знаний (*Knowledge representation*) [Мельчук, 1974], [Демьянков, 1985], [Рубашкин, 1989], [Шенк, 1980], [Витгенштейн, 2009], [Звенигородский, 2010], [Гаврилова, 1997], [Рыбина, 2010]. В этом направлении сделана попытка представить внутреннее содержание текста, или смысл, как некоторую формальную конструкцию в языковой Картине Мира. В частности, в работах [Гладун и др, 2008], [Гладун, Святогор, 2011] предложена модель онтологической Картины Мира в виде концептуального графа, а текст отображается на онтологию, возбуждая в ней множество подграфов, или смысловых траекторий. Множество смысловых траекторий предложено считать *онтологическим смыслом текста*. Процедура реализуется при помощи компьютерных средств.

Анализ текста человеком опирается, прежде всего, на понимание входящих в него *предложений*, каждое из которых выражает законченную мысль автора. Он оформляет мысль выразительными средствами живого языка в соответствии с личной коммуникационной задачей. Поэтому предложение субъективно по его рождению, но объективно по интерпретации в общей (социальной) среде знаний; оно несёт семантическую и прагматическую нагрузку. Однако смысловое усвоение информации потребителем – понимание предложения – есть творческое усилие личности, направленное на согласование сказанного с его личной базой знаний, и успех здесь зависит от интеллекта и опыта субъекта.

В то же время при чтении текста (дискурса) ограничиться пониманием отдельных предложений мы не можем – необходимо связать весь текст в единое целое, создать «образ текста», сформировать его смысл, дать общую оценку в соответствии с персональной базой знаний.

Движение от текста к смыслу происходит через образование ситуаций. В семиотических моделях репрезентации знаний «*универсальной единицей репрезентации знаний и смысла является ситуация*» [Штерн, 1998]. В процессе чтения мы выдвигаем гипотезу о некоторой текущей ситуации, затем следим за её развитием, чтобы в итоге синтезировать ту ситуацию, из которой вытекает смысл дискурса.

В данной работе под *ситуацией* мы будем понимать определённую, зафиксированную в языке *реализацию* событийной Картины Мира. Это семантическая *макроструктура*, которой в тексте соответствует грамматическая конструкция, отвечающая на вопросы: «Что случилось?», «Что происходит?».

Ситуационный анализ текстов не должен противоречить концептуальному анализу, а должен его дополнять. За счёт этого можно добиться более глубокой интерпретации и понимания текстов ЕЯ.

Цель данной работы в том, чтобы *формализовать* понятие «ситуация» и включить его в алгоритмическую обработку дискурсов.

Об онтологии ассоциаций

Процедура концептуального анализа ЕЯ текстов основана на представлении Картины Мира в виде семантической сети – онтологии. Предложенная в работах [Гладун и др, 2008], [Гладун, Святогор, 2011]. Онтология ассоциаций – как познавательная структура Знаний – представлена в виде иерархической трёхуровневой сети (**ИО*3**) и задана графом концептуальных зависимостей.

Понятийная Модель знаний основана на учении академика В.И. Вернадского (1863–1945) о ноосфере. Иерархия знаний распространяется по сети от верхней категории «Материя» через «Косное вещество», «Энергию» и «Живое вещество» на уровни с концептами меньшей общности (например, такие как «Разум», «Человек», «Общество», «Искусство», «Музыка», «Музыкальные инструменты» и т.д.), и далее – вплоть до нижнего уровня понятий конкретных предметных областей (ПрО). В онтологии наряду с родовыми связями широко используются ассоциативные отношения.

Онтология **ИО*3** предназначена для того, чтобы исследовать семантику текста в контексте общих знаний о Мире. Конкретно такое исследование представляет собой алгоритмическую процедуру, состоящую из трёх этапов: **(а)** выделение в тексте *значимых слов*; **(б)** построение для каждого значимого слова такой цепочки концептов в онтологической сети, которая ведёт от заданного слова (концепта) к вершине онтологии (*смысловой траектории*); **(в)** образование из множества цепочек *онтологического смысла* связанного текста.

Под онтологическим (концептуальным) смыслом текста понимается полная совокупность смысловых траекторий, полученных проецированием значимых слов текста на граф онтологии.

Затем лингвистический процессор преобразует множество траекторий в *литературный смысл* документа. Литературный смысл представляет собой одно или несколько предложений естественного языка – таких, что они в предельно сжатой форме передают замысел автора дискурса, освобождённый от литературно-художественного оформления.

Эффективность алгоритма концептуально-литературного анализа была продемонстрирована на примере художественного отрывка, взятого из рассказа «Триста спартанцев» [Гладун, Святогор, 2011]. В эксперименте проявились и недостатки данного алгоритма. Они заключаются в следующем.

1. Смысловые траектории не поддерживают *семантическое единство слов* внутри предложения, в результате чего семантика отдельного предложения сводится к семантике его членов. Связность текста достигается только на родовых и ассоциативных связях между значимыми словами текста.

2. Концептуальная онтология представляет *статическую* Картину Мира, между тем как осмысление текста есть последовательный *динамический процесс*, в ходе которого постепенно, во времени, выкристаллизовывается глубинный смысл.

Эти недостатки могут быть преодолены на пути создания **нового класса онтологий** – *концептуально-ситуационных онтологий*.

Ситуационное мышление (краткий обзор)

«В течение жизни сознание одного субъекта работает с огромным количеством ситуаций, которые повторяются с некоторыми изменениями и характеризуются постоянством образов и их поведением. Это постоянство сознанием определяется как смысл» [Звенигородский, 2010].

В данном высказывании содержится идея, которая кажется нам весьма конструктивной. Человек настолько хорошо ориентируется в реальной среде, насколько хорошо наблюдение совпадает с его личным опытом; опыт же человека – конечен и сформирован набором стандартных ситуаций. Опыт служит основой умозаключений.

Термин «ситуация» настолько широко применяется в человеческом общении, что это неизбежно приводит к его расплывчатому толкованию. Выяснением того, **что есть ситуация**, занимаются лингвисты, психологи, нейрофизиологи [Штерн, 1998], [Звенигородский, 2010], [Кисленко, 2008]. В последнее время понятия ситуационного мышления и ситуационного моделирования становятся ключевыми в проблеме понимания естественного языка и окружающей среды. Причина очевидна: человек формирует представление о Мире в значительной степени при помощи обобщённого описания его *состояний*, то есть, по сути, – ситуаций.

Для передачи *семантического* содержания, изначально присущего термину «ситуация», необходимо обратиться к категориям семантики.

Семантическая классификация высказываний восходит к Аристотелю (IV в. до н.э.), который выделил 10 категорий семантики: *субстанция, количество, качество, отношение* и др. В средневековой логике активно разрабатывается учение об «идеях вещей» [Арно, Николь, 1991], открывающее пути к определению понятия «знак» (Г. Фреге, 1848–1925).

Современные исследователи, исследующие язык на концептуальном уровне, вводят онтологические категории типа «образной схемы» (Дж. Лакофф) или «конфигурационной структуры» (Л. Талми), посредством которых естественный язык осуществляет концептуальное структурирование обобщённых представлений о действительности. Р. Шенк предложил описывать *смысловую структуру ситуации* при помощи «типовых базовых действий», количество которых ограничено, – например: «перемещать объект», «передавать информацию», «сфокусироваться на актуальном стимуле» и др. [Шенк, 1980].

Связь ситуации с её пониманием на смысловой основе отмечается многими исследователями. *«Сторонам общения должны быть ясны такие индикаторы речи как «кто», «кому», «что», «зачем», «как» говорит и какой вывод вытекает из сказанного»* [Звягинцев, 2009]. Эта идея в общем виде прослеживается в области стратификации Знаний. Например, Т.А. Гаврилова вводит в рассмотрение восемь страт знаний, в числе которых присутствуют: **«кто-знания»** (субъектный анализ), **«что-знания»** (объектный анализ), **«зачем-знания»** (целевой анализ), **«где-знания»** (локальный анализ), **«когда-знания»** (временной анализ) и другие [Гаврилова, Воинов, 1997].

Наиболее полный перечень глубинных падежей, использующихся для описания типов актантов, мы встречаем у Г.В. Рыбиной. Семантический объект характеризуется с разных сторон его изучения, причём ролевые функции актантов укладываются в формат *«Быть признаком действия»*. Полный список глубинных падежей содержит 16 признаков действия: *агент, объект, адресат, тема, инструмент, место, начальная точка, конечная точка, цель, условие, время, способ, срок, количество, результат, контрагент* [Рыбина, 2010]. Заметим, что этот ролевой список мы имеем право дополнить новыми семантическими координатами, которые характеризуют другие стороны ситуации, например: *оценка действия, состояние, мешающие факторы, последствия, сравнение ресурсов, результат сравнения, эмоции субъекта, прогноз* и т.д.

Что касается собственно понятия «ситуация», то чёткое определение ситуации как семантической структуры дано в энциклопедическом словаре И.Б. Штерн: «В основу семиотических моделей репрезентации знаний положена идея о том, что универсальной элементарной единицей репрезентации знаний и смысла является ситуация, и что каждая ситуация раскладывается на предикат – «носитель» самого действия, и актанты – объекты, которые создают ситуацию. Функциональные отношения между предикатом и актантами (смысловые роли, семантические категории), отображают семантическую структуру этой ситуации» [Штерн, 1998]. Опираясь на конечное множество ролевых функций актантов в естественном языке, можно создать предпосылки для описания ситуации как целостного семантического объекта.

Одна из серьёзных попыток применить кибернетический подход к определению понятия «ситуация» предпринята в работе А.С. Звенигородского. Он отмечает, что «Задачу понимания смысла текста мы рассматриваем на множестве небольших текстов из простых предложений. Они описывают ситуации из внешнего мира, и в этих ситуациях участвует ограниченное множество образов». Автор вводит понятие «Сенсорной проблемной области ситуаций» (СПОС), которое раскрывается через семантические компоненты, такие как, например: «Точка отсчёта на сцене», «Время», «Шкала свойства положения образа», «Свойства движения образа», «Сцена». Сравнение состояния текущей сцены с состоянием сцены из памяти производится при помощи математической процедуры «Операция» [Звенигородский, 2010].

Явное указание на возможность конструирования ситуации при помощи семантических метапонятий содержится в вышеупомянутой работе [Штерн, 1998]. «Ситуационная (синтагматическая) семантика характеризует, по определению, явные и неявные связи между текстовыми объектами, образующими семантическую сеть. В качестве семантических категорий выступают метапонятия, выражающие глубинные падежи».

Из приведенного обзора мы можем сделать вывод, что путь к глубокому анализу текста лежит через выявление ситуаций, которые служат его **смысловыми блоками**. Таким образом, мы подошли к основной задаче – дать формальное определение **Лексической структуры ситуации**.

Формат Лексической структуры ситуации

Определение 1.

Конструктом ситуации будем называть слово фиксированного метаязыка, которое указывает в предложении *ролевою функцию* автономного лексического выражения и является *семантическим элементом структуры ситуации*.

Конструкт =: <Субъект> / <Цель> / < Объект> / <Признак> / <Инструмент> / <Действие> / <Намерение> / <Причина> / <Условие> / <Сравнение> / <Мера> / < Среда> / <Помехи> / <Место> / <Время> / <Обстоятельство> / <Состояние> / <Следствие> / <Результат> / <Оценка> <Акцент> / <Ожидание> / <Прогноз> / <Пустой конструкт>;

Ролевая функция =: <Конструкт> <Значение>;

Значение =: <Лексическое выражение>.

Лексическое выражение – это семантически-целостный фрагмент предложения, который служит пояснением (*глубинным падежом*, характеристикой) выбранного члена предложения.

Здесь и ниже косая черта используется как разделитель «или». Пустой конструкт служит для вероятного пополнения множества конструктов. Конструкт является нетерминальным символом.

Ролевая функция, например, может иметь вид:

<Субъект> Ксеркс; <Цель> завоевать; < Объект> Греция.

Определение 2.

Конструктивная схема ситуации (КСС) есть кортеж:

$$КСС = \langle O, K, \& \rangle,$$

где: **O** – множество обозначений; **K** – множество конструкторов; **&** – множество связей.

КСС =: <Обозначение> <Последовательность конструкторов>;

Обозначение =: 1, 2, 3, ... ;

Последовательность конструкторов =: < Конструктор> – <Конструктор > – ... – <Конструктор>.

Таким образом, в общем виде Конструктивная схема ситуации представляет собой упорядоченную цепочку конструкторов, которая начинается номером ситуации.

Например, граф **КСС** для произвольной ситуации может иметь такие сочетания конструкторов:

1 – *Субъект – Цель – Действие – Объект – Результат*;

9 – *Цель – Инструмент – Обстоятельство – Результат*;

16 – *Субъект – Признак – Действие – Признак – Объект – Признак*

20 – *Результат – Оценка – Прогноз*.

Определение 3.

Лексическая структура ситуации (ЛСС) – это ориентированный граф, полученный при помощи дерева разбора предложения, вершинами которого служат имя ситуации и ролевые функции главных и второстепенных членов предложения. Базовым понятием ЛСС служит «Ядро ситуации».

ЛСС =: < Имя ситуации > < Ядро ситуации > < Контекст ситуации >;

Имя ситуации =: <Лексическое выражение>;

Ядро ситуации =: < *Субъект*> < *Значение*> & < *Действие*> < *Значение*> & < *Объект*> < *Значение*> &;

Контекст ситуации =: < *Ролевая функция субъекта*> & < *Ролевая функция действия*> & < *Ролевая функция объекта*> & < *Ролевая функция* > ... < *Ролевая функция*>;

& – это знак объединения в группу.

Значениями ядерных конструкторов и остальных конструкторов служат лексические выражения. Ролевые функции отдельных членов ядра могут быть пустыми.

Определение 3 подчёркивает **основную идею**: *ситуация в основном определяется её ядерной конструкцией*. Контекст ситуации несёт в себе вспомогательную или детализирующую информацию. Имя ситуации выбирается либо из текста, либо из базы знаний.

Для чего нужны эти построения?

Формат Лексической структуры ситуации, в котором объединены конструкторы и их лексические значения, даёт возможность разметить лексику текста конструкторами – своеобразными **ключами смысла**. Это означает, что появляется возможность сопоставлять, связывать и обобщать семантику разных предложений. По этим ключам можно проследить динамику развития ситуации относительно *субъекта, цели, результата* и т.д. Открывается путь от анализа простых (*proposal*) к синтезу обобщённых (*generalized*) ситуаций.

Действительно, динамика развития ситуации воспринимается субъективно как *связность текста*. Каждая текущая ситуация сопоставляется с предыдущими и корректирует тот смысл, который был накоплен ранее. Тем самым прослеживается динамика смысла (или нескольких смыслов).

Основываясь на формальном представлении ситуации (ЛСС), мы значительно приближаемся к решению амбициозной задачи – **компьютерного семантического анализа** естественно-языковых текстов со сложной семантической структурой и богатым тематическим содержанием.

Пример представления Лексической структуры ситуации

Текст:

В августе 480 г. до н.э. персидский царь Ксеркс поднял свою могущественную армию – около 200 тысяч опытных воинов, – чтобы покорить Грецию. Эллинскую армию, насчитывающую около 7 тысяч воинов, возглавлял спартанский царь Леонид.

Каждое из двух предложений подвергается следующим этапам обработки:

1. Строится *дерево разбора предложения*. Определяются главные и второстепенные члены, формируются группы подлежащего, сказуемого, дополнения. Фиксируются имена собственные и нарицательные, выявляются синонимические группы слов;
2. Выделяется *ядерная конструкция предложения*, это тройка: «подлежащее – сказуемое – дополнение»;
3. В каждой группе слов (подлежащего, сказуемого, дополнения) определяются *лексические фрагменты*, соответствующие глубинным падежам (прилагательные, обстоятельства, числительные и т.д.);
4. Каждому лексическому фрагменту ставится в соответствие конструкт, и образуются *ролевые функции* «конструкт – значение»;
5. Строится Лексическая структура ситуации (ЛСС) для текущего предложения – **ЛССп**. Это – *граф ситуации предложения*, вершинами которого являются ролевые функции (конструкты вместе с их лексическими значениями), а дугами – связи дерева разбора. Граф ситуации предложения обозначается Именем.

После выполнения указанных этапов обработки текста (в примере – двух предложений) образуется **Ситуативно-лексическое описание текста – СЛОТ**.

Для удобства линейной записи граф ситуации разбит на три подграфа – Я1, Я2, Я3, каждый из которых репрезентует одну компоненту ядра. Кроме того, в лексических функциях конструкт (в скобках) следует за его значением, что не принципиально.

СЛОТ:

ЛССп-1:

«Царь поднял армию»

- + Я1 = Царь (субъект) + Я1 – персидский (признак) + Я1 – Ксеркс (имя);
- + Я2 = Поднял (действие) + Я2 – в августе 480 г. (время) + Я2 – до н.э. (время) + Я2 – чтобы покорить Грецию (цель);
- + Я3 = Армию (объект) + Я3 – свою (признак) + Я3 – могущественную (признак) + Я3 – воинов (объект) + опытных (признак) + около 200 тысяч (мера).

ЛССп-2:

«Царь возглавил армию»

- + Я1 = Царь (субъект) + Я1 – спартанский (признак) + Я1 – Леонид (имя);
- + Я2 = Возглавил (действие);
- + Я3 = Армию (объект) + Я3 – эллинскую (признак) + Я3 – насчитывающую около 7 тысяч (мера) + Я3 – воинов (объект).

Примечание. Знак «Я» служит идентификатором члена ядра ситуации. Знак + обозначает присоединение ядра к имени ситуации, а ролевой функции – к ядру.

Таким образом, мы получили конструктивно-лексическое описание текста в виде двух отдельных ситуаций, соответствующих двум предложениям. Дальнейшая задача состоит в том, чтобы на основании

двух ситуаций сформировать общий **смысл дискурса**. Для этого потребуется выполнить онтологический (концептуальный) анализ текста.

Концептуальный анализ текста

Прежде всего, если судить по именам ситуаций, можно выдвинуть **семантическую гипотезу**, что речь пойдёт о вооружённом конфликте. Гипотеза получает имя – «*Война в древнем мире*».

Чтобы понять смысл текста с позиций семантической гипотезы, необходимо использовать как можно больше знаний о войне. Будем считать, что война является одним из видов человеческой деятельности. В таком случае знания о войне оформлены в онтологии Мира (в Базе Знаний о Мире) в виде предметной области (ПрО) «Война».

Структурно ПрО «Война» представляет собой, как часть онтологии, иерархическую сеть, содержащую концепты и связи. Концепты условно делятся на *конструкты интерпретации* и *значения* конструктов интерпретации. Любой концепт (конструкт или значение) может быть возбуждён словами текста.

Рассмотрим онтологию предметной области «Война». Опуская множество деталей, мы представим усреднённое концептуальное представление о войне следующим образом.

Концептуальное представление ПрО Война:

Война =

= **Субъекты** (участники) – А / – В

= **Время** (эпоха) – Древний мир / – Средние века / – современный мир

= **Место** (военный театр) – страна / – регион

= **Ресурсы** (А, В) – армия / – вооружение / – тыл

= **Цели** (А, В) – покорение / – завоевание / – захват ресурсов / – отмщение / – защита Родины / – победа / – оборона / – капитуляция / – ничья

= **Действия** (А, В) (тактика) – наступление / – сдерживание / – оборона / – отступление

= **Сражение** (А, В) – время / – место / – силы / – результаты

= Действия (А, В) в сражении – атака / – прорыв / – стойкая оборона / – отступление

= Результаты сражения (А, В) – прорыв / отход / гибель / потери / победа / поражение / ничья

= Оценка действий в сражении – успех / неудача / настойчивость / стойкость

= **Общая оценка** действий (А, В) – успех / неудача / ничья

= **Неожиданные обстоятельства** – природные катаклизмы / новая информация / – измена

= **Результаты** (А, В) – победа / – поражение / – ничья

= **Развитие** событий – переговоры / – капитуляция / – наступление / – отступление

= **Общие итоги** (А, В) – победа / – поражение / – ничья / – выигрыш / – потери / – гибель

= **Оценка результата** – подвиг / добровольная гибель / – мужество в атаке / – мужество в обороне / – талант полководца / – трусость / – предательство / – позор

= **Оценка** достижения целей (А, В) – цель достигнута / не достигнута

= **Прогноз** – поражение А или В / перемирие / мир / восстановление исходных позиций

Примечание. Нетерминальные символы применены для указания узлов и связей онтологической сети. Знак «=» отличает конструкт от «–» значения.

Для полного, всестороннего и абсолютно ясного понимания текста желательно ответить на все вопросы, которые закодированы *конструктами онтологии* (выделены шрифтом), а также получить

конкретные **факты** из множества *значений конструкторов*. Если текст удовлетворяет этому условию, то реализацию «Текст – ПрО» можно считать идеальной. Если не все конструкторы выявлены, реализация считается неполной.

Реализация ПрО Война =

Война =

= **Субъект А, Субъект В**

= **Время** (эпоха) – Древний мир

= **Место** (военный театр) – Греция

= **Ресурсы А** – армия – (мера) 200 тысяч

= **Ресурсы В** – армия – (мера) 7 тысяч

= **Цель А** – покорение В

= **Цель В** – ?

Значения конструкторов ПрО «Война» выявлены в заданном тексте. Неполнота реализации ПрО в данном примере обусловлена краткостью текста.

Что даёт реализация ПрО для семантического анализа текста?

Очевидно, что формат реализации ПрО частично совпадает с форматом ситуации ЛСС, поэтому реализацию ПрО возможно также называть «макроситуацией». Отличие двух форматов состоит в их назначении. Если Лексическая структура ситуации (ЛСС) служит только *шаблоном*, моделью некоторой *усреднённой* ситуации, то Концептуальное представление ПрО рисует *конкретную* макроситуацию под названием «Война». (ОК)

Кроме того, существуют и другие отличия: в макроситуации, во-первых, не существует ядра и, во-вторых, в ней опущены некоторые несущественные детали (признаки объектов и другие). Количественные характеристики – (*меры*) двух армий – в примере оставлены потому, что они выявляют **акцент** автора, который таким способом подчёркивает неравенство сил конфликтующих сторон.

Преимущества формата реализации ПрО состоят в следующем:

– Реализация ПрО *структурирована* по конструкторам макроситуации – смысловым позициям интерпретации;

– Конструкторы вместе с их значениями позволяют объяснить события текста в расширенном (выходящем за рамки текста) лексиконе, *обогащённом* онтологией;

– Концептуальная интерпретация текста стремится к лексическому *сокращению* изложения, она даёт *обобщённое* представление дискурса;

– **Реализация ПрО позволяет вести диалог потребителя с текстом на уровне конструкторов, при этом вопросы и ответы раскрывают существенные, глубинные характеристики макроситуации.**

Таким образом, в семантическом анализе текстов ЕЯ **участвуют две технологии**. Наряду с Лексической структурой ситуации – **ЛСС**, которая позволяет индексировать текст конструкторами ситуации, существует вторая форма представления текста – **Реализация ПрО**, которая допускает интерпретацию текста в терминах знаний. Оба способа представления семантики дополняют друг друга.

В конечном итоге мы хотим извлечь из дискурса его главный компонент, прагматический продукт и цель человеческой коммуникации – **смысл**.

На пути к обнаружению смысла дискурса

Смысл – он же бесплатен! (Михаил Жванецкий).

Категория познания – **смысл** – изучается в психологии и лингвистике, в соответствии с методами этих наук. Сегодня задача определения смысла стоит перед машинным интеллектом, где она должна быть сформулирована и решена на конструктивном уровне.

В лингвистической литературе отмечается, что смысл не сводится к значению слова, а является обобщением речевого потока. *«Значение языковой единицы – это элемент языковой системы, в то время как конкретный смысл – это явление речи, к тому же ситуативно обусловленное»* [Бондарко, 1978]. Категория «смысл» рассматривается здесь с гуманитарных позиций. *«Самым важным становится не столько понимание неких фактов и влияния на них тех или иных факторов, условий, механизмов и т.п., сколько отношение человека к этим фактам и воздействиям, смысл, который они для него приобретают»* [Братченко, 1999].

В современной психологии учитывается субъективный аспект познания мира, когда субъект *«ориентируется ... на выражение своего отношения к действительности, на выявление и обозначение смысла ситуации в символической форме»* [Балл, 2006].

Наиболее ясное представление о смысле как о категории понимания мы встречаем у известного американского учёного в области искусственного интеллекта Р. Шенка. Говоря о процедуре семантической обработки естественно-языковых текстов, он выделяет здесь три самостоятельных этапа: **(а)** отображение предложений в их *смысловое представление*; **(б)** хранение в памяти и осуществление умозаключений относительно полученного смысла; **(в)** перевод смыслового представления на *естественный язык* [Шенк, 1980].

Как указывалось выше, целью настоящего исследования является разработка процедур машинной обработки семантически-насыщенных текстов. Для этого привлекаются и формализуются базовые понятия: «ситуация», предметная область знаний (ПрО), онтология. Остаётся дать определение понятиям «смысл ситуации» и «смысл текста».

Ситуация представляет собой целостный семантический объект. **Под «смыслом ситуации» будем понимать краткое лексическое выражение естественного языка, сформированное из ядра ситуации с учётом акцентированных фрагментов текста.** Смысл ситуации, как правило, закрепляется в **имени ситуации**.

Текст, или дискурс, также является семантическим объектом с сильной прагматической нагрузкой, поскольку он должен быть понят и усвоен потребителем. Прагматика текста указывает на то, как ситуация соотносится с компендиумом знаний потребителя, а смысл служит инструментом восприятия и усвоения коммуникативной информации.

«Смысл связного текста» – это предложение естественного языка, которое в сжатой форме передаёт коммуникативную идею автора текста, его информационный замысел.

Подчеркнём, что смысл является лично мотивированной формой сокращения и сбережения больших текстов и служит единственной альтернативой механическому запоминанию неструктурированной информации.

Обобщая сказанное, мы приходим к следующему выводу.

Задача извлечения смысла из текста с **заменой человека компьютером** становится легитимной при выполнении следующих условий:

– в памяти компьютера имеются *концептуальная Картина Мира*;

- есть алгоритмы формального синтеза *текстовых ситуаций*;
- имеются процедуры, выделяющие смыслы отдельных ситуаций и синтезирующие на их основе обобщённый смысл.

Пример представления смысла текста

Исходный текст, состоящий из двух предложений, приведен выше. Для него построено **Ситуативно-лексическое описание текста – СПОТ. Преобразуем его в смысл текста.**

Смысл первого предложения (первой ситуации) возьмём из продукции: «Если субъект активизировал армию, то это – Война». С учётом акцента синтезируем смысл первой фразы: «Война в древнем мире». Теперь имеется основание выдвинуть семантическую гипотезу – «Война» и напрямую обратиться к онтологии предметной области – ПрО «Война».

Выбор гипотезы провоцирует ряд вопросов, которые активизируются **в режиме «Ожидание»**; при этом: открывается ряд конструктов ПрО «Война». Почти все они, за исключением шести конструктов, – не заполнены значениями. Субъекту ситуации (нападающим силам) присваивается обозначение А. Смысл первой фразы становится *накопленным смыслом-1*.

Переходим ко второму предложению. Оно имеет то же ядро. Однако СПОТ заполняется новыми значениями, касающимися стороны В. При этом учитываем акцент автора – он указывает численность двух армий. Следовательно, ресурсы А и В должны быть сравнены и сделан вывод. Сравнительная оценка выполняется логическим процессором, который выдаёт три продукции: 1) «Подавляющее превосходство А над В» и 2) «Из сравнения следует, что А – победа, В – поражение». Первое заключение снабжается конструктом ситуации *сравнение*, а второе получает статус гипотезы и отмечается конструктом *прогноз*. Кроме того, логический процессор сделает вывод, что победа В – нереальна и уточнит: «Цель В – оборона или ничья».

По результатам анализа двух предложений синтезируется *обобщённый смысл, который становится накопленным смыслом-2: «Война в древнем мире с подавляющим превосходством А над В»*. Данный смысл является окончательным. Он имеет форму факта.

На этом текст примера исчерпан, и смысловой анализ текста заканчивается.

В дальнейшем, если появится третье предложение, – будут заполнены новые конструкты ПрО «Война»; в результате этого прежний смысл (обобщённый по двум предложениям) будет изменён и появится смысл, обобщённый по трём предложениям – *накопленный смысл-3*.

Процесс накопления смысла протекает синхронно с развёртыванием текста и построением ситуаций. Периодически, через несколько предложений, производится коррекция накопленного смысла в сторону его сокращения. Сокращение возможно благодаря тому, что в нашем распоряжении имеются имена ситуаций – максимально короткие фразы, которые можно подставлять в *накопленный смысл-N*.

В конце концов, по мере продвижения анализа, накопленный смысл будет трансформироваться и уходить от факта к его глубинному раскрытию. В результате смысловое содержание дискурса почти наверняка выйдет за рамки узкой предметной области «Война» – оно получит **статус оценки** описанных событий.

Более того, мы можем выдвинуть рабочую гипотезу относительно ментального восприятия смысла человеком, которая может быть учтена при создании машинного интеллекта: человек аккумулирует опыт в виде **стандартных ситуаций и стандартных смыслов**. Следовательно, при разработке машинной онтологии ситуаций эти стандартные семантики должны быть обязательно учтены и зафиксированы.

Выводы

Семантический анализ текстов естественного языка требует одновременного использования общих Знаний о Мире, знаний в конкретных областях деятельности человека и распознавания ситуаций. Знания задаются концептуальной онтологией Мира и Предметной областью. Ситуация формируется в процессе разбора предложения из лексики текста под управлением конструкторов, которые «задают вопросы тексту» и выделяют из предложения те фрагменты, которые служат «ответами», в результате чего ситуация приобретает структуру и стандартный вид (в формате ЛСС).

Центральное место в формате ситуации занимают Ядро и Имя ситуации. Выбор конструкторов зависит от того, к какой области человеческой деятельности (ПрО) относится дискурс.

В процессе семантического анализа лексика ситуации заменяется понятиями онтологии, в результате чего происходит обобщение ситуаций и описание их концептами высокого уровня репрезентации. Результатом анализа текста является его смысл (или несколько допустимых смыслов).

Концептуально-ситуационный анализ является новым шагом к глубокому пониманию текстов и ориентирован на компьютерную обработку семантических структур.

Благодарности

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария www.ithea.org и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина www.aduis.com.ua.

Библиография

- [Мельчук, 1974] Мельчук И.А. Опыт теории лингвистических моделей «Смысл–Текст». – М.: Наука, 1974.
- [Демьянков, 1985] Демьянков В.З. Основы теории интерпретации и её приложения в вычислительной лингвистике. – М.: Изд-во Моск. Ун-та, 1985.
- [Рубашкин, 1989] В.Ш. Рубашкин. Представление и анализ смысла в интеллектуальных информационных системах. – М.: Наука, 1989.
- [Шенк, 1980] Шенк Р. Обработка концептуальной информации: Пер. с англ. – М.: Энергия, 1980.
- [Витгенштейн, 2009] Витгенштейн Л. Логико–философский трактат: Пер. с нем. – М.: Наука, 2009.
- [Звенигородский, 2010] А.С. Звенигородский. Кибернетические основы процесса понимания смысла текста. // Искусственный интеллект. – 2010. – № 4.
- [Гаврилова, Воинов, 1997] Гаврилова Т.А., Воинов А.В. Визуальный концептуальный анализ знаний и данных для интеллектуальных и обучающих систем // VI международная конф. «Знание – Диалог – Решение» (KDS – 97). Сборник научных трудов в 2-х томах. Т.1. – Ялта 1997.
- [Рыбина, 2010] Г.В. Рыбина. Основы построения интеллектуальных систем: учеб. пособие. – М.: Финансы и статистика; ИНФРА-М, 2010.
- [Гладун и др., 2008] В. Гладун, В. Величко, Л. Святогор. Структурирование онтологии ассоциаций для конспектирования естественно-языковых текстов // Advanced Research in Artificial Intelligence. International Book Series, Number 2. Supplement to the International Journal “Information Technologies & Knowledge”, Volume 2 / 2008. ITHEA, Sofia, 2008.
- [Гладун, Святогор, 2011] Гладун В.П., Святогор Л.А. Вычисление смысла в парадигме искусственного интеллекта // Интегрированные модели и мягкие вычисления в искусственном интеллекте. Сборник научных трудов VI

Международной научно-технической конференции (Коломна, 16 – 19 мая 2011 г.). В 2-х томах. Т.2. – М.: Физматгиз, 2011.

[Штерн, 1998] Штерн І.Б. Вибрані топіки та лексикон сучасної лінгвістики. Енцикл. Словник. – Київ: «АртЕк», 1998.

[Кисленко, 2008] Кисленко Ю.И. От мысли к знанию (нейрофизиологические основания): Монография. – Киев, изд-во «Український літопис», 2008.

[Арно, Николь, 1991] Антуан Арно, Пьер Николь. Логика, или искусство мыслить. Наука, 1991

[Звягинцев, 2009] А.И. Звягинцев. Логико-дискурсивная модель экспертной поддержки принятия решений в ситуациях неопределённости // Системи підтримки прийняття рішень. Теорія і практика: Збірник доповідей науково-практичної конф. – Київ: ІПММС НАНУ, 2009.

[Бондарко, 1978] Бондарко А.В. Грамматическое значение и смысл. – М.: Наука, 1978.

[Братченко, 1999] Братченко С.Л. Введение в гуманитарную экспертизу образования (психологические аспекты). – М., 1999.

[Балл, 2006] Г.А. Балл. Психология в рაციогуманистической перспективе. – К., 2006.

Информация об авторах



Святогор Леонід Александрович – *Ин-т кибернетики им. В.М. Глушкова НАН Украины, Киев-187 ГСП, 03680, просп. Акад. Глушкова, 40, e-mail: aduis@rambler.ru*



Гладун Виктор Поликарпович – *Ин-т кибернетики им. В.М. Глушкова НАН Украины, Киев-187 ГСП, 03680, просп. Акад. Глушкова, 40, e-mail: aduis@rambler.ru*