



OSTIS-2013

(Open Semantic Technologies for Intelligent Systems)

УДК 004.822:514

СЕМАНТИКО-СИНТАКСИЧЕСКАЯ МОДЕЛЬ ТАТАРСКОГО ПРЕДЛОЖЕНИЯ В КОНТЕКСТЕ РЕЛЯЦИОННО-СИТУАЦИОННОЙ СИСТЕМЫ

Сулейманов Д.Ш., Гатиатуллин А.Р., Вагапов Д.Р.

*НИИ «Прикладная семиотика»
Академии наук Республики Татарстан,
Казанский (Приволжский) Федеральный университет
г. Казань, Россия
dvdt.slt@gmail.com
agat1972@mail.ru*

В работе описывается семантико-синтаксическая модель для представления структуры простого татарского предложения, основанная на идеереляционно-ситуационной системы. Реляционно-ситуационная система представляет собой взаимосвязанную совокупность сценариев, называемых ситуациями. Сценарий используется для описания структуры простого предложения, где элементами сценария являются синтаксемы.

Ключевые слова: реляционно-ситуационная система; семантические универсалии, синтаксемы, многословные конструкции.

ВВЕДЕНИЕ

В настоящее время в области обработки текстов, написанных на языках тюркского семейства, задача синтаксического анализа предложения продолжает оставаться одной из наиболее актуальных задач, поскольку для них нет такого множества известных разработок в области синтаксиса как для индоевропейских языков.

Эта задача является основной составляющей для многих разработок, которые ведутся в Институте Прикладной семиотики Академии наук Республики Татарстан по компьютерной обработке татарских текстов. Это такие прикладные задачи, как многоязычный интеллектуальный поиск, машинный перевод для близкородственных языков, и ряд других.

Для моделирования использованы следующие ограничения: модели строятся только для простых предложений татарского языка без причастных и деепричастных оборотов.

Семантико-синтаксическая модель татарского предложения представляет собой прагматически ориентированный контекст объектно-предикатной системы, разработанной для описания семантического аспекта структурно-функциональной модели татарской аффиксальной морфемы [Сулейманов, 2003], как оказалось, концептуально весьма близкой к реляционно-ситуационной системе, разрабатываемой группой Г.С. Осипова [Осипов, 1997].

1. Семантико-синтаксическая модель в реляционно-ситуационной системе

Реляционно-ситуационная система состоит из совокупности семантических сценариев, называемых нами семантическими универсалиями. Эти семантические сценарии по сути являются неоднородными семантическими сетями [Осипов, 1997]. Текст предложения в этой модели представляется в виде ситуации, с определенным набором участников и отношений между ними.

Авторами статьи проведена семантическая классификация контекстов в зависимости от типов отношений, участвующих в формировании глубинного смысла данного контекста. Соответственно, формируются семантические сценарии, в которых участвуют объекты и отношения, причем, сценарии строятся от простого к сложному. Простые сценарии – описывают простые типы отношений, сложные сценарии – это комбинации нескольких взаимосвязанных семантических сценариев, называемых ситуациями. Соответственно, семантические сценарии являются теми самыми семантическими универсалиями, которые используются для описания глубинного смысла определенного контекста.

Предлагаемый авторами вариант реляционно-ситуационной системы имеет следующую структуру [Сулейманов, 2003]:

1. Атрибутивные отношения

- a. Пространственные отношения
 - b. Отношения классификации
 - c. Отношения принадлежности
 - d. Признаковые отношения
 - e. Количественные отношения
 - f. Отношения сравнения
2. Темпоральные отношения
- a. Состояние
 - b. Изменение состояния
 - c. Изменение пространственных отношений
 - d. Изменение отношений принадлежности
 - e. Изменение отношений классификации
 - f. Изменение количественных отношений
 - g. Изменение качественных отношений
 - h. Воздействия на объекты
 - i. Понудительные действия
3. Отношения над ситуациями
- a. Каузальные отношения
 - b. Временные отношения
 - c. Порядковые отношения
4. Образ человека
- a. Физическое восприятие
 - b. Речь
 - c. Желание
 - d. Эмоции
 - e. Мышление
 - f. Модальные отношения

Рассмотрим примеры из данной системы. Один из примеров это *Действие* по изменению пространственного отношения. Общий вид схематического представления подобных действий приведен на рисунок 1.

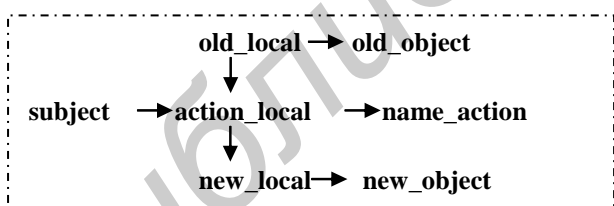


Рис. 1. Действие по изменению пространственного отношения

Подставим в общую схему конкретный пример:

Мин дивангайттым 'Ялегна диван'

Получим заполнение этой схемы, приведенное на рисунке 2.

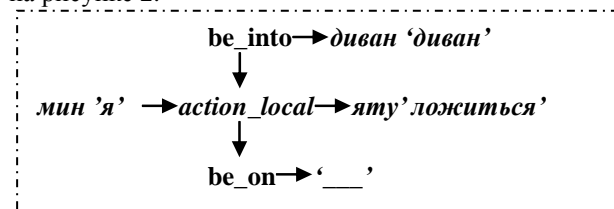


Рис.2. Пример действия по изменению пространственного отношения

Следующий пример – это действие типа воздействие. Схематическое представление действий подобного вида приведен на рис.3.

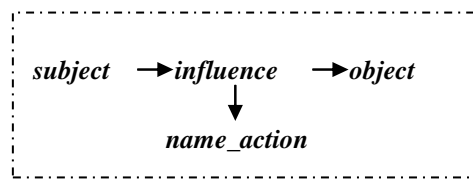


Рис.3. Действие типа воздействие

Заполним данную схему конкретным примером: *Малай кадаккасукты. 'Мальчик ударил по гвоздю'.*

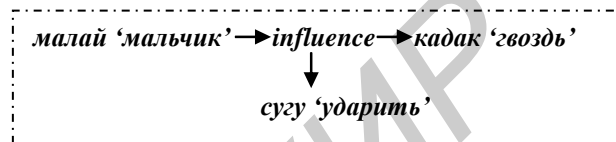


Рис.4. Пример действия типа - Воздействие

Представление текста в виде семантической сети позволяет находить смысловую близость двух разных текстов сравнением соответствующих им семантических сетей. Эта особенность используется как при интеллектуальном поиске, так и как механизм снятия неоднозначностей в задачах машинного перевода.

Участники ситуации в данной семантико-синтаксической модели представляются синтаксемами.

Согласно теории коммуникативной грамматики Золотовой [Золотова, 2004] предполагается тесная связь синтаксиса и семантики в языке, их взаимодействие при выполнении коммуникативной функции языка. Синтаксис играет важную роль в формировании осмысленных предложений. Слова-лексемы, единицы словарного состава языка в совокупности его конкретных грамматических форм, а также возможных конкретных смысловых вариантов, сами не передают смысл высказывания. Синтаксис имеет дело с осмысленными единицами, несущими свой не индивидуально-лексический, а обобщенный, категориальный смысл в конструкциях разной степени сложности. Эти единицы характеризуются взаимодействием морфологических, семантических и функциональных признаков, которые называются синтаксемами.

Синтаксема в тюркских языках характеризуется морфологической структурой – набором аффиксальных морфем. Значение синтаксемы является единицей передаваемого смысла, например, субъект – производитель действия. В конкретном предложении лексическая единица выступает в качестве единицы смысла именно как синтаксема. Таким образом, при работе с текстом необходимо оперировать не лексемами, а синтаксемами.

Общий вид ситуации, описываемой в нашей модели выглядит следующим образом:

Ситуация Q

Предикат: Синтаксема₀;

Роль 1: Синтаксема₁;

...

Роль k: Синтаксема_k.

Приведем пример заполнения этой структуры для ситуации типа Воздействие:

Influence: Синтаксема₀;

Subject: Синтаксема₁;

Object: Синтаксема₂.

Рассмотрим, что представляет собой структура синтаксемы:

Синтаксема: Название

Структура: Набор морфем, корневые и аффиксальные.

Семантика: Семантические ограничения, в виде лексико-семантической группы.

Грамматическая структура синтаксемы представляет собой ее морфемную структуру. Здесь описывается из каких корневых и аффиксальных морфем образована данная синтаксема.

Например:

Поверхностная форма: *урманнарга*

Морфемная структура: урман+ЛАр+ГА

Категориальная структура: N+Plu+Dir.

Структура синтаксемы в своем составе может иметь больше одной корневой морфемы. Это бывает в тех случаях, когда синтаксема представляет собой многословную конструкцию.

2. Синтаксемы, выражаемые многословными конструкциями

Синтаксемы в татарском языке по своей структуре могут быть образованы как отдельными словоформами, так и сочетанием из нескольких словоформ. В нашем случае, синтаксемы из нескольких словоформ – это несвободные словосочетания, составляющие единство с точки зрения семантики, которые в зарубежной литературе обозначаются как многословные выражения – MultiwordExpression, или многословные конструкции.

Под многословными выражениями нами понимаются аналитические конструкции, т.е. все конструкции, значение которых невозможно вывести из значения их составляющих.

В работе А.Н. Баранова [Баранов, 2008] отмечается, что в описании семантики, в целом наиболее правильным и единственно научным является атомистический подход. Однако этот подход не работает для описания семантики многословных выражений.

Компьютерная обработка многословных конструкций, значение которых не определяется из определения значений их составляющих, на

сегодняшний день продолжает оставаться одним из наиболее актуальных вопросов в NLP-задачах. Свидетельством этому служит ежегодный международный семинар Workshop on Multiword Expression, который действует с 2003 года [<http://multiword.sourceforge.net/>].

В нашей модели [Сулейманов, 2012] все многословные конструкции в соответствии со своими структурами делятся на два основных вида:

1. Многословные аналитические формы (МАФ).
2. Многословные аналитические конструкции (МАК).

К первому типу относятся все конструкции, в которых основа первого элемента не привязана к остальным элементам конструкции, то есть, на это слово накладываются не лексические, а только морфологические ограничения. Это означает, что для образования конструкции подобного типа на месте основного слова может быть любое слово определенной категории, но в строго заданной морфологической форме, то есть, должно завершаться на определенную морфему. Вместе с тем, вспомогательное слово имеет строго фиксированную корневую морфему. Это означает, что данная конструкция имеет универсальный характер, соответственно, в нее может быть подставлена корневая морфема такого же морфологического типа.

Ко второму типу относятся конструкции, в которых основы всех элементов конструкции фиксированы.

Структура синтаксемы, представляющей собой многословное выражение, будет выглядеть следующим образом:

Поверхностная форма: *урманга кадәр.*

Морфемная структура: урман+ГА кадәр.

Категориальная структура: N+DirS.

3. Семантически-признаки. Лексико-семантические группы

Ситуации состоят из объектов, значений и отношений между ними. Причем, на сочетаемость этих объектов и отношений внутри ситуации накладываются определенные ограничения. Так, например, есть ряд ситуаций, в которых исполнителем действия может быть только человек, в нашей модели это группа отношений “Образ человека”. Это означает, что определенную роль в ситуациях этого типа могут заполнять только участники, принадлежащие к определенной лексико-семантической группе.

Лексико-семантическая группа (ЛСГ) – группа слов одной части речи, объединенная одним словом-идентификатором или устойчивым словосочетанием, значение которого полностью входит в значение остальных слов группы и которое может заменять остальные слова в некоторых

контекстах.

Сходство слов, входящих в одну ЛСГ, не исчерпывается только наличием у них общих лексических признаков, а также проявляется в наличии типовой сочетаемости, связанности с определенными структурно-семантическими моделями предложений, в закономерностях соотношений парадигматики внутри группы, в единстве словообразовательных характеристик, с направлением развития многозначности, которое часто носит регулярный характер, и т.д.

Например:

абый сиңа хатында урман турында сөйләде 'в своем письме братрассказал тебе о лесе'

Say: S0: *сө йлә де (рассказал)*

Structure: V+(*)

LSG: излагать

Agent: S1: *абый (брат)*

Structure: N

LSG: человек

Address: S1: *сиң а (тебе)*

Structure: N+Dir(ГА)

LSG: человек

Tool: S3: *хатында (в своем письме)*

Structure: N+Loc([н]ДА)

LSG: средство передачи информации

Maintenance: S4: *урман турында*

Structure: Nтурында

LSG: сущность

Здесь V+(*) означает, что может быть глагол с любым набором аффиксов, морфотактически допустимых для глагола.

ЗАКЛЮЧЕНИЕ

В данной работе предложены семантико-синтаксические модели татарского простого предложения. Реализация семантико-синтаксического анализатора татарского предложения на базе этих моделей требует подготовки целого ряда лингвистических ресурсов. В их число входят базы данных с иерархическими классификациями Лексико-семантических групп, с семантическими сценариями, словарей синтаксем, а также словарей семантических ролей глаголов. На данный момент для татарского языка этих ресурсов не существует, или они находятся на стадии разработки.

Поддержан грантом РФФИ («Разработка моделей и программных средств прикладного синтаксического анализа татарских предложений»), проект № 12-07-00735-а.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

[Сулейманов и др., 2003] Сулейманов Д.Ш., Гатиатуллин А.Р. Структурно-функциональная модель татарских морфем. Казань, 2003. 212 с.

[Сулейманов и др.(1), 2012] Сулейманов Д.Ш., Гатиатуллин А.Р. Семантические универсалии в системе описания значений морфем // Сборник Трудов конференции «Открытые семантические технологии проектирования интеллектуальных систем» (OSTIS-2011), Минск: Издательство БГУИР, 2012. – С.391-396.

[Сулейманов и др.(2), 2012] Сулейманов Д.Ш., Гатиатуллин А.Р., Галеева А.М., Сулейманова Д.Д. Контекстно-ориентированные семантические универсалии // Сборник трудов Тринадцатой национальной конференции по искусственному интеллекту с международным участием. Белгород. Россия. Т.1. – Белгород: Изд-во БГТУ, 2012. – С.275-284.

[Золотова, 2004] Золотова Г.А., Ониненко Н. К., Сидорова М. Ю. Коммуникативная грамматика русского языка. Институт русского языка РАН им. В. В. Виноградова, М, 2004 – 544 с.

[Баранов, 2008] Баранов А.Н. Против «разложения смысла»: узнавание в семантике идиом // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 4–8 июня 2008 г.). Вып. 7 (14). – М.: РГГУ, 2008. – С.39-44.

[Осипов и др., 2006] Осипов Г.С., Тихомиров И.А., Смирнов И.В. Ехастус – система интеллектуального метапоиска в сети Интернет. // Труды десятой национальной конференции по искусственному интеллекту с международным участием КИИ-2006. М: Физматлит, 2006. т. 3. – С. 859–866.

[Сулейманов и др., 2003] Сулейманов Д.Ш., Гатиатуллин А.Р. Модель многословных конструкций татарского языка: аналитические формы // Журнал «Казанская наука». № 12. 2012. В печати.

SEMANTICO-SYNTACTIC MODEL OF THE TATAR SENTENCE IN THE CONTEXT OF RELATIONAL-SITUATIONAL SYSTEM

Suleymanov D. Sh., Gatiatullin A. R.,
Vagapov D.R.

*Scientific Research Institute "Applied Semiotics"
Academy of Sciences of the Republic of Tatarstan,
Kazan Federal University
Kazan, Russia*

dvd.t.slt@gmail.com

agat1972@mail.ru

In this work the semantico-syntactic model for representation of the structure of the simple Tatar sentence is described. This model is based on the idea of the relational-situational system, representing a system of scenarios, called situations. The scenario is used for the description of the structure of a simple sentence, where the syntaxemes are the elements of the scenario.

Key words: semantic universals, morphemic model, object-predicate system, semantic aspect