



OSTIS-2012

(Open Semantic Technologies for Intelligent Systems)

УДК 004.822

СЕМАНТИЧЕСКИЕ УНИВЕРСАЛИИ В СИСТЕМЕ ОПИСАНИЯ ЗНАЧЕНИЙ МОРФЕМ

Сулейманов Д.Ш., Гатиатуллин А.Р.

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

В работе описываются семантические универсалии, представляющие собой семантические сценарии и в совокупности составляющие реляционно-ситуационную систему. Реляционно-ситуационная система предназначена для описания значений аффиксальных морфем, представленных в структурно-функциональной модели морфем. Семантические универсалии рассматриваются на примере описания значений падежной морфемы –ГА (директив).

Ключевые слова: семантические универсалии, модель морфем, реляционно-ситуационная система, семантический аспект

ВВЕДЕНИЕ

Такие извечные и, одновременно, не имеющие до сих пор эффективных решений проблемы, как понимание текстов на естественном языке, автоматизированный поиск многоязычной информации, машинный перевод с языка на язык, и ряд других, настраивают исследователей на поиск «прорывных» идей и подходов в этой области. Практически является общепризнанным, что возможное решение лежит на плоскости создания семантических технологий, позволяющих проникновение на глубинный, «доязыковой», уровень значения (смысла) лингвистического объекта (морфемы, слова, предложения, фразы и т.д.). Вместе с тем, далее исследователи расходятся как в интерпретации самого понятия семантика, так и в методах и принципах ее описания и извлечения. Одни пытаются построить универсальный язык, систему кодирования смысла [Мартынов, 1994; Лезин и др., 1996], другие – контекстные семантические формулы, грамматики, семантические роли и падежи [Шенк, 1980; Филлмор, 1981; Апресян и др., 1982; Бухараев и др., 1990], третьи – семантические универсалии, как основы универсальной интерпретации, претендующие на роль инструментария, в совокупности наиболее полно описывающего смысловые конструкции ЕЯ-текстов независимо от

языка и предметной области [Поспелов, 1981; Осипов, 2006; Закиев, 1992].

В данной работе авторы описывают семантические универсалии в системе представления значений морфем, которые при их дальнейшем развитии могут быть использованы для семантического описания и более сложных лингвистических конструкций. Как известно, морфемы являются наименьшими значащими (семантическими) единицами языка, из которых составляется словоформа, а далее, соответственно, и лексема. А из лексем составляются предложения и фразы. Значение морфологического уровня при обработке ЕЯ-текстов, особенно для языков агглютинативного типа, общепризнанно и создание открытой концептуально-функциональной модели морфем, отражающей свойства морфем во всех проявлениях его в структуре текста, включая фонетический, морфологический, синтаксический и семантический уровни, является несомненно актуальной и важной задачей. Ценность и прагматическая ориентированность концептуально-функциональных моделей, отнесенных нами к лингвистическим моделям концептуально-функционального класса [Сулейманов, 1998], главным образом, заключается в том, что они являются универсальными, наиболее полными описаниями значимых единиц языковых уровней и их грамматик, и являются той базой

«лингвистического строительного материала», из которого на основе прагматически-ориентированной технологии строятся лингвистические модели как основа систем обработки ЕЯ-текстов определенного типа. К тому же, имея максимально полные концептуально-функциональные модели для составляющих элементов на разных языковых уровнях мы можем на их основе организовать тот самый эффективный асинхронный децентрализованный анализ продукта языковой деятельности, о которой говорится в работах Нариньяни А.С. [Труды, 1996]. Несмотря на методологическую, теоретическую и практическую ценность, такого рода модели крайне слабо исследованы и отражены в литературе. Также косвенным подтверждением важности и необходимости построения такого рода моделей являются аналогичные исследования, проводимые за рубежом [Casey и др., 1992]. В работе [Pulman, 1995] утверждается близкая мысль о том, что интеллектуальная обработка естественного языка в реальных приложениях требует наличия лексикона, который обеспечивает разработчика богатой информацией о морфологических, синтаксических и семантических характеристиках слов, хорошо структурированных, и которые могут быть эффективно применены. Эти цели могут быть достигнуты разработкой инструментария, который облегчает приобретение лексической информации с машиночитаемых словарей и корпусов текстов, также из баз данных и теоретических знаний о слове, предлагаемых в кодированном виде, необходимом для целей NLP (*Natural Language Processing*). В этих работах, как правило, исследуются языковые единицы и их характеристики на уровне слов (полисемия, описание значений, допустимые употребления слова). Учитывая особую важность морфологического уровня для татарского языка как языка агглютинативного типа, а также отсутствие такого рода исследований и технологических решений, авторами данной статьи разработана и описана компьютерная структурно-функциональная модель аффиксальных морфем, которая вполне может быть использована также для описания морфем других языков [Сулейманов и др., 2003].

Наибольшую сложность при создании структурно-функциональной модели морфем вызвала разработка формальной структуры для отражения значений (семантики) морфем. Практически, впервые возникла необходимость описания значения татарских морфем на некотором формальном языке «глубинных структур», т.е. на языке семантического уровня. В данной работе дается краткое описание структурно-функциональной модели морфем и описание реляционно-ситуационной системы, состоящей из совокупности семантических сценариев, называемых нами семантическими универсалиями, служащими для обозначения позиции той или иной морфемы в соответствующем контексте. С целью обеспечения компактности изложения и раскрытия

сути идеи семантических универсалий для описания значений морфем, здесь приводится не полное описание всей совокупности универсалий, а лишь той ее части, которая используется для контекстного описания значений морфемы –Г4, маркируемой в татарской грамматике как аффиксальная морфема направительного падежа (или директив).

1. Структурно-функциональная модель аффиксальных морфем

Структурно-функциональная модель аффиксальных морфем представляет собой концептуально-функциональное описание элементов морфологии и используется для построения прагматически-ориентированных морфологических модулей как основы лингвопроцессоров. Особенностью данной модели является то, что она позволяет осуществить полную «инвентаризацию» татарских аффиксальных морфем с описанием функций и ситуаций проявления их на всех языковых уровнях (формирования звука, словоформы, предложения, значения). Важными при построении компьютерной структурно-функциональной модели являются концептуальные (полнота описания характеристик), технологические (удобство обработки, прагматичность) и инструментальные аспекты разработки и реализации (открытость, наличие функций и средств для доступа и применения).

Из названия самой модели - структурно-функциональная, следует, во-первых, что она структурно, т.е. на разных языковых уровнях (фонологическом, морфологическом, синтаксическом, семантическом) отображает проявление свойств морфемы. Модель является открытой, что позволяет вносить в нее даже единичные, незначимые на первый взгляд, характеристики или проявления аффикса в словоформе, при необходимости модифицируя и саму структуру. Во-вторых, модель функциональная, т.е. после наполнения ее лингвистами-экспертами является информационно-справочной и учебно-консультационной базой при изучении и преподавании татарского языка, а также мета-моделью, используемой для разработки прагматически-ориентированных лингвистических моделей в качестве формальной базы для лингвопроцессоров. Модель имеет структуру, в которой каждая морфема характеризуется по пяти основным параметрам, называемым аспектами:

- 1) Функциональный;
- 2) Морфологический;
- 3) Синтаксический;
- 4) Семантический;
- 5) Морфонологический.

Каждый основной параметр подразделяется на подпараметры, а те, в свою очередь, также могут подразделяться на подпараметры и т.д. [Сулейманов

и др., 2003]. При описании семантики морфем мы исходим из утверждения, что каждая морфема используется для кодирования того или иного значения в некотором контексте, отражающем некую локальную "модель мира" или здравый смысл "говорящего", отображаемого в "слушающем". Использование аффиксальных морфем, внося определенные семантические оттенки, ограничения в значение корневой морфемы, позволяет существенно сократить количество лексем для передачи (кодирования) некоего смысла, то есть служит как элемент, редуцирующий лексическое пространство, необходимое для формирования контекста. Локальная модель мира представляет собой формализованное описание некоторого контекста, отражающего объекты и их отношения. Разделение лексем или групп лексем на объекты и отношения является достаточно условной процедурой и зависит от семантических ролей, исполняемых лексемами или группами лексем, отражающими некие значения в определенном контексте.

Примечание 1. Наша задача заключается не в том, чтобы построить семантическую модель, которая позволяет "понять", определить "смысл" некоторого текста, а в том, чтобы построить локальную концептуальную модель мира, которая может быть соотнесена со значениями конкретных морфем и, соответственно, размечена соответствующими им поверхностными выражениями аффиксальных морфем.

Как известно, в лингвистике разделяются такие понятия, как значение высказывания (или сущности) и его пресуппозиция [Моделирование, 1997]. Пресуппозиция определяется как предшествующий контекст, предшествующее знание, или как контекст, в котором происходит определение значения сущности. Очевидно, что понимание сущности, прежде всего, обеспечивается именно пресуппозицией. Следовательно, при определении значений аффиксальной морфемы важно рассматривать ее не только как часть словоформы, а также как составляющую контекста, образующего пресуппозицию.

Попытка описать значения каждой морфемы в подразделе «Семантический аспект» компьютерной модели татарских морфем в некотором унифицированном виде, с целью их дальнейшего использования в лингвопроцессорах, привела к необходимости систематизации отношений, выражаемых аффиксальными морфемами, и создания целостной картины этих отношений. Как известно, значение морфемы, зачастую, может быть установлено только исходя из контекста, в котором она встречается и является трудноформализуемым параметром. Тем не менее, создание некоего формального аппарата, возможно, нечеткого, позволяющего максимально полно отразить те отношения, в лексикализации которых участвуют рассматриваемые морфемы, повышает

репрезентативность, прозрачность и прикладную ценность модели. 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 – комбинация нескольких взаимосвязанных семантических сценариев, называемые ситуациями. Соответственно, семантические сценарии являются теми самыми семантическими универсалиями, которые используются для описания глубинного смысла определенного контекста и включаются в модель описания значений морфем. В дальнейшем происходит локализация конкретной морфемы в соответствующей позиции семантического универсалия.

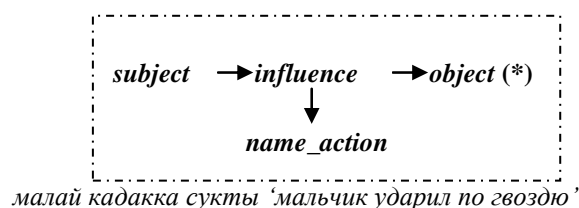
В этом разделе приводится описание семантических универсалий на примере аффиксальной морфемы *-ГА*, являющихся частью реляционно-ситуационной системы и используемых в структурно-функциональной модели для формального представления значений морфем.

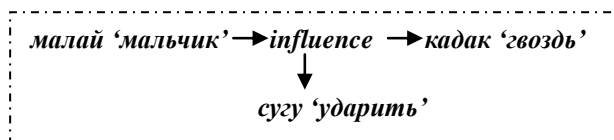
2. Описание семантических универсалий для формального представления значений аффиксальной морфемы *-ГА*

Аффиксальные морфемы по определению, как минимальные значащие единицы языка, имеют хотя бы одно значение, проявляющееся при использовании его в словоформе. В татарском языке, как правило, в зависимости от окружения, аффиксальные морфемы имеют различные интерпретации, точнее, используются для поверхностного оформления различного рода контекстов. В результате исследований нами выявлено 19 контекстных значений аффиксальной морфемы *-ГА*. Это означает, что данный аффикс в составе словоформы участвует в формировании (или поверхностном описании, лексическом проявлении) 19 контекстных значений.

Далее рассмотрим 10 семантических универсалий, в заполнении которых лексическими единицами по заданному тексту участвует морфема *-ГА*. В рассматриваемых универсалиях позиция, маркируемая (заполняемая) данной морфемой, обозначена символом звездочка (*).

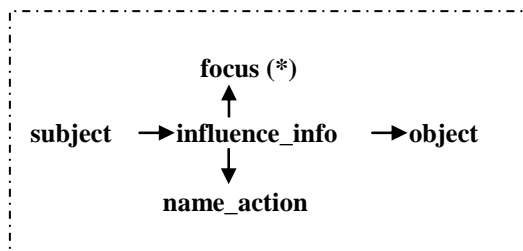
Значение 1. Зависимое имя с аффиксом *-ГА* выражает объект физического воздействия.



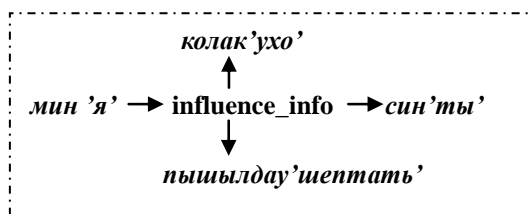


В реляционно-ситуационной системе рассматриваемая словоформа заполняет следующую ролевую ситуацию: **Отношение Действие: Действие-воздействие: Объект, испытывающий физическое воздействие.**

Значение 2. Зависимое имя с аффиксом **-ГА** выражает объект информационного воздействия

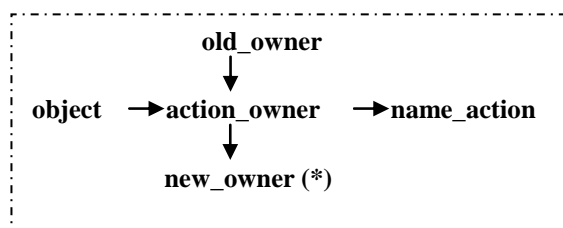


мин синец колагыңа пышылдадым 'я прошептал тебе на ухо'

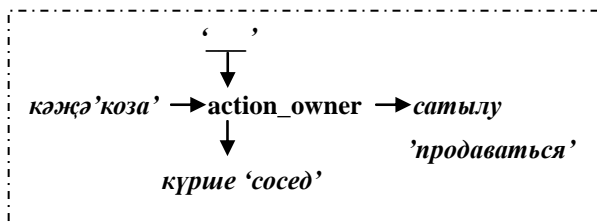


В реляционно-ситуационной системе рассматриваемая словоформа заполняет следующую ролевую ситуацию: **Отношение Действие: Отношение воздействия: Объект, испытывающий информационное воздействие.**

Значение 3. Зависимое имя с аффиксом **-ГА** выражает нового владельца объекта:



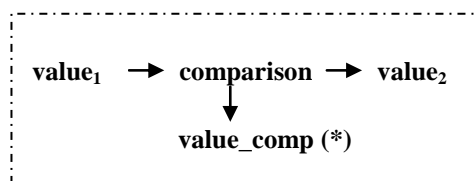
кәжә күршегә сатылды 'коза продана соседу'



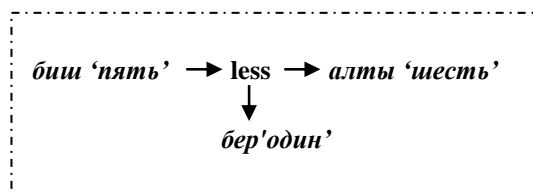
В реляционно-ситуационной системе рассматриваемая словоформа заполняет следующую

ролевую ситуацию: **Отношение Действие: Действие по изменению отношения принадлежности: Новый владелец.**

Значение 4. Зависимое имя с аффиксом **-ГА** выражает разницу (степень различия) между значениями.

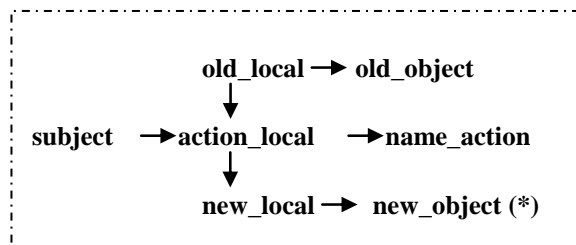


биш алтыдан бергә азрак 'пять на единицу меньше шести'

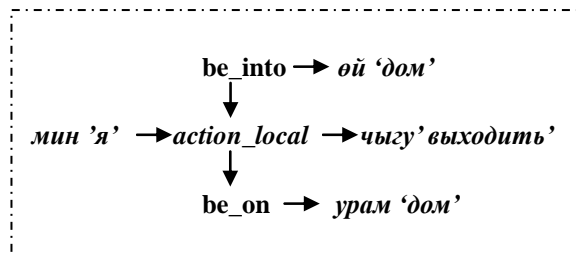


В реляционно-ситуационной системе рассматриваемая словоформа заполняет следующую ролевую ситуацию: **Отношение Сравнения: Разница между значениями.**

Значение 5. Зависимое слово с аффиксом **-ГА** выражает конечный пункт направления процесса, обозначенного глаголом движения.



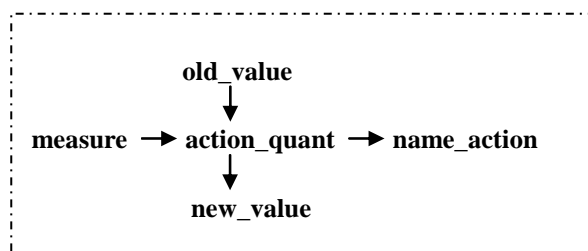
мин өйдән урамга чыктым 'я вышел из дома на улицу'



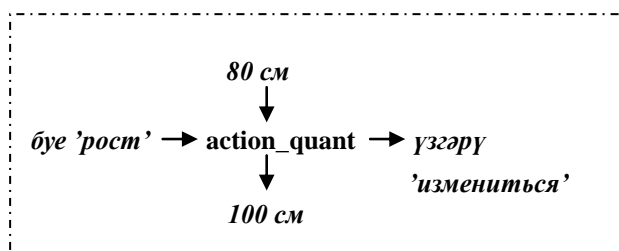
В реляционно-ситуационной системе рассматриваемая словоформа заполняет следующую ролевую ситуацию: **Отношение Действие: Действие по изменению пространственного отношения: Новое местоположение субъекта.**

Значение 6. Зависимое имя с морфемой **-ГА**, которое имеет при себе конкретизирующее

числительное, подчиняясь глаголам действия, выражает меру и степень действия в пространстве и во времени.

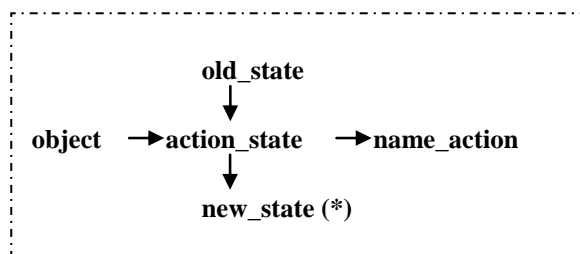


буе 80 сантиметрдан 100 сантиметрга кадәр үзгәрде 'рост изменился с 80 до 100 сантиметров'

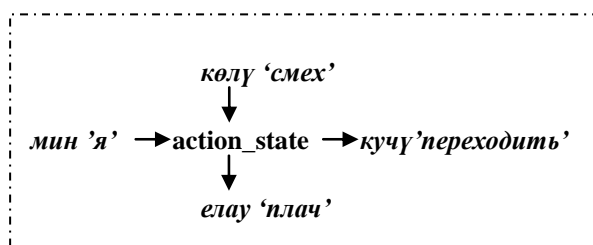


В реляционно-ситуационной системе рассматриваемая словоформа заполняет следующую ролевую ситуацию: **Отношение Действие:** Действие по изменению количественного отношения: Разница между значениями.

Значение 7. Зависимое слово с аффиксом -ГА выражает новое состояние объекта.

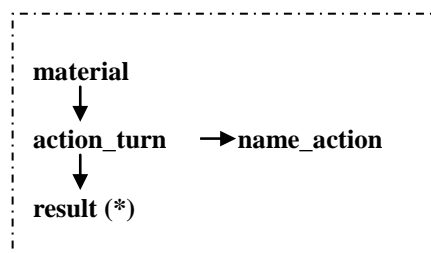


мин көлдән елауга күчтем 'я перешел от смеха к плачу'

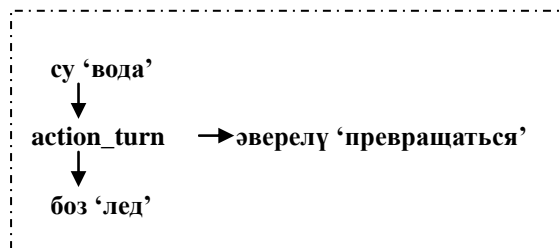


В реляционно-ситуационной системе рассматриваемая словоформа заполняет следующую ролевую ситуацию: **Отношение Действие:** Действие по изменению состояния: Новое состояние объекта.

Значение 8. Зависимое слово с аффиксом -ГА выражает новый объект.

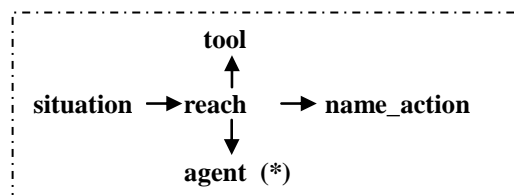


су бозга әверелде 'вода превратилась в лед'

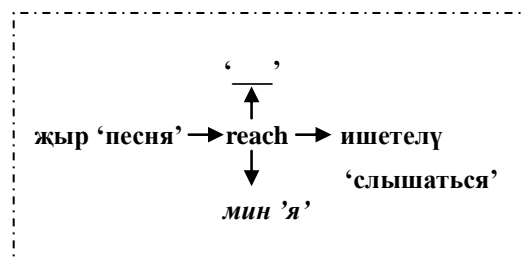


В реляционно-ситуационной системе рассматриваемая словоформа заполняет следующую ролевую ситуацию: **Отношение Действие:** Действие по изменению состояния: Новый объект.

Значение 9. Зависимое слово с аффиксом -ГА выражает субъекта восприятия ситуации.

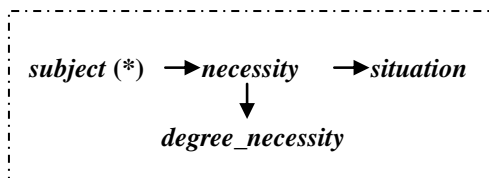


миңа жыр ишетелде 'мне послышалась песня'

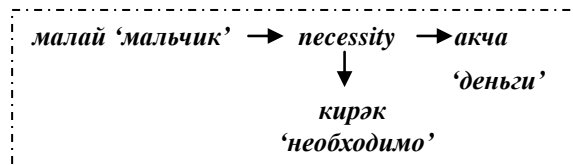


В реляционно-ситуационной системе рассматриваемая словоформа заполняет следующую ролевую ситуацию: **Отношение Действие:** Физическое восприятие: Объект восприятия некоторой ситуации.

Значение 10. Зависимое слово с аффиксом -ГА является субъектом, к которому обращена модальность.



малайга акча кирәк 'мальчику нужны деньги'



В реляционно-ситуационной системе рассматриваемая словоформа заполняет следующую ролевую ситуацию: **Отношение Модальность: Отношение необходимости.**

ЗАКЛЮЧЕНИЕ

Семантические универсалии, описанные в данной статье для представления значений морфем, при их дальнейшем развитии могут быть использованы для семантического описания и более сложных лингвистических конструкций.

Главная идея заключается в том, чтобы создать целостную реляционно-ситуационную систему, представляющую совокупность семантических универсалий, которые могут быть использованы как инвентарная база для формального кодирования семантических ситуаций в картине мира и возможного их дальнейшего описания на некотором естественном языке. Очевидно, семантические сценарии, составленные из таких языко- и предметно-независимых универсалий могут стать хорошей базой для создания интеллектуальных программ, таких как программы машинного перевода, извлечения знаний и машинного поиска.

Поддержан грантом РФФИ № 12-07-00615-а «Модели и методы реляционно-ситуационного анализа текстов тюркской группы языков для задач семантического поиска».

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

- [Мартьянов, 1984] Мартьянов В.В. Универсальный семантический код: УСК-3. – Минск: «Наука и техника». – 1984. – 131с.
- [Лезин и др., 1996] Лезин Г.В., Боярский К.К., Попов А.И. Систематизация информации: семантическое кодирование текстов // Труды межд. семинара ДИАЛОГ-96: компьютерная лингвистика и ее приложения (Пушино, 4-9 мая, 1996г.). – М., 1996. – С.131-136.
- [Шенк, 1980] Шенк Р. Обработка концептуальной информации. – М.: Энергия, 1980. – 361 с.
- [Филлмор, 1981] Филлмор Ч. Дело о падеже // В кн.: Новое в зарубежной лингвистике. Вып. X. Лингвистическая семантика. М.: Прогресс, 1981.
- [Апресян и др., 1992] Апресян Ю.Д., Богуславский И.М., Иомдин Л.Л. и др. Лингвистический процессор для сложных информационных систем. – М.: Наука, 1992. – 256 с.

[Бухараев и др., 1990] Бухараев Р.Г., Сулейманов Д.Ш. Семантический анализ в вопросно-ответных системах Казань: Изд-во Казан. ун-та. – 1990. – 124 с.

[Поспелов, 1981] Поспелов Д.А. Логико-лингвистические модели в системах управления. – М.: Энергоиздат, 1981. – 232 с.

[Осипов и др., 2006] Осипов Г.С., Тихомиров И.А., Смирнов И.В. Eхactus – 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.

[Закриев, 1992] Закриев М.З. Татарская грамматика. Т3. Синтаксис. – Казань: Таткнигоиздат. 1992. – 488 с.

[Сулейманов, 1998] Сулейманов Обработка ЕЯ-текстов на основе прагматически-ориентированных лингвистических моделей Сб.под ред. Соловьева В.Д.: Обработка текста и когнитивные технологии. – Вып.3. Труды научного семинара «Когнитивное моделирование» (Пушино, октябрь 1998 г.). – С.205-212.

[Труды, 1996] Труды Международного семинара Диалог-96: компьютерная лингвистика и ее приложения (Пушино, 4-9 мая, 1996г.). Под ред. А.С.Нариньяни. – М.: 1996. – 305 с.

[Casey и др., 1992] Casey, Nagy G. 1992. At the frontiers of OCR. Proceedings of the IEEE, (80(7):1093-1100.). – 1992.

[Pulman, 1995] Pulman Stephan G. Semantics // In Cole Ronald A., Mariani Joseph, Uszkoreit Hans, et al (editors). Survey of the State of the Art Human Language Technology. – 1995/. – P.122-129 / ftp://speech.cse.ogi.edu/pub/docs/HLT/.

[Сулейманов и др., 2003] Сулейманов Д.Ш., Гатиатуллин А.Р. Структурно-функциональная компьютерная модель татарских морфем. – Казань: Фэн, 2003. – 220с.

[Моделирование, 1987] Моделирование языковой деятельности в интеллектуальных системах. – М., Наука. Главная ред. физ.-мат.лит., 1987. – 280 с. /Под ред. А.Е. Кибрика А.С. Нариньяни; С предисловием А.П. Ершова.

SEMANTIC UNIVERSALS IN THE SYSTEM OF DESCRIPTION OF MORPHEME MEANINGS

Suleymanov D.Sh., Gatiatullin A.R.

«Scientific Research Institute of Applied Semiotics» of Academy of Sciences of the Republic of Tatarstan,
Kazan (Volga region) Federal University,
Kazan, Russia

dvd.t.slt@gmail.com

agat1972@mail.ru

The article offers a description of semantic universals which represent semantic scenarios and form in total an object-predicate system. The object-predicate system is aimed at describing the meanings of affixal morphemes presented in the structural-functional morphemic model. Semantic universals are studied upon the example of description of the meanings of a case morpheme -GA (directive).

Key words: semantic universals, morphemic model, relation-situation system, semantic aspect