

# Онтология силовых процессов

## An ontology of force interactions

**Кобозева И. М.** (kobozeva@list.ru),  
**Марушкина А. С.** (nastam@rambler.ru)

МГУ им. М. В. Ломоносова

В статье предпринимается попытка показать, как созданная на основе теории Динамики Сил онтология силовых процессов может быть использована для такого семантического представления текста на естественном языке, которое может служить базой для естественно-логического вывода о результатах и дальнейшем развитии описанных в нем ситуаций силового взаимодействия объектов. Семантическое представление текста задается, с одной стороны, в виде сценария, учитывающего основные силовые переходы, а с другой — в виде формализованного лексического описания.

### 1. Границы предметной области

Разработка коммерческих онтологий на данный момент сводится в основном к представлению четко очерченной и хорошо структурированной предметной области, например, онтологии медицинских препаратов или молекулярных функций, вин и продуктов питания и пр. Отношения между концептами в такой онтологии являются достаточно регулярными, а вопрос о границах предметной области не представляет особой сложности.

В случае создания универсальной онтологии, предназначенной для моделирования смысла естественно-языковых текстов по любой тематике на том уровне понимания, который отражает представления о мире рядового носителя языка, этот вопрос, безусловно, не является тривиальным, потому что мы имеем дело с бесконечно сложно организованным объектом моделирования. Во-первых, характер связей между концептами может быть самый разный, во-вторых, в семантике слов семы, соответствующие концептам онтологии, могут обладать разной степенью выраженности.

Цель данной работы — показать вариант анализа текста, описывающего силовые процессы, с использованием фрагмента универсальной онтологии. Организующим звеном для его построения является понятие Динамики сил, введенное в лингвистический обиход Леонардом Талми (см. [6]).

Следуя определению, которое Талми дает категории Динамики сил, обозначим границы предметной области следующим образом: **мы рассматриваем любые концепты, отражающие изменение состояния физических объектов в процессе сило-**

**вого взаимодействия.** Причем силовые процессы могут быть любого типа: это и действие силы, и сила противодействия (resistance to force), преодоление силы противодействия, блокирование силового воздействия (blockage of the expression of force), снятие блокирования и т. д.

Предлагаемая нами онтология в отношении формальной репрезентации семантической структуры концептов и лексических значений близка к универсальной онтологии «Онтологической семантики» С. Ниренбурга и В. Раскина [5], и поэтому может рассматриваться как дополнительный модуль этой онтологии, подключаемый при необходимости более точного «вычисления» смысла описываемых в тексте ситуаций силового взаимодействия объектов. На данном этапе она позволяет проинтерпретировать в аспекте динамики сил только тот пласт лексики (в основном глагольной), представители которого имеют в своем составе сему «сила», соответствующую онтологическому концепту СИЛЫ (FORCE) как физической величины. При этом пока не учитывается вклад граммом вида и залога в смысл языковых описаний ситуаций силового взаимодействия, а также метафорические употребления рассматриваемых единиц, хотя можно предположить, что структура выделенных концептов может помочь определить нетривиальные основания для метафорического переноса.

Особенность онтологии, построенной с помощью категории Динамики Сил, заключается в том, что с ее помощью можно учесть те аспекты означаемого, которые важны для естественно-логического вывода по тексту (иногда их рассматривают как следствия или импликации [2], [4]).

## 2. Онтология, лексикон и семантическое представление текста

На данном этапе мы выделяем два модуля, задействованных при семантическом анализе текста. Это онтология и лексикон.

Процедура обработки текста представляет собой процесс, протекающий на разных уровнях. В рамках данного исследования разрабатывается подход к построению семантического представления текста без привязки семантического блока к другим блокам обработки текста (морфологическому и синтаксическому парсерам).

Онтология Динамики Сил представляет собой набор концептов, репрезентирующих внутреннюю динамику силовых процессов, закономерности их протекания и позволяющих логически связать несколько примитивных ситуаций в единый сценарий.

Лексикон позволяет учесть дополнительные параметры ситуаций, которые будут влиять на выбор той или иной лексической единицы при описании ситуации.

## 3. Онтологический концепт и его структура

Основное отличие разрабатываемого нами фрагмента онтологии от других доступных онтологий состоит во внутреннем устройстве концептов и характере отношений между ними. Во многих лексических ресурсах, концептуализация законов силовых взаимодействий отражена номинативно, то есть внутренняя структура соответствующих концептов в полной мере не описана. Это происходит даже в тех случаях, когда создатели ресурса делают успешную попытку моделировать концептуализацию силовых процессов не в научной, а в языковой картине мира и делают это достаточно формально (см. [5], [7], [8], [9]). Как было показано раньше в работе одного из авторов данного исследования (см. [3]), языковая концептуализация силовых взаимодействий может быть представлена в терминах *тенденций* и *соотношений*, приводящих к определенному результату. На основе предложенных Талми концептуальных схем мы выделили 10 типов базовых ситуаций, которые отражают наивное представление о законах физической механики. Таким образом, в нашем распоряжении оказывается 10 онтологических концептов. К сожалению, описание каждого из них достаточно трудоемко, но поскольку для понимания устройства онтологии достаточно объяснить базовые принципы построения концептов, мы этим и ограничимся (для более подробного ознакомления со структурой каждого концепта отсылаем читателя к работе [3]). Каждый выделенный тип силового взаимодействия представлен в виде

одной или нескольких схем, структурными элементами которых являются:

- Участники ситуации: антагонист (Ant) — воздействующий, и агонист (Ago) — претерпевающий воздействие
- Внутренняя тенденция участников: импульс к действию (toward action) — тенденция к покою (toward rest)
- Соотношение сил: более сильный (stronger entity) — менее сильный (weaker entity)
- Результат взаимодействия: действие (action)<sup>1</sup> / движение (motion) — покой (rest)

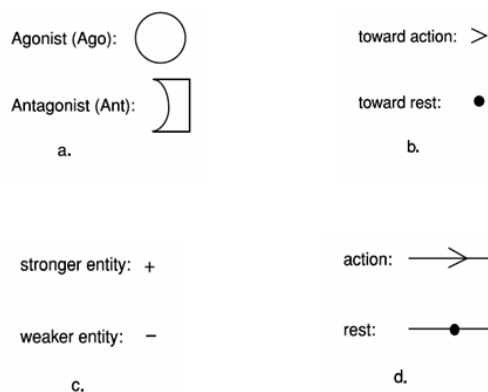


Рис. 1

В виде сценарной схемы силовых процессов представим следующее предложение в его «силовой» интерпретации (другие возможные для него интерпретации в рамках данного исследования игнорируются):

(87) Дверь не открывается

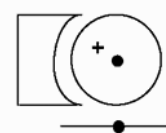


Рис. 2

Схема показывает, что на дверь воздействует некоторая сила, которой недостаточно, чтобы дверь открыть, поэтому результатом взаимодействия оказывается состояние покоя.

Не все концепты, однако, описываются одной схемой — некоторые имеют более сложную внутреннюю логическую структуру. Для примера рас-

<sup>1</sup> В рамках данной статьи не будем разграничивать понятия *движение (motion)* и *деформация (deform)*, состоянию *покой (rest)* будем противопоставлять *изменение (action)*, под которым следует понимать дизъюнкцию (*движение ^ деформация*).

смотрим глагол со значением импульсного воздействия *толкать*. Успешность воздействия неопределена, а значит, и сценарное описание данного силового процесса не может быть однозначным. По этой причине мы будем описывать соответствующий ему концепт через дизъюнкцию, которая показывает, что движение (деформация) может не последовать, если силы воздействующего для этого не достаточно:

- I. 1. Участники ситуации:  
 ANT (SEM (PHYSICAL\_OBJECT))  
 AGO (SEM (PHYSICAL\_OBJECT))
2. Внутренняя силовая тенденция:  
 (HAS\_TENDENCY (SEM ({ACTION, REST, MOTION\DEFORM})))
- II. 3. Соотношение сил:  
 FORCE\_BALANCE (SEM ({UNKNOWN, LETTING, STRONGER (SEM (ANT)),  
 STRONGER (SEM (AGO))}))
- III. 4. Начальное состояние Агониста:  
 FD\_INITIAL (SEM ({REST, MOTION\DEFORM}))
5. Результат взаимодействия:  
 FD\_RESULT (SEM ({UNKNOWN, ACTION, MOTION\DEFORM, REST}))

Необходимо оговориться, что выбор фасетов — наследие терминологии принятой в [5]. В фигурных скобках через запятую перечислены концепты, составляющие область значений слота. За первыми двумя слотами на данном этапе закреплено значение PHYSICAL\_OBJECT (физический объект) — метафоры мы пока не рассматриваем.

По законам Динамики Сил Агонист и Антагонист должны стремиться либо к изменению (движение, деформация), либо к покою. Это отражено в слоте HAS-TENDENCY. При этом тенденции Антагониста и Агониста в рамках одной схемы не могут совпадать, поэтому достаточно задать только тенденцию Агониста. Значение слота HAS-TENDENCY для Антагониста будет определено как обратное. Для фиксации этой закономерности в онтологии используем слот OPPOSITE.

Для обозначения соотношения сил между основными участниками внутри концепта вводим четыре значения слота FORCE\_BALANCE: STRONGER (SEM (ANT)), STRONGER (SEM (AGO)), LETTING, UNKNOWN. Значение LETTING соответствует снятию силового воздействия, которое Антагонист оказывает на Агониста, а значение UNKNOWN показывает, что в данном концепте соотношение сил не определено.

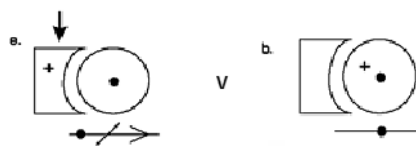


Рис. 3

Структурные элементы схем формально представлены в виде слотов. Набор слотов фиксирован для каждого из концептов и выглядеть он будет следующим образом:

Значения слота FD\_RESULT (результат взаимодействия) показывают конечное состояние Агониста (изменение или покой).

Осталось фиксировать только динамический контур ситуации — постоянный или переменный (то есть, меняется ли начальное состояние Агониста или нет). Для этого целесообразно ввести еще один слот, который будет отражать начальное состояние Агониста, а вместе с ним и системы объектов в целом. Такой слот назовем FD\_INITIAL. Если его значение одновременно совпадает со значением HAS-TENDENCY для Агониста и не совпадает со значением слота FD\_RESULT, мы имеем дело с ситуацией силового взаимодействия с **переменным** динамическим контуром.

#### 4. Создание лексикона и формата семантического представления анализируемого текста

Теперь на основе описанных выше концептов можно начать разработку лексических описаний для создания словаря. Предложенная структура лек-

сического входа разработана для описания лексических единиц русского языка с семантикой силового взаимодействия.

#### 4.1. Структура шаблона для лексического описания

Как было отмечено ранее, семантическое представление текста имеет двойной формат: с одной стороны, это сценарное описание с привлечением концептов онтологии, с другой — описание в формате словаря. В этом разделе рассмотрим шаблон лексического входа для единиц с силовым значением.

В первую очередь рассмотрим свойства Антагониста (воздействующего). Пока здесь выделены только два слота: это *одушевленность / неодушевленность* и характеристика *лицо / не-лицо*. Важно сразу сказать, что эти два параметра отвечают за разные области. *Одушевленность* во многом определяет тип Агониста (такие его характеристики, как, например, манипулируемость), а также распределение сил между Агонистом и Антагонистом. Эта характеристика описывает физическую сферу. *Лицо / не-лицо* — параметр, главным образом определяющий импликации, которые станут основой для дальнейшего семантического расширения. С ним неразрывно связано понятие цели действия (ср. глаголы *бросить* и *выбросить*). От того, является ли Антагонист лицом, могут также зависеть и такие существенные вещи, как выбор предлога. Например, *тащить* что-либо можно *за* собой и *на* себе, только последний предлог предпочтительнее именно в том случае, когда Антагонист — лицо.

В работе [1] также отмечается, что идея агенса напрямую связывается с идеей воли и цели. При этом акцентируется семантическая сопряженность с этой идеей значений граммы несовершенного вида и в качестве подтверждения этому наблюдению приводятся деструктивные глаголы (*рвать*, *разбивать*, *ломать* и др.), которые не имеют актуально-длительного значения, когда человек выступает не в роли агенса, а в роли «невольной причины» разрушения: ср.

(88) Его толкают, и он разбивает окно,

где деструктивное действие не намеренное, и соответствующие глаголы употреблены нарративно, и

(89) Посмотри-ка, они уже и витрины разбивают,

где агенс уже действует целенаправленно, и реализуется актуально-длительное значение глагола НСВ (примеры взяты из [1]).

В общую концептуальную структуру также вводится описание физических и топологических свойств Антагониста.

Описание Агониста (претерпевающего воздействие) устроено похожим образом.

Остальные слоты являются факультативными. Это значит, что их наличие в той или иной степени зависит от онтологического класса. Так, например, не всегда необходимо исчерпывающее описание физических параметров Антагониста (таких как размер и вес). Для онтологических концептов, которые фиксируют соотношение сил между Антагонистом и Агонистом, к которым, например, относятся глаголы *двигать*, *держать*, описание размера и веса обычно необязательно.

Еще один факультативный параметр — форма. Форма Агониста играет решающую роль при выборе лексической единицы в ограниченном числе случаев. Самый яркий пример — предикат *катить*, который в общем случае предполагает, что Агонист круглый / шарообразный.

Далее задается описание свойств материала, из которого сделан Агонист (в случае, если он неодушевленный). Этот раздел лексического описания важен для тех ситуаций, которые потенциально влекут за собой деформацию Агониста. В других случаях эти параметры также можно опустить.

Для ряда случаев (а именно, для ситуаций с переменным динамическим контуром, где Антагонист препятствует движению Агониста) также важно указывать, с какой скоростью Агонист движется. На данном этапе мы предлагаем три значения для этого слота:

VELOCITY (high, medium, low).

От типов объектов перейдем к характеристике действия силы. Здесь мы имеем принципиальные различия в описании для концептов деформации и концептов движения.

Мы выделяем несколько направлений приложения силы: это направления вверх-вниз, вправо-влево, а также четыре диагональных направления. Вектор приложения силы и наклон поверхности, по которой происходит движение, взаимозависимые величины.

Как мы уже говорили, внутренний семантический потенциал предикатов силового взаимодействия может быть разным: неодинаково распределение значимости между признаками, характеризующими ситуацию, значение одного и того же признака в одних случаях выводится из контекста, в других — оно будет фиксированным. Так, на примере глаголов сопротивления хорошо видно, что в семантической структуре этих глаголов присутствует дополнительный признак «точка зрения» или внутренний фокус эмпатии. Возьмем пару глаголов *опираться* vs *подпирать*. Практически идентичная ситуация дается сначала с точки зрения Агониста, потом — с точки зрения Антагониста. Это явление отражено в слоте FOCUS\_VALUE.

Остальные три слота, описывающие природу приложения силы, универсальны. Интенсивность приложения силы определяется по шкале от 0 до 1. От того, какое значение получает данный слот в том или ином случае, может зависеть выбор лексемы (ср. *стукнуть* — *шарахнуть*).

Сила может действовать на Антагониста постоянно или импульсно (ср. *тащить vs. толкать*). Эта характеристика отражена в бинарном слоте APPLICATION (*permanent, impulse*).

Также выделен слот CONTACT\_PATH, с помощью которого кодируется тип контакта Агониста с поверхностью, по которой происходит движение — может иметь место либо касание только одной стороной (*single\_side*), как для глаголов *двигать, тащить*, либо касание по всей поверхности Агониста (*comprehensive*), как для глаголов *катить, перекачивать*. Следующий слот заполняется только при значении предыдущего — *comprehensive*, то есть контакт по всей поверхности Агониста. Это слот PROMPTNESS, или количество оборотов, со-

вершенных Антагонистом, каковое представляется в виде численного значения.

Наконец, последний слот в этом разделе — CONTACT\_ANT — описывает опосредованность контакта между Антагонистом и Агонистом, т. е. специфицирует, подразумевается ли в данной ситуации дополнительный инструмент воздействия. Его значения: *direct* (непосредственный) и *instr* (опосредованный). При выборе второго значения необходимо указать, какой именно инструмент использован при опосредованном контакте. Для этого существует слот INSTR\_TYPE, областью значения которого по умолчанию является множество объектов, являющихся артефактами.

В последней части описания описано взаимодействие Агониста с окружающей средой (чаще всего — поверхностью, по которой происходит или из-за физических свойств которой не происходит движение). Сюда же как внешний признак попадет траектория движения Агониста.

### Структура лексического описания

#### [1] Participant Type

##### [1.1] Antagonist

```
[Ant_type ANIM(anim / inanim)
  PERS (pers / non-pers)
  VOLITIONAL (+,-) {намеренно/ненамеренно}]
  PHYSICAL_TYPE
```

##### [1.2] Agonist

```
[Ago_type
  ANIM(anim / inanim)
  PHYSICAL_TYPE
  SHAPE(shape)
  SIZE (0<>1)
  WEIGHT (0<>1)
  MATERIAL (x, (
    ELASTICITY (0<>1))
    SOLIDITY (0<>1))
  ]
  {В этом разделе в описание попадает ограниченное количество
  параметров: акцент может делаться как на материале (одной
  или нескольких характеристиках), так и на форме}
```

#### [2] Focus

```
[Focus_Value (Ant, Ago)]
```

#### [3] Force Application

```
[Force_Application
  VECTOR (towards_Ant, from_Ant, down, up, towards_Ant_down,
    from_Ant_down, towards_Ant_up, from_Ant_up)
  INTENSITY (0<>1)
  APPLICATION (permanent, impulse)
  CONTACT_ANT (direct, instr)
  INSTR_TYPE (artifact)
  CONTACT_PATH (comprehensive, single-side) — {контакт
    с поверхностью: по всей поверхности агониста
    или точка соприкосновения с пов-тью фиксирована}
```

```
PROMPTNESS (n >= 0) {число оборотов}}
[4] Environment
[4.1] Ago_trajectory
    [Ago_trajectory
        TRAJECTORY (default {= vector value}, circled, wigwag{туда-
        сюда}, oscillation {вращения}, stop)
        TRAJECTORY_LENGTH (n >=0)]
        VELOCITY (high, medium, low, n >=0) {скорость движения Агониста}}]
[4.2] Environment
    [Path/Environment
        ROUGHNESS (0<>1) {шероховатость}
        INSTABILITY (+,-) {нестабильность}
        BIAS (towards_Ant_down (0<>1), from_Ant_down (0<>1),
            towards_Ant_up (0<>1), from_Ant_up (0<>1))]
```

### 4.2. Пример построения семантического представления текста

Попробуем представить общий вариант анализа входящего письменного текста.

Для простоты возьмем текст задачи из сборника задач по физике для 7–9 классов. Употребление лексических единиц с силовой семантикой здесь, как правило, не противоречит обыденному употреблению, а каждый из текстов представляет собой некоторую, обычно завершённую, цепь сменяющих друг друга силовых переходов, для которых удобно показать, как на практике выглядит их последовательный анализ.

Обработка текстов задач существенно отличается от обработки любого другого вида текста только в двух взаимосвязанных аспектах. Во-первых, с дискурсивной точки зрения задачи имеют четкую двухчастную структуру: основную часть и вопрос. Причем в основной части всегда дается развернутый сценарий, а при формулировании вопроса либо дается дополнительная информация для расчетов, либо альтернативные условия протекания силовых процессов. Во-вторых, на уровне конечного семантического представления текста будет заполнено значительно большее количество слотов, чем у обычного текста.

Рассмотрим такой пример:

(90) Мяч, который уронили с некоторой высоты Н в неподвижном лифте, подпрыгивает на высоту h. Изменится ли эта высота, если лифт равномерно движется навстречу уроненному в нём с той же высоты мячу?

Мы делим текст задачи на две части по границе между предложениями. Сначала система обращается к словарю, с помощью которого определяются лексические единицы со значением силовых процессов. Таким образом обнаруживаются два предиката: *уронили* и *подпрыгивает* и концепты, которым они соответствуют. На основе концептов строим сцена-

рий, который графически будет выглядеть следующим образом (формальное описание концептов, состоящее из слотов, представленных в разделе 3, полностью соответствует графическому, но является довольно громоздким, поэтому мы не будем приводить его в рамках данной работы):

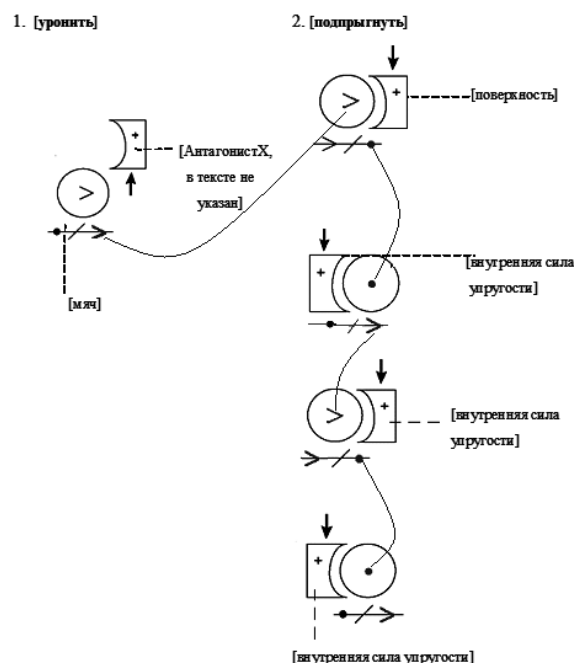


Рис. 4

Глагол *подпрыгивать* (с неодушевленным Агонистом) подразумевает наличие упругого удара, что отражено конъюнкцией сразу четырех концептуальных схем. Причем в двух из них в роли Антагониста выступает физическая сила — внутренняя сила упругости (далее INNER\_ELASIC\_FORCE).

Теперь на основе лексических входов для этих глаголов начинается построение семантического представления текста задачи, ориентированного на отражение силовой динамики.

«Уронить»

**Antagonist** [X]

```
[Ant_type ANIM(anim)
    PERS (pers)
    VOLITIONAL(-)]
```

**Agonist<sub>1</sub>** [«мяч»]<sup>2</sup>

```
[Ago_type
    ANIM(inanim)
    PHYSICAL_TYPE
    SHAPE(round)
    SIZE (0<>1)
    WEIGHT (0<>1)
    MATERIAL (x, (
        ELASTICITY (0.4<>0.5))
        SOLIDITY (0.8<>0.9))
    VELOCITY (high, medium, low or n >=0)]
```

**Focus**

[Focus\_Value (Ant)]

**Force Application**

```
[Force_Application
    VECTOR (down)
    CONTACT_ANT (direct)
]
```

**Ago\_trajectory**

```
[Ago_trajectory
    TRAJECTORY (default [= vector value])
    TRAJECTORY_LENGTH (n >=0)]
```

**Environment** [«лифт»]<sub>E</sub> {нижний индекс позволяет показать, что лифт выступает в качестве окружающей среды в данном сценарии, но в следующем его роль меняется, и поверхность лифта будет играть уже роль Антагониста}

```
[Path/Environment
    ROUGHNESS(0<>1)
    INSTABILITY (-)]
```

«Подпрыгнуть»

**Antagonist**

[Ant\_type<sub>1</sub> SURFACE<sub>E</sub> Ant\_type<sub>2</sub> INNER\_ELASTIC\_FORCE Ant\_type<sub>3</sub> INNER\_ELASTIC\_FORCE]

**Agonist<sub>1</sub>** [«мяч»]

**Focus**

[Focus\_Value (Ago)]

**Force Application**

```
[Force_Application1
    VECTOR1 (down_squeeze)
    CONTACT_ANT1 (direct)
Force_Application2
    VECTOR2 (up)
    CONTACT_ANT2 (direct)
Force_Application3
    VECTOR3 (up)
    CONTACT_ANT3 (direct)
]
```

**Ago\_trajectory<sub>2</sub>**

```
[TRAJECTORY (stop)]
```

<sup>2</sup> Здесь и далее физические свойства Агониста (кроме свойства VELOCITY), когда он назван в тексте, по умолчанию берутся из других концептов. Так, например, свойства Агониста мяч взяты из соответствующего концепта, присутствующего в универсальной онтологии.

```

TRAJECTORY_LENGTH (n >=0) ]
Ago_trajectory,
  [TRAJECTORY (vector3 value)
  TRAJECTORY_LENGTH (H) ]
Environment [«лифт»]
  [Path/Environment
  ROUGHNESS (0)
  INSTABILITY (-)
  ]

```

Поскольку в вопросе к задаче изменяется статус стабильности окружающей среды, в альтернативном представлении поменяется значение слота INSTABILITY в разделе Environment с «-» на «+»

## 5. Заключение

В данной статье мы попытались показать, как можно построить семантическое представление текста с использованием онтологии силовых про-

цессов. Конечно, в ряде случаев пришлось огрублять и упрощать предложенные описания, кроме того, для качественного анализа текста еще предстоит провести большую работу. Однако основные принципы применения данной онтологии при обработке текста на естественном языке уже достаточно четко прослеживаются. Получаемое формальное сценарное представление, отражающее динамику силовых процессов, обозначаемых глаголами, может служить входом для правил естественно-логического вывода о тенденциях и результатах описанных в тексте ситуаций взаимодействия физических объектов.

## Литература

1. Апресян В. Ю., Апресян Ю. Д., Бабаева Е. Э., Богуславская О. Ю., Иомдин Б. Л., Крылова Т. В., Левонтина И. Б., Санников А. В., Урысон Е. В.; Отв. ред. Апресян Ю. Д. Русская языковая картина мира и системная лексикография // М.: Языки славянских культур, 2006
2. Кустова Г. И. Типы производных значений и механизмы языкового расширения. М.: Языки славянской культуры. 2004
3. Марушкина А. С. «Наивная механика» в языке и онтологии // Сборник трудов международной конференции Диалог, 2006
4. Падучева Е. В. Динамические модели в семантике лексики. М.: Языки славянской культуры, 2004
5. Nirenburg S., Raskin V. Ontological semantics // Cambridge.: MIT press, 2004.
6. Talmy L. Toward a cognitive semantics. Vol. 1 // Cambridge.: MIT press, 2000.
7. [www.ontologyportal.org](http://www.ontologyportal.org) (SUMO)
8. [www.openency.org](http://www.openency.org) (CYC)
9. <http://wordnet.princeton.edu> (WordNet)