

На правах рукописи

МАРУШКИНА АНАСТАСИЯ СЕРГЕЕВНА

**СЕМАНТИКА ЛЕКСИКИ, ОБОЗНАЧАЮЩЕЙ ФИЗИЧЕСКИЕ
ВЗАИМОДЕЙСТВИЯ, В ОНТОЛОГИЧЕСКОМ АСПЕКТЕ**

Специальность 10.02.21 – прикладная и математическая лингвистика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата филологических наук

Москва – 2012

Работа выполнена на кафедре теоретической и прикладной лингвистики филологического факультета ФГОУ ВПО «Московский государственный университет имени М.В. Ломоносова».

Научный руководитель:

доктор филологических наук,
профессор
Кобозева Ирина Михайловна

Официальные оппоненты:

Рябцева Надежда Константиновна
доктор филологических наук,
зав. Сектором прикладного языкознания
ФГБУН «Институт языкознания РАН»,

Сичинава Дмитрий Владимирович
кандидат филологических наук
старший научный сотрудник
ФГБУН «Институт русского языка
им. В. В. Виноградова РАН»,

Ведущая организация:

ФГОУ ВПО «Российский государственный
гуманитарный университет»

Защита состоится 30 мая 2012 года в 14.30 часов на заседании диссертационного совета Д 501.001.24 при ФГОУ ВПО «Московский государственный университет имени М. В. Ломоносова» по адресу: 119991 ГСП-1, Москва, Ленинские горы, МГУ, 1-й учебный корпус, филологический факультет.

С диссертацией можно ознакомиться в библиотеке 1-го учебного корпуса ФГОУ ВПО «Московский государственный университет имени М.В. Ломоносова»

Автореферат разослан 27 апреля 2012 года.

Ученый секретарь
диссертационного совета

А.М. Белов

Актуальность диссертационного исследования обусловлена тем, что в последнее время растет популярность использования онтологий для представления знаний. В области искусственного интеллекта онтологии обеспечивают поддержку разнообразных экспертных систем, требующих построения логического вывода или принятия решений. Кроме того, появляется все больше проектов в русле концепции Semantic Web, так как недостаток семантизации и доступности Web-контента в Интернет остается на данный момент серьезной проблемой. Среди огромного количества онтологий можно назвать такие известные доступные в сети Интернет ресурсы, как CYC, SUMO, Wordnet, онтология Дж. Сóвы и др. В России заметные разработки подобного рода ведутся в Москве (АНО ЦИИ, «Авикомп сервисез»), Санкт-Петербурге (коллектив под руководством Т.А. Гавриловой), Новосибирске (коллективы под руководством Ю.А. Загорулько и Н.Г. Загоруйко), а также в Казани и Владивостоке (см. обзор в [Ена, Ефименко, Хорошевский 2011]).

Одним из направлений в проектировании онтологий является разработка ресурсов для автоматизированной обработки текстов на естественном языке, которые позволяют построить семантическое представление текста с учетом содержащихся в нем импликаций [Леонтьева 2006, Лукашевич 2011]. Такие онтологии моделируют общие понятия и реальный мир в его многообразии и состоят из неформально заданных концептов в виде словарных толкований описательного характера. Их называют вербальными (verbal ontologies) или универсальными (universal ontologies). Естественный язык как метаязык дает достаточную свободу и гибкость описания, но одновременно снижает четкость и однозначность. Часто концепты обнаруживают зоны пересечения на предметной области, и случается, что в силу расплывчатости определения затруднен выбор концепта при разработке лексической статьи. В данном исследовании мы предприняли попытку более подробно рассмотреть ограниченную предметную область – область силовых взаимодействий – и избежать неоднозначности в ее описании. Выбор предметной области, в частности, продиктован тем, что характер силовых взаимодействий между физическими объектами до сих пор не был отдельным объектом рассмотрения при создании универсальных онтологий (так, например, в онтологии, разработанной С. Ниренбургом и В. Раскиным, для

описания силовых процессов существует недостаточно детально разработанный концепт FORCE-APPLICATION).

Учитывая вышесказанное, обозначим предмет исследования: им является разработка концептуального представления силовых процессов в их отношении к наивной языковой картине мира.

Объектом исследования являются языковые единицы (прежде всего глаголы), в семантическую структуру которых входят семы 'сила', 'взаимодействие', 'изменение', 'каузация'. В рамках диссертационного исследования рассматриваются только прямые пространственно-физические значения этих единиц, хотя в дальнейшем полученные результаты могут быть использованы для описания гораздо более широкой сферы взаимодействий. В идеографическом словаре О.С. Баранова такие единицы попадают в самые разные разделы: ВЗАИМОДЕЙСТВИЕ СИСТЕМ, ДВИЖЕНИЕ МАТЕРИИ, ТВЕРДОЕ ТЕЛО, МЕХАНИЧЕСКИЕ КОЛЕБАНИЯ, ПЕРЕДВИЖЕНИЕ, ДИНАМИКА ДВИЖЕНИЯ и пр.

Цель исследования – построение фрагмента универсальной онтологии, описывающего силовые процессы между физическими объектами, разработка формата лексического описания и, наконец, определение основных принципов для построения семантического представления текста, описывающего силовые процессы разной природы.

В соответствии с целью исследования обозначим конкретные задачи, которые должны быть решены в ходе данной работы:

1. обозначить и исследовать основные типы силового взаимодействия;
2. разработать средства их формального описания;
3. составить базовую классификацию предикатов силового взаимодействия;
4. выделить структуру концептов, составляющих фрагмент универсальной онтологии;
5. построить фрагмент универсальной онтологии сферы силовых взаимодействий;
6. создать формат лексического описания языковых единиц, обладающих силовой семантикой;

7. разработать алгоритм построения семантического представления входящего текста на базе онтологии и лексикона.

Вышеперечисленные задачи последовательно решаются в главах диссертации.

Материалом для исследования послужили данные словарей разных типов, таких, как «Семантический словарь русского языка» под общей редакцией Н.Ю. Шведовой, «Словарь русского языка» под редакцией А.П. Евгеньевой (МАС), «Идеографический словарь русского языка» (О.С. Баранов), «Русский ассоциативный словарь» (Ю.Н. Караулов, Г.А. Черкасова, Н.В. Уфимцева, Ю.А. Сорокин, В.Н. Ярошинская.), «Русский семантический словарь: опыт автоматического построения тезауруса: от понятия к слову» (Ю.Н. Караулов, В.И. Молчанов, В.А. Афанасьев, Н.В. Михалев), толковые словари русского языка под редакцией С.И. Ожегова и Н.Ю. Шведовой, Д.Н. Ушакова, «Словарь синонимов и сходных по смыслу выражений» (Н. Абрамов), данные лексического ресурса WordNet, доступных в сети интернет онтологий (CYC, SUMO), а также онтологии, разработанной в рамках теории онтологической семантики [Nirenburg, Raskin 2004]. Кроме того, значительное количество проанализированных контекстов было получено благодаря работе с Национальным корпусом русского языка.

В качестве теоретической базы исследования была выбрана теория, разработанная в 1970-е годы американским лингвистом Леонардом Талми, который предложил для описания семантики естественного языка ставшую уже достаточно известной понятийную категорию **силового взаимодействия** (Force Dynamics). Как утверждает автор, она применима и при описании более широкой области лексики. В работах Талми это хорошо показано, в частности, на примере различных модальных глаголов. Следует также отметить, что изучение языковых способов кодирования силового взаимодействия продолжается на материале различных языков [Brandt 1992], [Lai & Chiang 2003], [Chun & David A Zubin 1990] и др. По нашему мнению, рассмотренная теория Л. Талми, несмотря на справедливую во многом критику в ее адрес [Goddard 1998], обладает достаточной объяснительной силой и будет полезна при описании языковых значений и деривационных отношений между лексическими единицами. Таким образом, именно она взята за основу в нашем исследовании.

Помимо теории Талми, мы будем учитывать и последний масштабный труд Е.В. Падучевой «Динамические модели в семантике лексики» [Падучева 2004]. Из представленных здесь тематических классов, выделенных для глагольной лексики, значительная часть в той или иной степени имеет отношение к понятиям «движение», «физическое воздействие», «перемещение объекта», «каузация», что представляет непосредственный интерес для предлагаемого описания. Определенные аналогии с теорией динамики сил прослеживаются в вышедшей в том же году работе Г.И. Кустовой «Типы производных значений и механизмы языкового расширения» [Кустова 2004] (см. главу 5 “Семантические модели производных значений глаголов”).

В работе были использованы следующие методы: корпусный метод для сбора данных, метод классификации, метод когнитивного моделирования значения лексических единиц, а также метод опроса информантов для верификации некоторых результатов, полученных путем интроспекции.

Научная новизна исследования обусловлена тем, что впервые сфера силовых взаимодействий стала объектом отдельного детального описания как часть универсальной лексической онтологии. На базе теории Динамики Сил Л. Талми был разработан новый принцип формального описания, ориентированный на проекцию данной предметной области в языке.

С привлечением языкового материала осуществлена классификация ситуаций силового взаимодействия и предикатов русского языка с силовой семантикой, которая стала основой для создания фрагмента универсальной онтологии. Онтология насчитывает **10** концептов.

Разработанный инструментарий позволил спроектировать онтологию силовых взаимодействий не в виде упорядоченного набора «строительных» элементов, а как единую концептуальную базу для создания лексических описаний или фактологических высказываний.

Помимо онтологической структуры в работе также предложен вариант лексического описания, которое учитывает основные роли и атрибуты ситуаций силового взаимодействия (условия протекания ситуации, топологические характеристики участников ситуации, направление действия силы и пр.).

Предложен общий алгоритм построения семантического представления текста на естественном языке. В результате обработки текста генерируется общий контур силовых переходов, а также значения атрибутов, дающих, насколько это возможно в каждом конкретном тексте, полное представление о ситуации.

Представленное в работе описание может быть адаптировано к широкому спектру приложений в зависимости от тех задач, которые они выполняют, что является его основным достоинством.

Теоретическая значимость состоит в том, что на основе существующих подходов к описанию семантики языковых единиц предикаты силового взаимодействия классифицируются как отдельная группа слов со своеобразным значением. Это дает возможность представить предметную область силовых взаимодействий в универсальной онтологии не как один или несколько отдельных концептов, а в качестве самостоятельного систематически устроенного сегмента. В свою очередь, использование чисто лингвистической теории при построении фрагмента онтологии способствует более точному построению семантического представления текста.

Практическая значимость исследования обусловлена прежде всего тем, что его результаты не нацелены на конкретный программный продукт. Это позволяет широко использовать данные результаты в разного рода автоматизированных приложениях для обработки текстов на естественном языке, особенно, если речь идет о специализированном дискурсе, для которого данная предметная область должна быть детально описана. Предложенное описание силовых значений как отдельной области, отличной от смежной области пространственных отношений и перемещений, можно использовать в дальнейших исследованиях по семантике и типологии.

Структура исследования. Структура работы соответствует целям и задачам исследования. Работа состоит из введения, трех глав, заключения и списка литературы.

Во **введении** определяются объект и предмет исследования, обосновывается актуальность, новизна, цели и задачи работы, методология, теоретическая и практическая значимость диссертации, а также описывается структура работы.

Первая глава («Наивная механика») посвящена исследованию предметной области силовых взаимодействий.

В первом параграфе дается описание теории Динамики сил, предложенной Леонардом Талми, которая детализирует общепринятое в лингвистике понятие каузации и является основополагающей для настоящего исследования. Представлены десять базовых сценариев силового взаимодействия, которые будут в дальнейшем использоваться нами при построении модели предметной области силовых взаимодействий. Основными структурными элементами сценариев силового взаимодействия, предложенных Талми, являются:

- Антагонист и Агонист (воздействующий – претерпевающий воздействие)¹
- импульс к действию / движению – тенденция к покою
- более сильный – менее сильный
- действие / движение – покой

Графически эти элементы выглядят следующим образом:

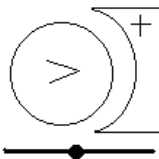


Далее проиллюстрирован сценарий ситуации силового взаимодействия, составленный из описанных выше элементов. В качестве примера рассматривается предложение

¹ Не следует путать такие понятия, как Агонист и Агенс. Описываемая терминология используется Талми исключительно для обозначения ролей участников ситуации силового взаимодействия.

Дверь не открывается.

Для него строится следующая сценарная диаграмма:



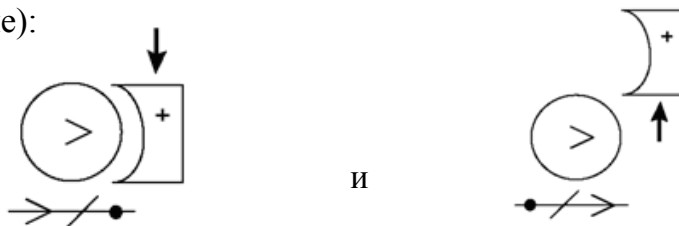
Она показывает, что дверь – Агонист, которому сообщен импульс к движению и на который действует достаточно слабый Антагонист, чтобы противодействие импульсу оказывалось сильнее; результат взаимодействия – покой. Такого рода диаграммы описывают ситуации «с постоянным динамическим контуром» (steady-state force dynamics patterns в терминологии Талми).

Другой тип ситуации – «с переменным динамическим контуром» (shifting force-dynamic pattern) – будет описываться такими фразами, как:

Он остановил машину,

Мальчик отпустил мяч.

Им соответствуют следующие диаграммы, которые показывают, что начальное состояние Агониста меняется (объект перестает двигаться под воздействием некоторой силы и наоборот – после того, как объект перестали удерживать, он пришел в движение):



Второй параграф содержит классификацию предикатов русского языка с силовой семантикой. Основу данной классификации формируют модели, предложенные Л. Талми. Предикаты объединяются в классы на основе их принадлежности к тому или иному динамическому сценарию, базирующемуся на трех основных параметрах: начальное состояние объекта, тип взаимодействия и

результат взаимодействия. Также существенным является потенциальный исход динамической ситуации.

В процессе создания нашей классификации мы старались прежде всего отразить имплицативный потенциал того или иного предиката с силовой семантикой. Для этого при описании некоторых классов мы решили объединять несколько сценариев развития ситуаций силового взаимодействия, предложенных Талми, в одну более сложную логическую структуру. Например, глаголы импульсного воздействия и давления (например, *толкать*, *ударять*, *давить* и др.) имеют следующую общую черту: в актуальном употреблении они могут обозначать ситуацию как с переменным, так и с постоянным динамическим контуром. В их значении заложен лишь потенциальный результат обозначенного ими действия. Мы предполагаем, что в итоге Агонист придет в состояние движения, но этот результат может так и не быть достигнут, ср.:

Потом ещё напряглись, толкали и упирались руками, подпирали камень плечами и спинами. Но как ни старались, он так и не сдвинулся с места.

Таким образом, вне контекста мы можем установить лишь соответствие данной ситуации статической или динамической модели с исходным состоянием покоя. Для фиксации результата необходима дополнительная информация.

На основе таких схематических обобщений выделяется 7 широких классов предикатов русского языка со значением силового взаимодействия.

Устанавливается, что четкой корреляции между видом и принадлежностью к тому или иному классу не существует.

В третьем параграфе осуществляется переход от классификации лексических единиц естественного языка к концептам онтологии; вводится еще один структурный элемент, не предусмотренный теорией Л. Талми, для обозначения деформации:

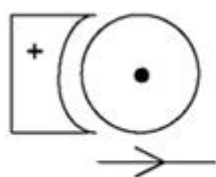


10 выделенных концептов формируются, с одной стороны, на основе всех возможных типов физического взаимодействия, присутствующих в сознании наивного носителя языка, а с другой стороны, на базе естественноязыковых значений.

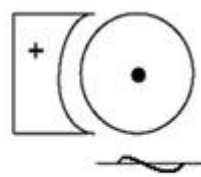
ТИПЫ СИТУАЦИЙ СИЛОВОГО ВЗАИМОДЕЙСТВИЯ²

1 Модель *a* в классификации Талми (концепт **Постоянного изменения**)

<p>[Антагонист (Ant): Субъект X Агонист (Ago): Объект Y</p> <p>Начальное состояние: Антагонист – тенденция к движению Агонист – тенденция к покою</p> <p>Тип взаимодействия: F (ant) > F (ago)</p> <p>Результат: движение]</p>	<p><i>двигать, катить, тащить, оттягивать, волочь, пропихивать, растягивать, и др.</i></p>
--	--

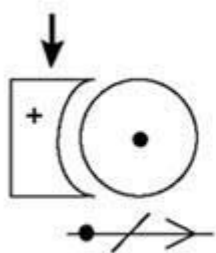


∨

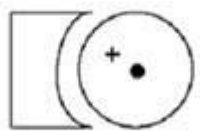


2 Модели (*b,e*) в классификации Талми (концепт **Порога силы воздействия**)

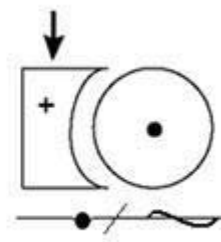
<p>[Антагонист (Ant): Субъект X Агонист (Ago): Объект Y</p> <p>Начальное состояние: Антагонист – тенденция к движению Агонист – тенденция к покою</p> <p>Тип взаимодействия: -</p> <p>Результат: -]</p>	<p><i>толкать, подталкивать, ударять, сопротивляться, давить, тянуть, подуть, растягивать, сжимать, рвануть, пихнуть, дернуть, стукнуть, садануть, шарахнуть и др.</i></p>
--	--



∨



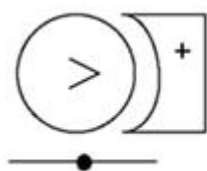
∨



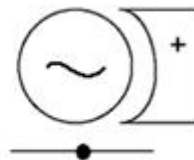
² знак «∨» соответствует логическому оператору «или», знак «^» соответствует логическому оператору «и»

3 Модель *d* в классификации Талми (концепт Сохранения устойчивости системы)

<p>[Антагонист (Ant): Субъект X Агонист (Ago): Объект Y</p> <p>Начальное состояние: Антагонист – тенденция к покою Агонист – тенденция к движению</p> <p>Тип взаимодействия: F (ago) < F (ant)</p> <p>Результат: покой]</p>	<p><i>держат, опираться, упираться, устоять, подпирать, поддерживать, сцеплять, скреплять (как статические явления), поддерживать, подпирать и др.</i></p>
---	--

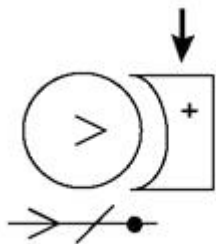


v

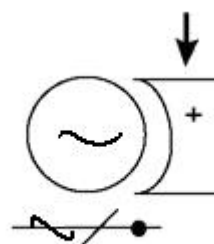


4 Модель *f* в классификации Талми (концепт Лишения подвижности)

<p>[Антагонист (Ant): Локализация X Агонист (Ago): Объект Y</p> <p>Начальное состояние: Антагонист – тенденция к покою Агонист – тенденция к движению</p> <p>Тип взаимодействия: F (ant) > F (ago)</p> <p>Результат: покой]</p>	<p><i>остановить, зацепить(ся), прилипнуть, зафиксировать, скрепить и др.</i></p>
---	---

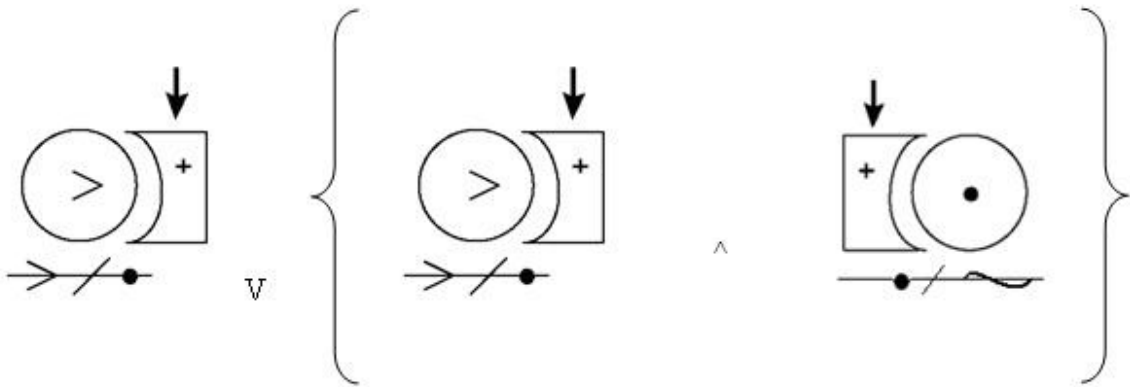


v



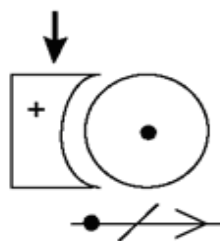
4а Модель *f* в классификации Талми (концепт **Преграждения)**

<p>[<u>Антагонист (Ant)</u>]: Локализация X <u>Агонист (Ago)</u>: Объект Y</p> <p>Начальное состояние: Антагонист – тенденция к покою Агонист – тенденция к движению</p> <p>Тип взаимодействия: $F(ant) > F(ago)$</p> <p>Результат: покой]</p>	<p><i>Застрять, увязнуть, уткнуться и др.</i></p>
---	---



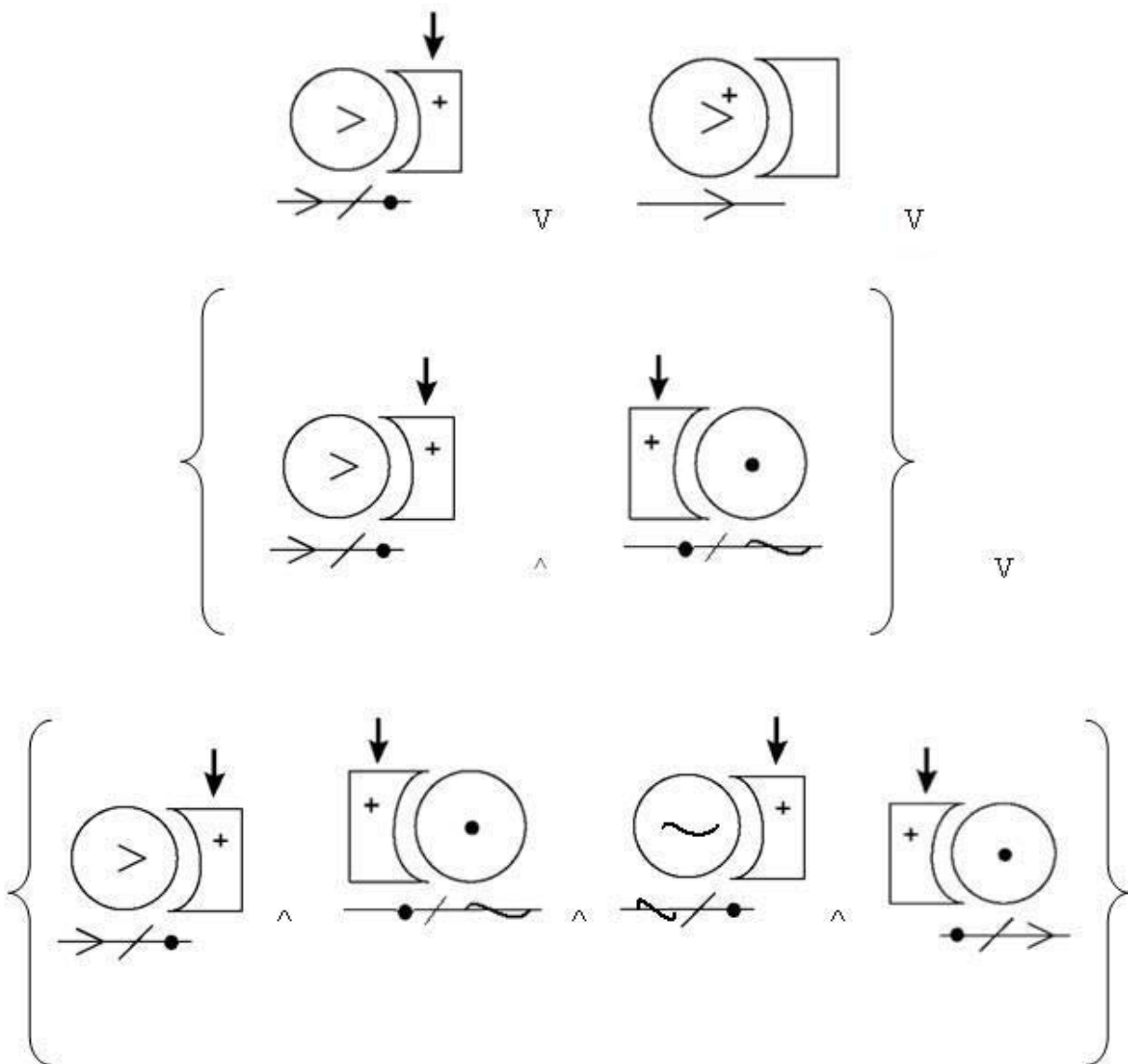
5 Модель *e* в классификации Талми (концепт **Инциптивного изменения)**

<p>[<u>Антагонист (Ant)</u>]: Локализация X <u>Агонист (Ago)</u>: Объект Y</p> <p>Начальное состояние: Антагонист – тенденция к движению Агонист – тенденция к покою</p> <p>Тип взаимодействия: $F(ant) > F(ago)$</p> <p>Результат: движение]</p>	<p><i>оттолкнуть, уронить, придвинуть и др.</i></p>
--	---



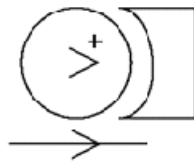
6 Модели (f, c) в классификации Талми (концепт Столкновения)

<p>[Антагонист (Ant): Субъект X <u>Агонист (Ago):</u> Объект Y</p> <p>Начальное состояние: Антагонист – тенденция к покою Агонист – тенденция к движению</p> <p>Тип взаимодействия: -</p> <p>Результат: -]</p>	<p><i>встать на пути, преградить, врезаться в, столкнуться с, столкнуться (друг с другом), подтолкнуть и др.</i></p>
--	--



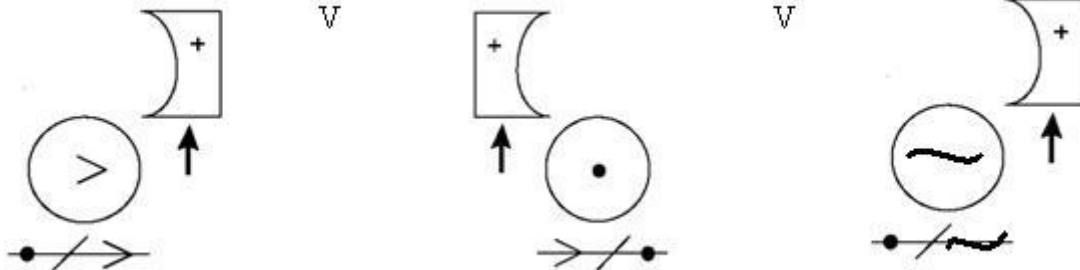
6a Модели (c) в классификации Талми (концепт Преодоления преграды)

<p>[Антагонист (Ant): Субъект X Агонист (Ago): Объект Y</p> <p>Начальное состояние: Антагонист – тенденция к покою Агонист – тенденция к движению</p> <p>Тип взаимодействия: -</p> <p>Результат: -]</p>	<p><i>Протаранить, пробить, пройти через и др.</i></p>
--	--



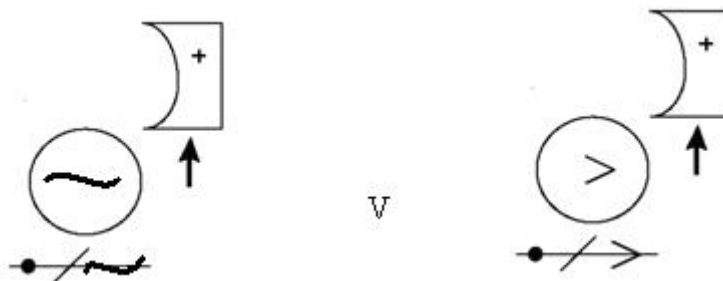
7 Модели (g, h) в классификации Талми (концепт Снятия воздействия)

<p>[Антагонист (Ant): Субъект X Агонист (Ago): Объект Y</p> <p>Начальное состояние: Антагонист - Агонист -</p> <p>Тип взаимодействия: снятие воздействия</p> <p>Результат: -]</p>	<p><i>пустить, отпустить, выпустить, перестать P (где P – гл движения/силового воздействия), дать (X) (где X – глагол движения) и др.</i></p>
--	---



7a Модели (*g, h*) в классификации Талми (концепт **Снятия опоры**)

<p>[Антагонист (Ant): Субъект X</p> <p>Агонист (Ago): Объект Y</p> <p>Начальное состояние: Антагонист – Агонист –</p> <p>Тип взаимодействия: снятие воздействия</p> <p>Результат: -]</p>	<p><i>дать (X)</i> (где X – глагол движения), <i>уронить, отколоть, отвинтить,</i> <i>открепить, повалить, опрокинуть,</i> <i>отделить, и др.</i></p>
--	--



Во **второй** главе («**Разработка онтологии**») дается описание непосредственно онтологии; выделяются концепты, составляющие онтологию, и разрабатывается их внутренняя структура.

В первом параграфе обозначен подход к построению концептов силового взаимодействия: сначала разрабатывается модель предметной области, а затем исследуется ее проекция на язык. Данный подход характеризуется нами как наиболее эффективный, так как при разработке универсальной онтологии появляется возможность избежать неоднозначности, присущей семантике слов естественного языка. Кроме того, в данном параграфе сравниваются понятия онтологии и тезауруса.

Второй параграф характеризует общую структуру онтологии и ее составляющих. Онтологическая структура представляет собой не простую иерархию понятий; скорее, она демонстрирует состав и взаимосвязи свойств понятий предметной области.

В третьем параграфе задаются границы предметной области, для которой осуществляется построение онтологии силовых взаимодействий. Определение границ предметной области в случае работы с фрагментом универсальной

онтологии – достаточно сложная задача, поскольку мы имеем дело с бесконечно сложно организованным объектом моделирования. В диссертационном исследовании границы предметной области обозначены следующим образом: рассмотрению подлежат любые концепты, отражающие состояние физических объектов и изменения, происходящие в процессе силового взаимодействия. Типы силовых процессов разнообразны: действие силы, сила противодействия, преодоление силы противодействия, блокирование силового воздействия, снятие блокирования и т.д.

В четвертом параграфе характеризуются общие принципы функционирования онтологии и лексикона для построения семантического представления текста. Семантическое представление имеет двойной формат. На основе онтологии создается базовый динамический контур силовых переходов, который представляет собой результат логического выбора из предусмотренного каждым концептом множества потенциальных вариантов развития ситуации. Лексикон помогает восстановить в СемП те аспекты референтной ситуации, которые влияют на выбор той или иной лексической единицы при описании действительности. Каждая единица лексикона соотносится с определенным концептом. С одной стороны, принадлежность к концепту онтологии влияет на распределение статусов слотов внутри лексической статьи, а с другой стороны, позволяет системе обработки текста в тексте получить доступ к онтологии через употребление той или иной лексемы. Формат лексического описания отличается от формата представления концепта.

Далее в параграфе рассматривается формальное представление концепта силового взаимодействия и словарной статьи лексикона, чьи структурные элементы представлены в виде слотов.

Набор формально описанных концептов соответствует семантическим классам, выделенным в главе I, и графической нотации. Метаязык описания концептов имеет формат, использованный С. Ниренбургом и В. Раскиным в проекте «Онтологическая семантика». Формально фиксируется значение каждого слота. Дочерние концепты наследуют слоты и их значения от верхних узлов онтологии.

Набор слотов фиксирован для каждого из концептов и выглядит следующим образом:

I. 1. Участники ситуации (роли Агониста и Антагониста присваиваются физическим объектам):

```
ANT (SEM (PHYSICAL_OBJECT))
```

```
AGO (SEM (PHYSICAL_OBJECT))
```

2. Внутренняя силовая тенденция (возможные значения: стремление к изменению (либо движение, либо деформация), покою, движению или деформации):

```
(HAS_TENDENCY (SEM ({ACTION, REST, MOTION, DEFORM})))
```

II. 3. Соотношение сил (возможные значения: неизвестно, снятие воздействия, Антагонист сильнее, Агонист сильнее):

```
FORCE_BALANCE (SEM ({UNKNOWN, LETTING,  
STRONGER (SEM (ANT)),  
STRONGER (SEM (AGO))}))
```

III. 4. Начальное состояние Агониста (возможные значения: покой, движение, деформация):

```
FD_INITIAL (SEM ({REST, MOTION, DEFORM}))
```

5. Результат взаимодействия (возможные значения: неизвестен, изменение, движение, деформация, покой):

```
FD_RESULT (SEM ({UNKNOWN, ACTION, MOTION, DEFORM, REST}))
```

Также в данном параграфе схематически представлена структура онтологии силовых процессов.

Формат фреймового описания каждого концепта проиллюстрируем на примере концепта Постоянного изменения:

```
(FD_STEADY_STATE
```

```
(IS-A
```

```
(VALUE  
FD_EVENT))
```

```
(SUBCLASSES
```

```
(VALUE  
(ACTION_STEADY_STATE, REST_STEADY_STATE))
```

(AGO

(HAS_TENDENCY
(SEM
({ACTION, REST})))

(FORCE_BALANCE

(SEM
(STRONGER (SEM(ANT))))

(FD_INITIAL

(SEM
({REST, MOTION, DEFORM})))

(FD_RESULT

(SEM
({ACTION, REST})))

Формат лексической статьи имеет фреймовую структуру, и состав слотов определяется отнесенностью статьи к тому или иному концепту. Выделяются два типа лексического описания на основе существенного различия в характере приложения силы к объекту: перемещения в пространстве и деформации.

Оба типа лексических статей делятся на четыре раздела: физические свойства Агониста и Антагониста (*Participant_type*), фокус внимания³ (*Focus*), характер приложения силы (*Force_application*), свойства окружающей среды (*Environment*). В каждом разделе перечислены релевантные свойства референтной ситуации и ее участников. Физические свойства Агониста и Антагониста включают: одушевленный / неодушевленный, лицо / не-лицо, размер, вес, топологический тип, свойства материала, из которого сделан объект (*ANIM, TOPOLOGICAL_TYPE*). Характер приложения силы описывается следующими параметрами: вектор, интенсивность, постоянность приложения силы, тип контакта с поверхностью⁴ (*VECTOR, INTENSITY, APPLICATION, CONTACT_ANT, INSTR_TYPE, CONTACT_PATH*⁵, *PROMPTNESS*). Кроме того, в данной статье описаны траектория движения Агониста или характер его деформации

³ показывает участника, с точки зрения которого показана ситуация (ср. *опираться* vs. *подпирать*)

⁴ значения слотов этого раздела отличаются в зависимости от того, описывает лексическая статья предикат деформации или перемещения в пространстве

⁵ контакт с поверхностью: по всей поверхности агониста или точка соприкосновения с поверхностью фиксирована

(Ago_trajectory (TRAJECTORY, TRAJECTORY_LENGTH, VELOCITY), Deformation_type (DEFORM_VECTOR, DEFORM_TYPE)). Свойства окружающей среды включают: устойчивость, вязкость/шероховатость, наклон поверхности (ROUGHNESS, INSTABILITY, BIAS).

Для того, чтобы определить, какие атрибуты будут обязательными в той или иной лексической статье и какие из них будут иметь больший вес в сравнении с другими при построении СемП, введено понятие семантической доминанты. В диссертационном исследовании это понятие формулируется следующим образом: под **семантическими доминантами** понимаются семантические признаки или слоты, которые в той или иной степени влияют на выбор средства лексического выражения. На наш взгляд, понятие семантической доминанты напрямую связано с когнитивными процессами, происходящими в сознании говорящего при порождении высказывания. При описании реальности говорящий всегда делает выбор между квазисинонимичными языковыми средствами; за этот процесс отвечают определенные механизмы – такие, как, скажем, механизм распределения внимания между важными и второстепенными элементами референтной ситуации. Однако, поскольку мы имеем дело с моделированием реальности, представленной в сознании рядового носителя языка в виде естественноязыковых значений, то их игнорирование могло бы привести к нежелательному огрублению всей системы в целом. Результаты работы с материалами толковых словарей и корпусов текстов показывают, что достаточно часто один и тот же референтный сценарий силового взаимодействия может быть выражен разными способами. В одних случаях можно говорить о том, что один способ лучше другого, в других – о том, что не вполне понятен механизм взаимодействия, поэтому затруднен выбор языкового средства (в частности, могут быть недостаточно ярко представлены значения определяющих признаков описываемой ситуации, что смущает говорящего). Это наблюдение возвращает нас к мысли о том, что лексические значения устроены прототипически. Для описания одной и той же ситуации могут быть более или менее успешно использованы сразу несколько предикатов. Видимо, это явление зависит не только от системы языка, но и от того, насколько наблюдаемая ситуация соответствует прототипической для каждой из имеющихся в распоряжении говорящего языковых единиц. Так, например, относительный размер участников силовой ситуации может

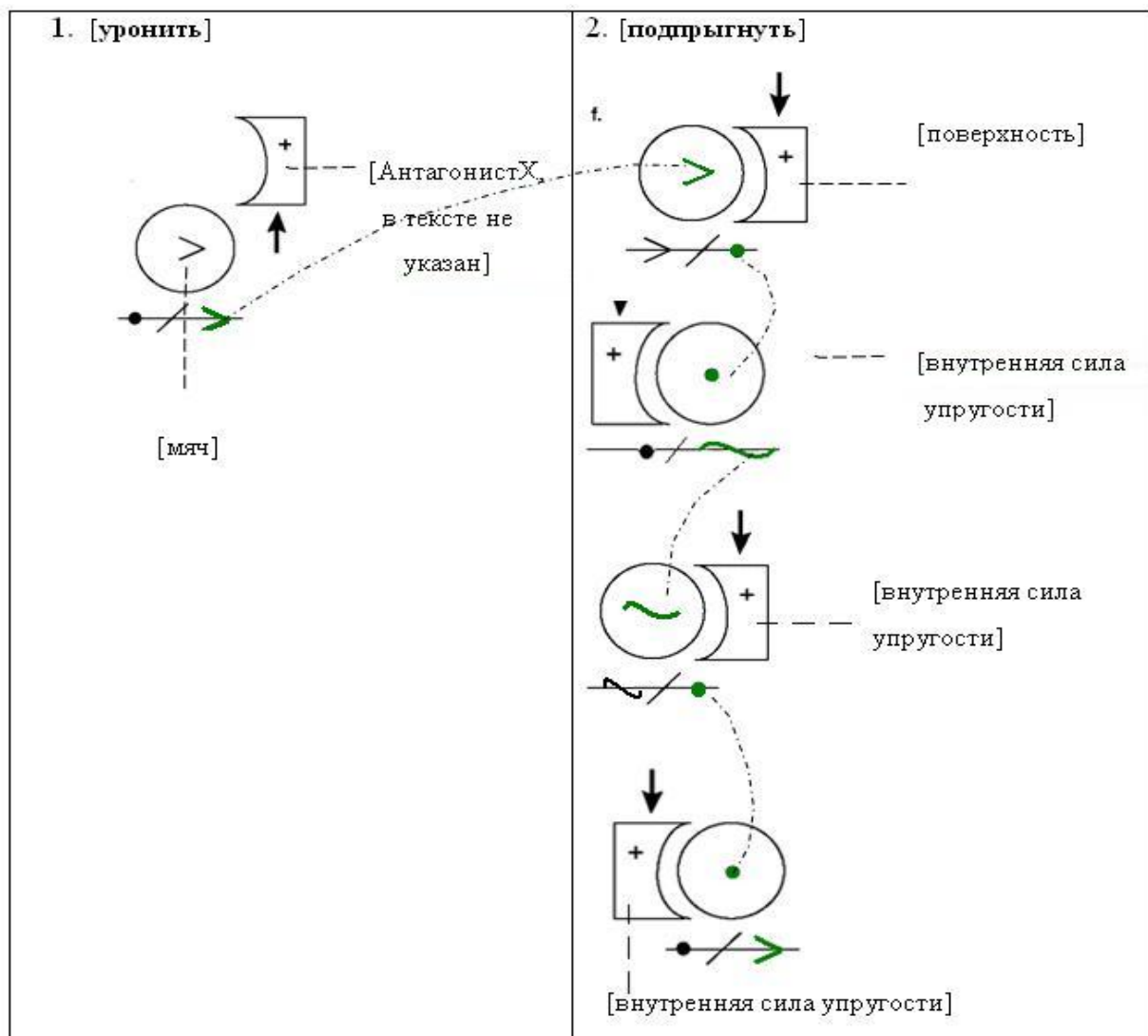
быть разным: если один из участников значительно больше другого, то это яркое соотношение поможет уверенно подобрать необходимое языковое выражение (например, предикат *повалить* подразумевает объект внушительного размера по сравнению с субъектом и легко употребляется и интерпретируется именно в таком контексте). Если же эта разница неочевидна, то при отсутствии прочих ярких признаков, которые диктовали бы употребление другого предиката, выбор адекватного языкового выражения будет несколько затруднен. Данный факт, впрочем, не означает, что выбранное в конечном счете выражение будет расценено как неправильное, так как наряду с прямым оно может обладать и метонимическим прочтением (так, например, в контексте *сжимать перчатку в руках* глагол означает ‘крепко держать’, хотя основной интерпретацией этого глагола можно считать ‘сдавливать’, ср. *сжимать резиновый шарик*). Л. Талми пишет о том, что при порождении высказывания говорящий пользуется системой схематизации недискретной ситуации реального мира, когда при репрезентации референтной сцены учитываются лишь определенные ее аспекты, а другие остаются неучтенными. Механизм схематизации делает возможным концептуальное структурирование, т.е. внутреннюю организацию смыслов в языковой системе. Результатом действия этого процесса является так называемая “рамка события” или “событийный фрейм” (в терминологии Талми “event frame”), в котором будут присутствовать приоритетные или доминантные элементы.

Каждая лексическая статья, в зависимости от того, к какому концепту она относится, имеет свои семантические доминанты. С целью их выявления было проведено анкетирование носителей русского языка. Опрос основывался на определении семантических аномалий в предложениях. Для каждого предиката силового взаимодействия нами строилось простое предложение. Мы предложили анкетироваемым оценить предложения как “приемлемые”, “неприемлемые” и “неполные”. В случае, если предложение оценивалось как “неприемлемое” или “неполное”, анкетироваемый должен был аргументировать свое решение. В результате для каждой лексической статьи, относящейся к тому или иному концепту в онтологии, были определены 3-4 доминантных семантических признака, которым поставлены в соответствие те или иные атрибуты.

В третьей главе показано, как строится семантическое представление текста на примере задач по физике для учащихся 7-9 классов. Употребление лексических единиц с силовой семантикой в таких текстах, как правило, не противоречит обыденному, а каждый из текстов представляет собой некоторую, обычно завершённую, цепь сменяющих друг друга силовых переходов, на примере которой достаточно просто показать, как на практике выглядит их анализ. Рассмотрим следующую задачу:

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

Ниже представлен сценарий силовых переходов, построенный на базе концептов:



Далее приведен пример семантического представления текста задачи, который построен с помощью разработанного нами формализма:

FD_LETTING

Antagonist [X]

[Ant_type **PERS (pers)**
INTENSIONAL (-)]

Agonist [«мяч»]

[Ago_type
ANIM(inanim)
TOPOLOGICAL_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, var)]

Focus

[Focus_Value (Ant)]

Force Application

[Force_Application
VECTOR (down)
CONTACT_ANT (direct)
]

Ago_trajectory

[Ago_trajectory
TRAJECTORY (default {= vector value})
TRAJECTORY_LENGTH (H)]

Environment [«лифт»]_E

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

FD_BARRIER

Antagonist

[Ant_type₁ **SURFACE_E**
Ant_type₂ **INNER_ELASTIC_FORCE**
Ant_type₃ **INNER_ELASTIC_FORCE**]

Agonist [«мяч»]

[Ago_type
ANIM(inanim)

TOPOLOGICAL_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, var)]

```

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)]
Deformation_type1
  [DEFORM_VECTOR (default {= vector1 value})
  DFORM_TYPE (elastic)]
Ago_trajectory2
  [TRAJECTORY (stop)
  TRAJECTORY_LENGTH (var)]
Ago_trajectory3
  [TRAJECTORY (vector3 value)
  TRAJECTORY_LENGTH (h)]
Environment [«лифт»]
  [Path/Environment
  ROUGHNESS (0)
  INSTABILITY (-)]

```

В заключении обобщаются основные выводы исследования.

Была рассмотрена предметная область различного рода физических взаимодействий, и на основе теории Динамики Сил Л. Талми разработан принцип ее формального описания, ориентированного на проекцию данной предметной области в языке.

Осуществлена классификация ситуаций силового взаимодействия и предикатов русского языка с силовой семантикой. На ее основе разработан фрагмент универсальной онтологии для ситуаций силового взаимодействия (FD_EVENT), который насчитывает 10 концептов.

Предложено лексическое описание на основе онтологии, которое учитывает основные роли и атрибуты ситуаций силового взаимодействия (условия протекания ситуации, топологические характеристики участников ситуации, направление действия силы и пр.).

Разработан алгоритм построения семантического представления текста на естественном языке с использованием онтологии и лексикона.

Апробация работы. Основные положения диссертации были представлены и обсуждены на кафедре теоретической и прикладной лингвистики МГУ им. Ломоносова, а также на ежегодной Международной конференции по компьютерной лингвистике «Диалог» («Бекасово» 2006 и «Бекасово» 2010).

Основное содержание диссертационной работы отражено в следующих публикациях:

1. **Марушкина А.С. Семантическое представление текста, описывающего механику силовых процессов, в универсальной онтологии // НТИ, серия 2 «Информационные процессы и системы» №4, 2012; С. 13-18.**
2. *Марушкина А.С. "Наивная механика" в языке и онтологии // Сборник трудов международной конференции Диалог, 2006 (Бекасово, 31 мая– 4 июня 2006 г.); С. 375-382.*
3. *Кобозева И.М., Марушкина А.С. Онтология силовых процессов // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 26-30 мая 2010 г.). Вып. 9 (16)., Москва: Изд-во РГГУ, 2010. С. 192-200.*