

Онтологическая семантика текста: форматирование лексики в семантическом словаре

© Г.В. Лезин

Санкт-Петербургский экономико-математический институт РАН
lezin@emi.nw.ru

Аннотация

В статье рассматривается задача выработки единого представления информации в семантической сети текста, лексиконе и в базовой онтологии. Предлагается дополнить стандартные описания классов в онтологии форматами многокомпонентных толкований, подготовленных для использования в лексиконе. Рассматривается семантика форматов и их связь с онтологией и лексиконом.

1 Введение

Интерпретация текста набором формальных утверждений, записываемых на предварительно строго определенном формальном метаязыке, является одной из центральных и не решенных в настоящее время задач автоматического анализа текста. Многообразие требований, предъявляемых к такому метаязыку со стороны различных приложений, делает практически невозможным его создание в законченном, раз и навсегда фиксированном виде. Но можно попытаться использовать в качестве метаязыка строго формализованную систему стандартных правил определения средств описания семантики текста для нужд конкретных приложений. Примером такого подхода, но для формирования описаний семантики данных во всемирной семантической паутине (Semantic Web) является предложенная W3C Консорциумом система формальных языков RDF/OWL [1, 2]. Опора на хорошо исследованную математическую модель (дескриптивные логики), наличие доступной и тщательно разработанной документации, имеющаяся уже в настоящее время программная поддержка (редакторы баз знаний, связи с языком программирования и современными базами данных) делают эту систему языков привлекательной и для экспериментов, нацеленных на получение формальной интерпретации текста.

Целесообразность интерпретации текста семантической сетью отношений между

сущностями, упоминаемыми в тексте, стала общепризнанным фактом [3]. Сеть текста строится на основе описаний лексики языка, получаемых из семантического словаря (лексикона). Согласование описаний лексики в лексиконе с представлением семантических сетей текстов, формируемых на основе этих описаний, создание единой семантической основы, в терминах которой описываются как лексика языка, так и семантические сети текстов, написанных на этом языке – это задача, без решения которой говорить об автоматической интерпретации текстов просто не приходится.

Попытка определения такой общей семантической основы в виде сильно ограниченного универсального набора семантических примитивов так и не привела к созданию общепризнанного семантического языка [4]. Рассматривая известные проекты создания семантических словарей с формальными толкованиями лексики [5-8], можно отметить следующие характеристики семантической основы:

а) Определению формул толкования предшествует выделение подсловаря исходных (атомарных) понятий, используемых в толкованиях оставшейся части лексикона (например, в словаре В.А. Тузова [7] более 3000 атомарных понятий).

б) Важной компонентой толкования лексического значения слова является указание места лексемы в общей таксономической иерархии понятий лексикона.

в) Наличие развитого списка свойств (ролей, атрибутов), позволяющих связывать атомарные понятия в общую формулу толкования лексемы.

г) В общей таксономической иерархии лексикона выделяются классы лексем, формулы толкования которых имеют общие части. Исследованию связей между таксономической характеристикой слова и форматом его толкования посвящена монография [6].

Отмеченные характеристики лексикона, ориентированного на получение семантической интерпретации текста, позволяют надеяться на возможность оформления метаязыка системы лексикон/текст средствами RDF/OWL. Подчеркнем, что речь идет о создании лингвистического ресурса для решения прикладных задач, в частности задач извлечения информации из текста. Рассматриваемые в статье задачи и возможные

методы их решения относятся, главным образом, к инженерной технике совместного представления информации в лексиконе и лежащей в его основе онтологии.

Будем исходить из следующих предположений [9].

1) Семантическая основа лексикона оформляется в виде базовой OWL-онтологии, содержащей описания атомарных понятий.

2) Семантическая сеть анализируемого текста описывается конъюнкцией RDF-триплетов. Узлами семантической сети могут быть:

- имена классов базовой OWL-онтологии;
- имена классов – лексических значений слов текста, полученные из лексикона;
- имена-референты (индивиды и подклассы классов базовой OWL-онтологии, упомянутые в тексте) автоматически сформированные в процессе анализа текста;
- узлы-переменные (blank nodes), сформированные в соответствии с модельной семантикой RDF.

Ребрами семантической сети могут быть исключительно имена свойств, заданные базовой OWL-онтологией.

3) Каждое из лексических значений слова (лексема) в лексиконе толкуется как подкласс одного или нескольких классов из базовой онтологии. Возможно уточнение заданного общим классом толкования лексемы в виде высказывания на атомарных элементах OWL-онтологии. Толкования в лексиконе оформляются в виде RDF-сетей, подготовленных для включения в общую семантическую сеть текста. Явно выделены входной и выходные узлы сети толкования; узлы-переменные сети толкования согласованы с актантными переменными, указанными для синтаксических актантов лексемы.

Важно отметить, что базовая онтология в сочетании с лексиконом образует схему примитивных знаний о мире, нашедших отражение в лексике языка. Создана такая база может быть лишь в результате опыта, как результат исследований разных коллективов, но представляется принципиально важным, чтобы в ее основе лежали общепринятые средства описания знаний. Мы в нашей работе рассматриваем всего лишь возможный подход к решению этой задачи – общую технику описания лексики языка, позволяющую увязать в единый комплекс информацию в онтологии, лексиконе и анализируемом тексте.

Согласование информации в базовой OWL-онтологии, лексиконе и семантической сети текста связано с необходимостью поиска ответов на следующие вопросы:

а) Как определить в OWL-онтологии атомарные элементы, допустимые для использования в толкованиях лексикона (возможные ограничения, правила преобразования OWL-описания элемента в его же RDF-представление в составе толкования)?

б) Как увязать отобранные для толкования атомарные элементы в параметризованное

высказывание (определение общей структуры толкования, выделение списка семантических актантов и их связь с синтактикой лексемы, возможность автоматизации процесса увязывания)?

в) Как организовать коллекцию общих форматов толкований (связь формата толкования с OWL-онтологией и его использование в конкретных толкованиях)?

В статье мы пытаемся найти возможный вариант ответа на поставленные вопросы.

2 Атомарные элементы онтологии.

Наша задача – оформление толкования лексемы в виде RDF-фрагмента семантической сети, точнее, в виде шаблона, переменные которого при подключении фрагмента к семантической сети текста, как правило, подлежат замене на константные значения. В RDF семантическая сеть интерпретируется как конъюнкция триплетов [13]. Логично потребовать, чтобы толкование лексемы представляло собой конъюнкцию атомарных элементов, получаемых из базовой OWL-онтологии и тоже оформленных в виде фрагментов семантической сети. (Мы ограничиваемся однозначными толкованиями значений слова. О возможности использования дизъюнктивно организованных толкований см. [14].)

Исходной информацией для построения атомарного элемента толкования служит описание класса в базовой OWL-онтологии. Согласно [15] описание класса содержит указание класса и список ограничений на свойства, сопоставляемые индивидам класса. Класс в описании может быть задан либо явно, либо как результат применения операций пересечения, объединения или дополнения классов. В любом случае описание класса позволяет выделить список свойств, явно сопоставленных классу указаниями ограничений на свойства. Полученный таким образом список свойств будем называть списком свойств класса.

Свойство *prop* может быть сопоставлено классу C_1 любым сочетанием из трех видов ограничений:

а) Ограничение "only V": любой индивид u из класса C_1 может иметь свойство *prop* с областью значений, ограниченной исключительно классом V , т.е.:

$u \text{ rdf:type } C_1; \text{ prop } v \text{ влечет } v \text{ rdf:type } V.$

б) Ограничение "some V": факт принадлежности индивида u классу C_1 свидетельствует о наличии у него некоторого значения v свойства *prop*, т.е.:

$u \text{ rdf:type } C_1 \text{ влечет } \exists (u \text{ prop } v. v \text{ rdf:type } V)$

в) Ограничения кардинальности: каждый индивид u из класса C_1 должен иметь в семантической сети значения свойства *prop* в количестве, удовлетворяющем ограничению (например, при *owl:minCardinality=2* индивиду u в сети должно быть сопоставлено не менее двух значений свойства *prop*, при *owl:maxCardinality=3* – не более трех значений этого свойства, а при *owl:cardinality=2* – в точности два значения)

Предположим, что описание класса с именем *ont:ClassId* сопоставляет этому классу список из N

свойств: `ont:prop1`, ..., `ont:propN`. Атомарным шаблоном класса `ClassId` будем называть RDF-граф вида

```
_:x rdf:type ont:ClassId; ont:prop1 _:x1; ... ; ont:propN _:xN.
```

В определении использована нотация `N3`; `_:x`, `_:x1`, ..., `_:xN` – обозначения переменных; `ont:` – префикс пространства имен OWL-онтологии. Шаблон получен в соответствии с правилами трансляции OWL-описания класса `ont:ClassId` в RDF-представление этого описания [15] с исключением из результатов трансляции ограничений на свойства, заданных в OWL-описании.

Пример 1.

В нашей базовой OWL-онтологии класс `ont:Нахождение_Место` описывает местонахождение объекта в заданном интервале времени.

```
ont:Нахождение_Место
  rdf:type owl:Class;
  rdfs:subClassOf ont:СостояниеОбъект;
  rdfs:subClassOf
    [rdf:type owl:Restriction ;
     owl:onProperty ont:c_Уровнем_высоты;
     owl:allValuesFrom ont:РазмерДистанции;
     rdfs:subClassOf
       [rdf:type owl:Restriction ;
        owl:onProperty ont:c_Уровнем_высоты;
        owl:someValuesFrom ont:РазмерДистанции
       ];
     rdfs:subClassOf
       [rdf:type owl:Restriction ;
        owl:onProperty ont:c_Местом ;
        owl:someValuesFrom ont:Место_Объекты
       ];
     rdfs:subClassOf
       [rdf:type owl:Restriction ;
        owl:onProperty ont:c_Местом;
        owl:allValuesFrom ont:Место_Объекты
       ];
     rdfs:subClassOf
       [rdf:type owl:Restriction ;
        owl:onProperty ont:c_Дистанцией;
        owl:allValuesFrom ont:РазмерДистанции
       ]
  ] .
```

В приведенном описании класса `ont:Нахождение_Место` указаны лишь непосредственно заданные для этого класса ограничения на свойства. Ряд других свойств, также заданных ограничениями, наследуется от объемлющего класса `ont:СостояниеОбъект` (свойство `ont:c_Объектом` определяет субъекта, находящегося в определяемом состоянии; свойствами `ont:c_Началом` и `ont:c_Концом` – интервал времени существования состояния и свойством `ont:c_Модальностью` – фактивность состояния). В результате атомарным шаблоном класса `ont:Нахождение_Место` является граф:

```
_:x rdf:type ont:Нахождение_Место;
  ont:c_Объектом _:x1;
  ont:c_Местом _:x5; ont:c_Дистанцией _:x6;
```

```
ont:c_Уровнем_высоты _:x7;
  ont:c_Началом _:x2; ont:c_Концом _:x3;
  ont:c_Модальностью _:x4.
```

Переменную, определенную на классе атомарного шаблона (в Примере 1 это `_:x`), мы будем называть переменной шаблона, прочие переменные – параметрами шаблона.

Отметим, что использование конкретных обозначений переменных в атомарном шаблоне, вообще говоря излишне. Можно было бы воспользоваться парой скобок '[]', считая каждую такую пару скобок указанием переменной с обозначением, отличным от других переменных. Мы будем использовать эту возможность в последующих примерах.

Отображение времени текста, регистрация временных отношений между ситуациями, описываемыми в тексте, является одной из основных компонент его семантической интерпретации. Учет времени текста осуществляется на основе информации, поставляемой толкованиями лексических значений слов [10]. Соответственно, базовая OWL-онтология должна обеспечивать необходимый набор атомарных понятий. Мы в нашем проекте следуем подходу, разрабатываемому в онтологии DOLCE [11].

Все классы OWL-онтологии делятся на две категории: времязависимые и постоянные. Любому из индивидов времязависимого класса сопоставляется интервал времени его существования. Отметим, что свойства любой онтологии, описанной на языках RDF или OWL по определению постоянны, не имеют индивидов и не могут, в свою очередь, иметь свойств. В то же время индивид, представляющий сущность из времязависимого класса, может в разное время иметь различные значения одного и того же свойства. (Например, какой-либо предмет может в разное время находиться в разных местах.) Отсюда возникает потребность сопоставлять время существования не только индивидам предметного плана, но и состояниям этих индивидов, процессам, участниками которых они являются. В общем классе времязависимых сущностей выделяется иерархия подклассов, позволяющих описать процессы и состояния с указанием времени их существования. Описание отношений между временными интервалами, их протяженности и положения на временной оси может быть выполнено, например, средствами временной онтологии [12].

3 Форматы (сценарии) толкований.

Достаточно широкому спектру лексических значений слов свойственна многокомпонентность толкований. Типична ситуация, когда некоторое множество лексем характеризуется наличием общей структуры толкования и именно эта общность служит основанием для объединения этих лексем в общий класс онтологии. Например, для глаголов, описывающих перемещение объекта из исходной

точки в конечную (*лететь, передвигать, идти* и др.), характерен сценарий:

- нахождение объекта в исходной точке;
- затем: движение объекта из исходной точки в конечную;

– затем: нахождение объекта в конечной точке.

Явно выделяются семантические участники этого сценария: Агенса, Объекта, Исходная точка, Конечная точка, интервал времени, охватываемый сценарием.

Возникает задача дополнения базовой OWL-онтологии достаточно тесно увязанным с нею и тем не менее новым информационным ресурсом – описаниями форматов толкований. Одним из возможных способов подключения такого ресурса мог бы быть следующий:

а) Формат толкования сопоставляется классу OWL-онтологии. В качестве свойств этого класса, явно заданных в его описании или наследованных, указываются имена ролей, задающих параметры формата толкования.

б) Определяется набор шаблонов, образующих толкование. Шаблоны задаются указанием соответствующих классов базовой OWL-онтологии.

в) Нормализация (переобозначение) переменных в шаблонах. В результате этой операции переменные в шаблонах не имеют совпадающих обозначений.

г) Параметризация шаблонов. Переменным в шаблонах, обозначающим общие узлы графа толкования, присваиваются общие обозначения (слияние узлов графа, обозначенных этими переменными). Полученный в результате действий в) и г) словарь обозначений переменных представляет собой полный (в рамках заданной базовой OWL-онтологии) список семантических параметров формата – набор узлов графа толкования, допустимых для слияния с узлами общей семантической сети текста. Этот словарь аморфен в том смысле, что в нем отсутствуют деление на актанты и сирконстанты, а также указания коммуникативного ранга указанных им параметров.

Пример 2.

Отмеченный выше сценарий перемещения может быть представлен следующим форматом:

```
<script>
<header>
_x rdf:type ont:СценарийПеремещения;
 ont:c_Объектом _:Y1; ont:рольИточка _:Y2;
 ont:рольИдистанция _:Y3;
 ont:рольИвысота _:Y4; ont:рольКточка _:Y5;
 ont:рольКдистанция _:Y6;
 ont:рольКвысота _:Y7; ont:рольНачало _:Y8;
 ont:рольКонец; _:Y9".
</header>
<body>
_x1 rdf:type ont:Нахождение_Место;
 ont:c_Объектом _:Y1; ont:c_Местом _:Y2;
 ont:c_Дистанцией _:Y3;
 ont:c_Уровнем_высоты _:Y4;
 ont:c_Началом [ ]; ont:c_Концом _:Y8.
_x2 rdf:type ont:ДвижениеПроцесс;
 ont:c_Агенса [ ]; ont:c_Объектом _:Y1;
 ont:c_Скоростью [ ]; ont:мимо [ ];
 ont:следом [ ]; ont:c_Началом _:Y8;
 ont:c_Концом _:Y9.
_x3 rdf:type ont:Нахождение_Место;
 ont:c_Объектом _:Y1; ont:c_Местом _:Y5;
 ont:c_Дистанцией _:Y6;
 ont:c_Уровнем_высоты _:Y7;
 ont:c_Началом _:Y9; ont:c_Концом [ ].
</body>
</script>
```

Формат имеет две явно выраженные части: заголовок и тело формата. Заголовком формата является шаблон класса общей OWL-онтологии. В Примере 2 таким классом является `ont:СценарийПеремещения`. Тело формата содержит RDF-граф, интерпретирующий класс, заданный заголовком. Формат оформлен в виде XML-контейнера.

RDF-граф тела формата представляет собой результат объединения шаблонов из заданного набора с предварительной нормализацией переменных шаблонов и последующей параметризацией полученного объединения. В Примере 2 заголовок интерпретируется набором из трех атомарных шаблонов: `ont:Нахождение_Место`, `ont:ДвижениеПроцесс` и `ont:Нахождение_Место`. Парамы скобок '[']' обозначены места вхождения переменных с уникальными в пределах формата обозначениями, полученными в результате нормализации. Явно обозначенные Y-переменные отмечают результаты параметризации. В совокупности имена неявно обозначенных переменных и Y-переменных образуют словарь параметров формата. Имена `_:x1`, `_:x2` и `_:x3` использованы для обозначения переменных шаблонов.

Одно из основных назначений формата – многократное использование в толкованиях конкретных лексем. Принципиально важной является семантическая эквивалентность между заголовком формата и его телом. Содержательно эквивалентность заголовка и тела формата должна проявляться в следующем:

а) Указывая в толковании лексемы шаблон класса, которому сопоставлен формат, мы тем самым объявляем тело формата компонентой толкования лексемы.

б) Выявив в семантической сети текста фрагмент, из которого следует истинность высказывания, заданного телом формата, мы тем самым выявляем в тексте наличие примера сценария, которому сопоставлен этот формат.

Формальное определение эквивалентности основано на предположении, что полные списки параметров заголовка формата и его тела должны совпадать. Иначе говоря, мы должны иметь возможность сослаться на любой параметр тела формата в его заголовке. Отметим, что параметры в шаблонах идентифицируются свойствами,

значениями которых они являются. Обозначения переменных не могут служить идентификаторами параметров в силу локальности их использования и постоянной готовности к переименованию в процессе нормализации. При взаимно однозначном соответствии пары "свойство" – "параметр", выявленном в теле формата, соответствующее свойство удобно считать сопоставленным классу формата и явно не указывать в списке свойств его заголовка. В Примере 2 классу `ont:СценарийПеремещения` наряду со свойствами, перечисленными в заголовке, сопоставлены также свойства `ont:c_Агенсом`, `ont:co_Скоростью`, `ont:мимо`, `ont:следом`; при этом в заголовке формата обеспечено взаимно однозначное соответствие между параметрами формата и идентифицирующими их свойствами.

Пусть F – любой из классов общей OWL-онтологии, которому сопоставлен формат толкования; X_F – общий список параметров формата F ; x_F – переменная, определенная на классе F ; V_F – набор переменных, определенных на классах шаблонов, образующих толкование (в нашем примере $V_F = \{ _x1, _x2, _x3 \}$); $\exists(V_F)$ – факт существования кортежа значений для переменных из набора V_F ; $H_F(x_F, X_F)$ – RDF-высказывание, заданное заголовком формата F и $S_F(V_F, X_F)$ – RDF-высказывание тела F . Тогда заголовок формата F и его тело связаны отношением эквивалентности:

$$\forall(F)[\forall(x_F)[H_F(x_F, X_F) \rightarrow \exists(V_F) S_F(V_F, X_F)] \& \\ \forall(V_F)[S_F(V_F, X_F) \rightarrow \exists(x_F) H_F(x_F, X_F)]]$$

при взаимно однозначном соответствии между значениями x_F и конкретными кортежами значений V_F .

Рассмотренное отношение имеет статус аксиомы, т.е. исходного требования, которому должен удовлетворять вновь определяемый формат. Следствием этой аксиомы является возможность сопоставления форматов в условиях, когда один формат отличается от другого лишь тем, что некоторые из шаблонов этого формата являются подклассами соответствующих шаблонов другого.

Пусть $S1$ и $S2$ – классы общей OWL-онтологии, причем каждому из них сопоставлен собственный формат толкования:

$$S1 \leftrightarrow P_1, \dots, P_i, \dots, P_k \text{ и } S2 \leftrightarrow P_1, \dots, C_i, \dots, P_k$$

где C_i, P_1, \dots, P_k – шаблоны классов OWL-онтологии, C_i является подклассом P_i .

Тогда $S2$ – подкласс $S1$.

Рассмотрим произвольный кортеж $(p_1, \dots, c_i, \dots, p_k)$ значений переменных шаблонов формата $S2$: $(p_1, \dots, c_i, \dots, p_k) \in V_{S2}$. В соответствии с отношением эквивалентности ему сопоставлен единственный индивид $s2$ класса $S2$. Но c_i , будучи индивидом класса C_i , принадлежит и классу P_i . Следовательно, $(p_1, \dots, c_i, \dots, p_k) \in V_{S1}$, и кортежу сопоставлен также и индивид $s1$ из класса $S1$. Это невозможно ввиду требования взаимно однозначного соответствия между индивидами класса, заданного заголовком формата, и кортежами значений переменных шаблонов. Таким образом, $s1 \equiv s2$, т.е.

любой индивид из класса $S2$ принадлежит также классу $S1$.

Определяя формат как средство описания семантики лексики, дополняющее общую OWL-онтологию, полезно рассмотреть вопросы, связанные с корректностью параметризации тела формата и попытаться выявить возможности наследования форматов для классов, связанных отношением "класс–подкласс".

В общем случае процедура параметризации тела формата сводится к попарному слиянию узлов графа, причем каждый из объединяемых узлов имеет в базовой онтологии собственное описание области возможных значений, заданное описанием шаблона. Объединяться могут как узлы, представленные переменными шаблонов, так и параметры шаблонов. В любом случае объединять можно лишь узлы с непротиворечивыми описаниями. Условия корректного слияния естественным образом следуют из семантики языков RDF/OWL:

а) Основным элементом описания узла является указание класса его возможных конкретных значений. Пусть узел $_x_i$ определен на непустом классе C_i , узел $_x_j$ – на непустом классе C_j , а $_y$ – обозначение результата слияния этих узлов. Тогда классом возможных значений узла $_y$ будет пересечение $C_i \cap C_j$. Слияние узлов $_x_i$ и $_x_j$ корректно, если это пересечение не пусто. В частности, при наличии в базовой онтологии указания на непересекаемость C_i и C_j слияние узлов $_x_i$ и $_x_j$ не корректно.

б) Параметры шаблона наряду с указанием класса возможных значений (only-ограничение) могут иметь также some-ограничение и явно заданную кардинальность. Some-ограничение для заданного в шаблоне свойства устанавливает факт существования значения свойства из заданного в ограничении класса (см. п.2). Если хотя бы один из объединяемых узлов имеет some-ограничение, результат слияния также имеет some-ограничение, но класс значений для этого ограничения – пересечение классов, заданных для значений объединяемых узлов. При объединении узлов, имеющих ограничения на кардинальность, для результата устанавливается сильнейшее из заданных ограничений.

Итак, слияние параметров шаблонов может привести к необходимости определения новых ограничений на значения свойств, зафиксированные в исходных шаблонах. В OWL-онтологии мы можем зафиксировать ограничение, которому должно удовлетворять конкретное значение v свойства p в условиях, когда это свойство сопоставлено конкретному классу C_1 (см. п.2). Таким классом в нашем случае является класс, интерпретируемый форматом и заданный его заголовком. Фиксация ограничения сводится к явному определению для этого класса свойства, значения которого обладают требуемым ограничением. В заголовке формата этому свойству сопоставляется соответствующий параметр.

Подводя итог, заметим следующее:

- 1) Полное множество классов OWL-онтологии может быть поделено на две категории: общие классы, полностью определенные средствами онтологии, и классы, которым наряду с их описанием в онтологии сопоставлен формат толкования – F-классы.
- 2) Шаблон F-класса является заголовком сопоставленного классу формата толкования и строится по общим правилам (см. п.2).
- 3) Список параметров шаблона F-класса совпадает со списком параметров, заданных телом его толкования. Ограничения на значения параметров в описании F-класса в онтологии не должны противоречить ограничениям на эти же параметры в теле толкования класса.
- 4) Список свойств F-класса должен обеспечивать в рамках описания класса взаимно однозначное соответствие между свойством и параметром, указанным в качестве значения свойства.
- 5) Вхождение шаблона F-класса в какое-либо RDF-высказывание (например, в состав толкования лексемы) равносильно ссылке на тело толкования в формате этого класса.

4 Структура толкований в лексиконе

При определении структуры описания лексического значения слова в лексиконе мы будем исходить из предположения, что в составе общей OWL-онтологии всегда имеется класс, к которому можно отнести определяемую лексему. Таким классом может быть либо класс, описанный стандартными OWL-средствами, либо F-класс, которому наряду со стандартными средствами описания сопоставлена также формула толкования. Выделив непосредственные описания смысла лексем в общую онтологию, мы в лексиконе можем сосредоточиться на определении условий выбора лексемы из набора возможных значений слова и на информации, поставляемой лексемой для решения задачи композиции семантической сети предложения и далее текста из фрагментов, поставляемых описаниями лексем.

Мы рассматриваем семантическую сеть предложения как результат трансформации синтаксической структуры этого предложения. Синтаксис предложения задается деревом сочинительно-подчинительных связей между его словами и словосочетаниями. Толкованиями лексических значений слов поставляются исходные заготовки, фрагменты семантической сети предложения. В процессе трансформации эти заготовки могут изменяться в соответствии с правилами объединения фрагментов в общую сеть [6]. В любом случае на фрагмент семантической сети толкования наложено следующее ограничение: фрагмент сети должен иметь единственный входной узел; входной и выходные узлы должны быть явно обозначены.

В Примере 3 показан фрагмент описания глагола ПЕРЕДВИНУТЬ. Фрагмент содержит две части: синтактику (описание валентности), в рамках которой выделены синтаксические актанты глагола

при его использовании в предложении, и семантику глагола в значении "передвинуть физический объект с одного места на другое". Синтактика примера (в несколько упрощенном виде) взята из словаря В.А. Тузова [7].

Пример 3.

ПЕРЕДВИНУТЬ

Синтактика:

Z1: !Им,Z2: !Вин,Z3: НЕЧТО\$1~!Откуда,Z4: НЕЧТО\$1~!Куда)

Семантика лексемы:

```
<interpret voc:input="_:x" voc:instance="_:x" >
  voc:ПЕРЕДВИНУТЬ rdf:subClassOf
    ont:СценарийПеремещения;
  _:x rdf:type voc:ПЕРЕДВИНУТЬ;
  ont:c_Агентом :Z1; ont:c_Объектом_:Z2;
  ont:рольИточка _:Z3; ont:рольКточка _:Z4;
  ont:рольНачало _:Y8; ont:рольКонец;_:Y9".
  _:Y8 rdf:type ont:НачалоСитуации.
  _:Y9 rdf:type ont:КонецСитуации.
</interpret >
```

Семантика лексемы в лексиконе оформляется в виде XML-контейнера, содержанием которого является текст RDF-графа, отражающего информацию, привносимую лексемой в семантическую сеть предложения. Атрибутами контейнера определяется входной узел графа. Набор его выходных узлов совпадает со списком параметров лексемы.

В толковании лексемы указываются:

- класс, подклассом которого является определяемая лексема (в Примере 3 – ont:СценарийПеремещения);
- параметры лексемы, увязываемые с ее синтактикой;
- возможные уточнения ограничений, накладываемых данной лексемой на параметры (в приведенном примере это указание на замкнутость интервала времени для ситуации описываемой лексемой).

В лексиконе каждому из лексических значений слова (его лексеме) сопоставляется собственная семантика. Синтактика может быть общей для нескольких лексем. Семантическими актантами лексемы принято называть список параметров ее толкования. В толковании набор семантических актантов лексемы совпадает со списком параметров ее класса, описанного в общей OWL-онтологии. В онтологии для каждого из параметров определен класс возможных значений и указано свойство (актантная роль) определяемое этим параметром. В составе описания лексемы упоминаются лишь те из ее актантных ролей, которые требуют согласования с ее синтактикой.

Таким образом, полная информация о семантике лексемы распределена по трем информационным ресурсам:

- общая OWL-онтология;
- форматы толкований;
- лексикон.

В задачу онтологии входит описание словаря основных понятий, в терминах которых

описываются смысловые значения слов естественного языка и обеспечение связи между описаниями лексики в лексиконе и форматами толкований.

Форматами толкований решается задача увязывания компонент толкования в единое целое. На этом уровне описаний лексики удается преодолеть статичность чисто онтологических описаний, отразить развитие ситуации во времени, определить пред- и постусловия реализации различных ситуаций. На этом же уровне достаточно обусловленным образом формируется список семантических актантов, свойственных той или иной ситуации. Формат дает схематическое описание смысла, образуя тем самым новое понятие, регистрируемое в онтологии как особый класс.

Литература.

- [1] OWL Web Ontology Language Guide. W3C Recommendation 10 February 2004 <http://www.w3.org/TR/2004/REC-owl-guide-2004021>.
- [2] RDF Primer. W3C Recommendation 10 February 2004 <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>.
- [3] Леонтьева Н.Н. Автоматическое понимание текстов: системы, модели, ресурсы. – М.: Изд. центр "Академия", 2006. 304 с.
- [4] Перцов Н.В. К проблеме построения семантического метаязыка. // Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции "Диалог 2006". – М.: Изд-во РГГУ, 2006. С. 419-425.
- [5] Кустова Г.И., Падучева Е. В. Словарь как лексическая база данных // Вопросы языкознания, М., 1994, N 4. С. 96-106
- [6] Падучева Е.В. Динамические модели в семантике лексики. – М. "Языки славянской культуры", 2004. 608 с.
- [7] Тузов В. А. Компьютерная семантика русского языка. – СПб.: Издательство СПб ГУ, 2004. 400 с.
- [8] Nirenburg S., Raskin V. Ontological Semantics. <http://crl.nmsu.edu/Staff.pages/Technical/sergei/book/>
- [9] Лезин Г.В. Онтологическое представление семантики предложений в системе языков RDF/OWL. //Языковая инженерия: в поисках смыслов. Доклады семинара "Лингвистические информационные технологии в Интернете". – СПб.: СПб ГУ, 2008. С. 5-21.
- [10] Падучева Е.В. Семантические исследования: семантика времени и вида в русском языке. Семантика нарратива. – М.: "Языки русской культуры", 1996.
- [11] Masolo C., Borgo S., Gangemi A, Guarino N., Oltramari A., Schneider L. DOLCE: a Descriptive Ontology for Linguistic and Cognitive Engineering // DOLCE documentation: <http://www.loa-cnr.it/DOLCE.html>.

- [12] Time Ontology in OWL. W3C Working Draft. <http://www.w3.org/TR/2006/WD-owl-time-20060927/>
- [13] RDF Semantics. W3C Recommendation 10 February 2004. <http://www.w3.org/TR/2004/REC-rdf-mt-20040210/>
- [14] Урынсон Е.В. "Несостоявшаяся полисемия" и некоторые ее типы // Семиотика и информатика.– М.: "Яз. рус. культ., Рус. словари", Вып. 36,1998. С. 226-261.
- [15] OWL Web Ontology Language. Semantics and Abstract Syntax. W3C Recommendation 10 February 2004. <http://www.w3.org/TR/2004/REC-owl-semantics-20040210/>

Ontological semantics of the text: formatting of interpretation in the semantic dictionary

Lezin G.V.

In article the problem of development of uniform representation of the information in a semantic network of the text, a lexicon and in a base ontology is considered. It is offered to add standard descriptions of classes in an ontology with formats of the multicomponent interpretation prepared for use in a lexicon. Semantics of formats and their co-ordination with an ontology and a lexicon is discussed.