

Разработка формата разметки и принципов аннотирования для семантического корпуса русского языка на материале «Маленького принца»*

Фролова Т. И.¹ и Рыгаев И. П.²

Лаборатория компьютерной лингвистики ИППИ РАН,
127051 Москва, Большой Каретный переулок, 19, стр.1
¹tfrolova@gmail.com и ²irygaev@gmail.com

Аннотация. В предлагаемой работе изложены основные принципы построения корпуса семантически размеченных структур русских текстов. Эти принципы включают в себя 1) установление соответствия между словами текста и узлами структуры, 2) сохранение в семантической структуре направления синтаксических связей, 3) заполнение валентностей предикатов, как явно, так и имплицитно заполненных в предложении. Кроме того, описана работа по автоматизации построения этого корпуса, перечислены основные группы правил, участвующих в построении семантической структуры. Работа над семантической разметкой ведется коллективом Лаборатории Компьютерной Лингвистики ИППИ РАН. Первая очередь корпуса составляет сказка «Маленький принц» в русском переводе. Объем этого текста около полутора тысяч предложений. Выбор переводного текста объясняется наличием семантических корпусов для аналогичного текста в английском переводе проектах UNL и AMR, что позволит легко сопоставить и взаимно обогатить разметку, выполняемую разными коллективами. Построение семантических структур опирается на результат синтаксической разметки, ведущейся в лаборатории уже более двадцати лет, результаты которой доступны на сайте русского национального корпуса. Таким образом, задача построения семантического корпуса с технической точки зрения сводится к построению правил преобразования синтаксической структуры в семантическую.

Ключевые слова: Корпусная лингвистика, семантика, модель «Смысл ⇔ Текст».

* Работа выполнена при поддержке гранта Министерства науки и высшего образования № 075-15-2020-793 «Компьютерно-лингвистическая платформа нового поколения для цифровой документации русского языка: инфраструктура, ресурсы, научные исследования».

1 Вводные замечания

В лаборатории компьютерной лингвистики ИППИ РАН ведется работа над созданием корпуса семантически размеченных текстов. Первая очередь корпуса представлена текстом перевода Норы Галь на русский язык сказки А. де Сент-Экзюпери «Маленький принц»; этот текст объемом около полутора тысяч предложений был выбран потому, что его английский перевод стал базой для семантической разметки в двух других проектах: разметка в рамках проекта UNL (<http://www.unlweb.net/wiki/LPP>) и проекта AMR (<https://amr.isi.edu/download.html>). Имея в распоряжении один и тот же текст, можно будет легко сравнить подходы и в дальнейшем, если понадобится, построить правила пересчета разметки одного корпуса в другой для взаимного обогащения корпусов с семантической разметкой. Текст сказки ранее был размечен синтаксически и включен в состав корпуса СинТагРус.

Построение семантически размеченного корпуса стало очередной частью семантического проекта, который идет в лаборатории уже несколько лет. Этапы развития этого семантического проекта отражены в работах [1-5]. На текущей стадии работа ведется в двух основных направлениях: во-первых, выработка принципов устройства семантических структур, во-вторых, обеспечение возможности построения этих структур в полуавтоматическом режиме. Разделы 1 и 2 ниже соответствуют этим направлениям работы.

2 Как устроены семантические структуры

Как и для других проектов лаборатории, теоретическим источником для построения семантического представления является теория «Смысл \leftrightarrow Текст» Мельчука [6] и теория интегрального описания языка Апресяна [7]. Семантическая структура строится для каждого предложения отдельно, связи между предложениями в текущей версии корпуса не отражаются. В последующих версиях разметки предполагается добавление этого типа информации.

В нашем проекте реализовано два уровня представления семантики: базовый и расширенный. Базовый уровень представляет смысл предложения. На расширенном уровне в структуру добавляются семантические разложения концептов и выводы. Подробнее о различиях между базовой и расширенной структурой см., например [8]. Семантически размеченный корпус состоит из базовых семантических структур.

Семантическое представление обоих уровней представляет собой граф, в узлах которого стоят семантические элементы, узлы связаны семантическими отношениями. В отличие от синтаксических структур к семантическим графам не предъявляется требование древовидности.

Семантические элементы представляют собой литералы (цепочки символов и числа) или индивиды. Индивиды бывают двух типов: некоторые из них это именованные индивиды, соотносящиеся с уникальными объектами в реальном мире.

Таков, например, индивид France, он соотносится с единственным объектом – Францией. Другие индивиды, анонимные, не представлены в онтологии, но соотносятся с концептами. Концепты – это классы сущностей: предметов, свойств или событий. Например, концепт Rose используется для описания класса роз. Если в предложении встречается слово «роза», то узлом, соответствующим этому слову в семантической структуре, будет индивид класса Rose (такой индивид обозначается в структуре названием класса с числовым индексом, например, Rose_1 или Rose_18). Концепт Seeing используется для описания класса событий видения. Если в предложении встречается глагол «видеть», то узлом, соответствующим этому глаголу в семантической структуре, будет анонимный индивид класса Seeing, не представленный в онтологии, а созданный «на лету» при анализе предложения (обозначается названием класса с индексом, Seeing_2 или Seeing_15). Концепты и именованные индивиды организованы в онтологию при помощи отношения «класс-подкласс» и «класс-индивид». Так, концепт Rose является подклассом концепта более высокого уровня FloweringPlant и наследует все свойства, которые могут быть приписаны надклассу. Индивид France представлен в онтологии как именованный индивид класса EuropeanNation и наследует свойства этого класса. Концепт Seeing является подклассом концепта более высокого уровня Perception и наследует все его свойства. Основные свойства концепта записываются в онтологии в виде списка отношений, характерных для него. Так, концепту Event, надклассу всех событийных концептов, приписано свойство иметь время, которое записывается при помощи отношения hasTime. Это свойство – иметь время – наследуется всеми подклассами класса Event. Валентности концепта также заданы отношениями. Концепту Perception, соответствующему классу событий восприятия, в онтологии приписаны свойства иметь носителя и стимул восприятия – hasExperiencer и hasStimulus, эти свойства наследуются всеми его подклассами, в том числе подклассом Seeing. Список семантических отношений, валентных и невалентных, задан в соответствующем разделе онтологии.

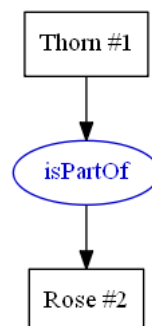
Существенно, что противопоставление концептов и отношений не соответствует противопоставлению имен и глаголов в языке. Классы событий, свойств и явлений, так же как и классы предметов соотносятся с концептами онтологии, а онтологические отношения описывают отношения между концептами.

Подробнее об устройстве онтологии см. [1].

В примере (1) записана базовая семантическая структура словосочетания «шип розы»:

(1) *Шип розы*

Thorn_1
isPartOf Rose_2



Запись Thorn_1 в этой структуре означает, что в структуре имеется индивид класса Thorn. Индекс обозначает, что узлом в структуре является индивид, соответствующий слову с соответствующим номером в предложении. Запись Rose_2 читается аналогично. Запись isPartOf означает, что между узлами Thorn_1 и Rose_2 проведено семантическое отношение с таким именем. В семантической структуре (как и в онтологии) принята запись названий классов с заглавной буквы, семантические отношения записываются со строчной.

Мы придерживаемся следующих принципов построения базовой семантической структуры: 1) по возможности сохранять соответствие между словами в предложении и узлами семантической структуры; 2) по возможности сохранять направление семантических связей, соответствующих синтаксическим отношениям; 3) заполнять валентности всех концептов, представленных в структуре, если эти валентности заполнены в предложении. Рассмотрим эти принципы подробнее.

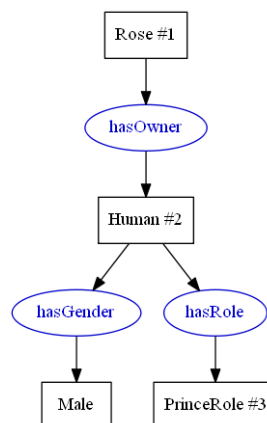
2.1 Соответствие между словами и узлами

Чаще всего каждому значимому слову в предложении соответствует один узел. Но существует несколько основных источников несоответствия между словами и узлами базовой семантической структуры.

Во-первых, для некоторых слов семантические разложения делаются уже на этапе построения базовой структуры, особенно если эти разложения регулярны и тривиальны. Так, в примере (2) ниже слово *принц* представлено в структуре подграфом из трех узлов: Human_2, Male и PrinceRole_3, соединенных отношениями hasGender и hasRole, который следует читать как «человек мужского пола, имеющий социальную роль принца». Из этих трех узлов только вершинный узел подграфа, Human_2, соотнесен в разметке со вторым словом в словосочетании. Остальным узлам присваиваются номера, следующие за номером последнего слова во фразе, и они не соотнесены в разметке ни с каким словом. Узел Male не имеет индекса в структуре, поскольку трактуется как именованный индивид из онтологии.

(2) *Роза принца*

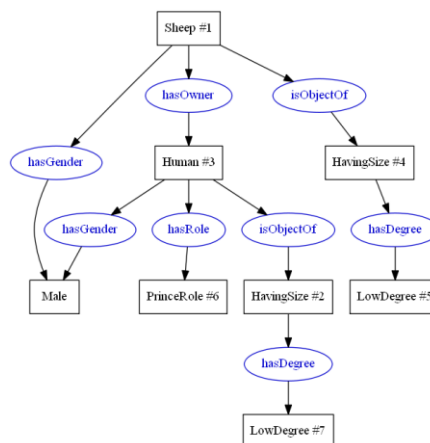
Rose_1
hasOwner Human_2
hasRole PrinceRole_3
hasGender Male



В примере (3) ниже каждое из трех слов в словосочетании представлено более, чем одним узлом. Слово *принц* представлено тем же подграфом, что и в примере (2), слово *маленький* представлено подграфом HavingSize_2 hasDegree LowDegree_9 ('иметь малое значение параметра «размер»'). слово *барашек* представлено подграфом из четырех узлов: Sheep_1, Male, HavingSize_4 и LowDegree_5. Можно заметить, что прилагательное *маленький* и уменьшительный компонент значения слова *барашек* в базовой семантической структуре представлены одинаково.

(3) *Барашек маленького принца*

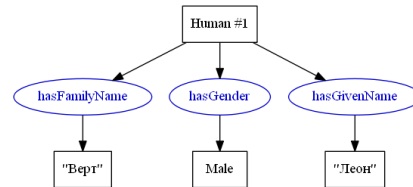
Sheep_1
hasGender Male
isObjectOf HavingSize_4
hasDegree LowDegree_5
hasOwner Human_3
hasRole PrinceRole_6
hasGender Male
isObjectOf HavingSize_2
hasDegree LowDegree_7



Похожее разложение можно видеть в примере (4), только в этом случае среди добавленных узлов есть не только индивиды онтологических классов, но и литералы. Литералы не соотносятся со словами предложения, то есть не имеют номера.

(4) *Леон Верг*

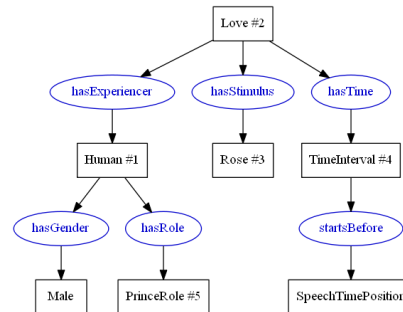
Human_1
 hasGender Male
 hasGivenName "Леон"
 hasFamilyName "Верг"



Кроме того, отдельными узлами в структуре обозначены значения времени, в том числе грамматические. В примере (5) прошедшее время глагола *любил* соответствует в структуре подграфу TimeInterval_4 startsBefore SpeechTimePosition, зависящему от узла, представляющего лексическое значение глагола, по отношению hasTime. Эту часть структуры можно прочесть как «имеет место на таком отрезке временной оси, который начался до момента речи». Узел SpeechTimePosition не имеет индекса, так как трактуется как именованный индивид из онтологии. Выработка стандарта представления временных и видовых значений находится в стадии активной разработки. Некоторые аспекты этой темы обсуждаются в [9].

(5) *Принц любил розу*

Love_2
 hasExperiencer Human_1
 hasRole PrinceRole_5
 hasGender Male
 hasTime TimeInterval_4
 startsBefore SpeechTimePosition
 hasStimulus Rose_3

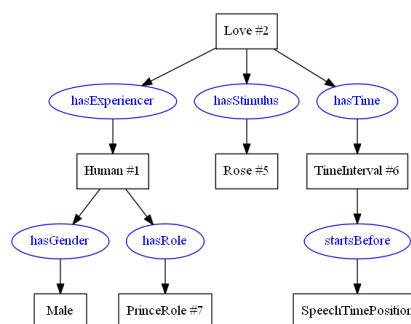


В примерах (2) – (5) были рассмотрены случаи несоответствия количества узлов и слов в предложении, происходящие из-за появления дополнительных узлов в структуре. Есть также случаи, когда источником несоответствия становится то, что слова предложения не соответствуют никакому узлу в семантической структуре. Так происходит с некоторыми служебными словами, например, во многих случаях предлоги не получают соответствующего им узла, можно сказать, что их смысл передается через отношение. Так, в примере (6) предлог К не имеет соответствия в структуре. Другой случай, когда слова в предложении не отражаются в структуре это лексико-функциональные сочетания со значениями лексических функций класса OPER-FUNC-LABOR. В примере (6) *испытывать любовь* является сочетанием слова аргумента ЛЮБОВЬ со значением лексической функции OPER1 от этого аргумента – ИСПЫТЫВАТЬ. Такое лексико-функциональное сочетание представлено в структуре одним узлом Love, соотнесенным по номеру с синтаксической вершиной сочетания, глаголом ИСПЫТЫВАТЬ. Подробнее о

лексических функций и работе с аппаратом лексических функций в проектах лаборатории см [10]. Таким образом, семантическая структура предложения (6) *Принц испытывал любовь к розе* совпадает со структурой предложения (5) *Принц любил розу* во всем, кроме номеров узлов. Порядковые номера слов 3 и 4 не приписаны никаким узлам.

(6) *Принц испытывал любовь к розе.*

Love_2
 hasExperiencer Human_1
 hasRole PrinceRole_7
 hasGender Male
 hasTime TimeInterval_6
 startsBefore SpeechTimePosition
 hasStimulus Rose_5



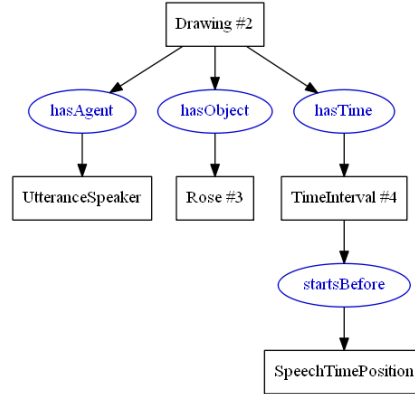
2.2 Сохранение направления связей

Семантические отношения между узлами в структуре имеют то же направление, что у синтаксического отношения между соответствующими словами. Для соблюдения этого принципа удобно пользоваться инверсными отношениями. Отношения в структуре могут записываться двумя способами, прямым и инверсным, так для `hasObject` инверсное отношение записывается как `isObjectOf`. Записи `A hasObject B` и `B isObjectOf A` эквивалентны.

Сравним связи между узлами `Drawing` и `UtteranceSpeaker` в структурах (7) и (8). В примере (7) стрелка `hasAgent` идет от узла `Drawing`, соответствующего слову *рисовал*, к узлу `UtteranceSpeaker`, соответствующему слову *я*, направление стрелки такое же, как у предикативного синтаксического отношения, связывающего узлы *рисовал* и *я* в синтаксической структуре. В примере (8) слово *рисующего*, соответствующее узлу `Drawing`, зависит от слова *меня*, соответствующего узлу `UtteranceSpeaker`, по определительному синтаксическому отношению. Семантическое же отношение между двумя узлами сохраняется, говорящий остается субъектом рисования, поэтому в структуре (8) используется инверсное отношение `isAgentOf`, таким образом направление связи в базовой структуре соответствует направлению синтаксической связи.

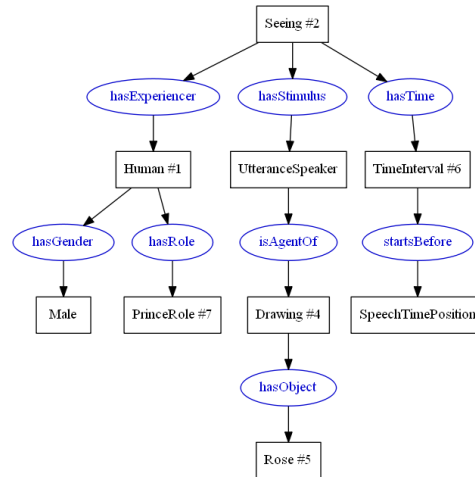
(7) Я рисовал розу

Drawing_2
 hasTime TimeInterval_4
 startsBefore SpeechTimePosition
 hasAgent UtteranceSpeaker
 hasObject Rose_3



(8) Принц видел меня, рисующего розу.

Seeing_2
 hasTime TimeInterval_6
 beforeSpeechTimePosition
 hasExperiencer Human_1
 hasRole PrinceRole_7
 hasGender Male
 hasStimulus UtteranceSpeaker
 isAgentOf Drawing_4
 hasObject Rose_5

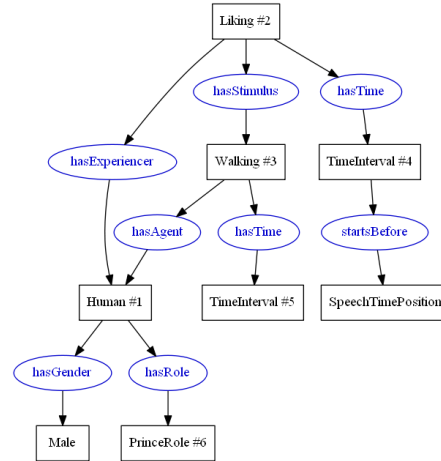


2.3 Заполнение валентностей предикатов, имеющих в предложении

Если в структуре встречается индивид класса, которому в онтологии приписаны какие-то валентности, и при этом из предложения ясно, чем эти валентности имплицитно заполнены, то их заполнение должно быть отражено в структуре. Рассмотрим пример (9). В предложении встречается слово *гулять*, в структуре этому слову соответствует индивид класса *Walking*. В онтологии для этого класса указано, что у него есть валентность, присоединяемая при помощи отношения *hasAgent*, соответствующая тому, кто гуляет. Из предложения ясно, что речь идет о прогулках принца. Значит, в базовой семантической структуре должна присутствовать стрелка *hasAgent* между узлом *Walking* и узлом *Human_1_1*, соответствующим слову *принц*, несмотря на то, что синтаксической связи между словами *гулять* и *принц* не устанавливается.

(9) *Принц любил гулять*

Liking_2
hasTime TimeInterval_4
beforeSpeechTimePosition
hasExperiencer Human_1
hasRole PrinceRole_6
hasGender Male
hasStimulus Walking_3
hasTime TimeInterval_5
hasAgent Human_1



3 Техника выполнения разметки

3.1 Общие сведения о технике разметки

Одной из целей текущей стадии проекта является достижение того уровня автоматизации разметки, который сейчас имеется для СинТагРус: сначала осуществляется первичная **автоматическая разметка**, затем проверка и при необходимости **корректировка вручную построенной структуры** лингвистом-аннотатором.

На вход блока построения семантической структуры подается предложение вместе с результатом синтаксической разметки в том же виде, в каком эта разметка присутствует в СинТагРус'е, то есть в виде дерева зависимостей, в узлах которого стоят слова, а ветви помечены именами синтаксических отношений. Подробнее о синтаксической разметке см [10]. Важно, что синтаксическая структура, которая служит отправной точкой для семантического анализа в силу подробности и разработанности содержит немало семантической информации. Так в словосочетаниях *вопрос принца* и *вопрос наследства* между первым и вторым словом проведены разные синтаксические отношения, в первом случае квазиагентивное, во втором случае первое комплетивное. Синтаксические правила устанавливают в этих случаях разные синтаксические отношения, соответствующие разным семантическим ролям при слове *вопрос*, опираясь на семантическую информацию в комбинаторном словаре ЭТАПа. В каком-то смысле можно сказать, что семантическая разметка не только опирается на синтаксическую, но и начинается еще на этапе построения синтаксической структуры (который в настоящей работе не рассматривается). Кроме синтаксической разметки, на входе семантического блока доступен результат лексико-функциональной разметки – во фразах выделены и связаны различными лексико-функциональными связями

лексико-функциональные словосочетания – и результат восстановления эллипсиса; подробнее об этих двух типах разметки также см. [10].

3.2 Автоматическая разметка

При обработке синтаксической структуры прежде всего восстанавливаются анафорические и кореферентные связи, причем не только связи между вершинами выражений, отсылающих к одному и тому же объекту, но и связи, ведущие к контролерам нулевых актантов, не выраженных в предложении. Так, восстановление такой связи в предложении (9) *Принц любил гулять* происходит сразу после синтаксического анализа, еще в терминах слов и синтаксических отношений, а не семантических узлов и семантических отношений. В рассматриваемом случае это значит, что проводится «квазисинтаксическое» предикативное отношение между словами *гулять* и *принц*, и лишь на более поздних этапах преобразования по правилам перевода предикативное отношение в данном лексическом контексте превращается в *hasAgent*, а слова *принц* и *гулять* заменяются соответствующими семантическими узлами. Описание правил такого восстановления нулевого субъекта приводится в [11]. Об автоматическом разрешении других типов анафоры в ЭТАПе подробнее см. [12]; об анафорической разметке см. также [10].

В начале основного этапа обработки работают правила порождения узлов, соответствующих временным грамматическим значениям и привязка этих узлов к моменту речи. В это же время производится первичная обработка лексико-функциональных сочетаний и некоторые перестройки структуры, например, в конструкциях с притяжательными прилагательными; так словосочетание с определительным синтаксическим отношением *Васина победа* превращается в сочетание *победа Васи*, где от *победа* к *Васи* проведено квазиагентивное синтаксическое отношение.

Затем следует основной этап перевода слов и отношений. Перевод слов может заключаться в замене слов на один узел или на несколько узлов, связанных семантическими отношениями. Так, слово РОЗА заменяется на один узел – индивид класса *Rose*, а слово ПРИНЦ заменяется на комбинацию из узлов *Human*, *PrinceRole* и *Male*, связанных соответствующими отношениями, подробнее см комментарий выше перед примером (3). Перевод валентных отношений зависит от слова и соответствующего ему семантического узла, поэтому не существует универсального правила перевода валентных отношений. Так, например, для глагола ВИДЕТЬ соответствующее правило переводит отношение с прямым дополнением как *hasStimulus*, а для глагола РУБИТЬ связь с прямым дополнением переводится как *hasObject*.

На заключительном этапе осуществляется дежурный перевод непереуведенных узлов и отношений там, где это необходимо, а затем стирание слов исходного предложения, оставшихся без перевода, и связанных с ними отношений. В первую очередь речь идет о вводных словах, которые не снабжены переводом на текущем этапе проекта. Так, во фразе *Итак, я исправляю посвящение* слово ИТАК остается без перевода и на заключительном этапе построения базовой структуры стирается вместе с идущим к этому слову вводным синтаксическим отношением.

Семантическая структура этой фразы будет совпадать со структурой фразы *Я исповляю посвящение*.

3.3 Формат хранения семантической разметки

Семантическая разметка сохраняется вместе с синтаксической в специальном XML-формате, известном как Tagged Text [13]. Чтобы это было возможно, формат был дополнен новыми тэгами.

Внутри тэга S (предложение) добавился тэг SEM. Сам этот тэг не содержит собственных атрибутов и служит лишь для группировки семантической структуры в одном месте и ограничения её от описания синтаксиса.

Тэг SEM внутри себя содержит теги N, каждый соответствует одному семантическому узлу. Тэг N имеет следующие атрибуты:

1. ID – номер узла. Для узлов, соответствующих словам, совпадает с номером слова (в тэге W). Для остальных – номер по порядку, начиная с номера последнего слова + 1.
2. TYPE – тип узла. Может содержать значения:
 - a. named – узел представляет собой именованный индивид, представленный в онтологии
 - b. anonymous – узел представляет собой индивид класса (концепта), представленного в онтологии. Это самый частый случай. Большинство узлов в семантической структуре обозначают анонимных индивидов.
 - c. literal – узел представляет собой числовое или символьное значение.
3. VALUE – значение узла. Для узлов типа named указывается имя индивида из онтологии. Для узлов типа anonymous – имя класса из онтологии, которому принадлежит индивид. Узлы типа literal могут быть двух видов – число (целое или вещественное) или строка символов в двойных кавычках.

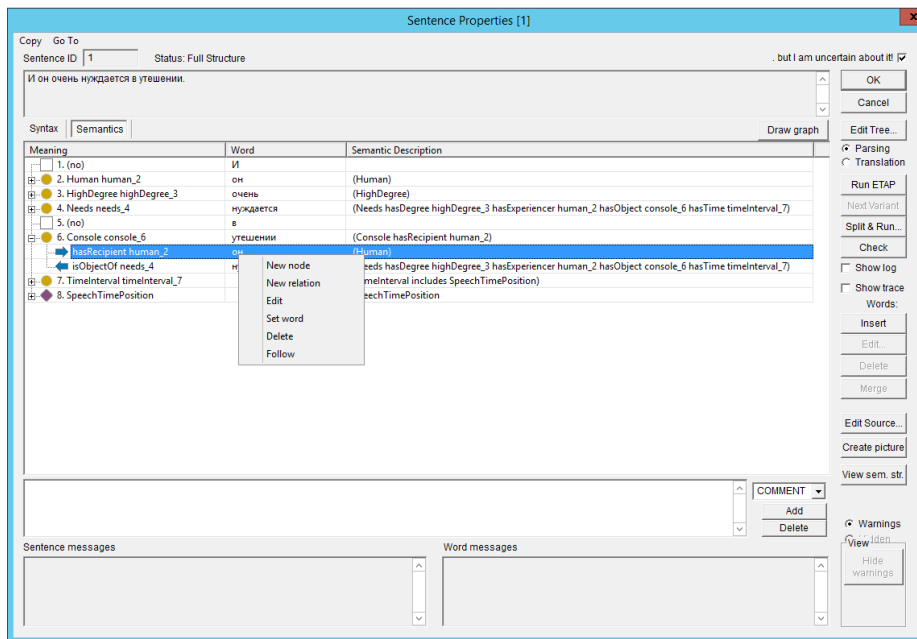
Тэг N внутри себя может содержать теги R, соответствующие отношениям, идущим от данного узла к другим узлам семантической структуры. Тэг R имеет два атрибута:

- a. LINK – имя отношения из онтологии
- b. TO – номер узла, в который направлено отношение

3.4 Редактирование структур

Для редактирования построенных автоматически семантических структур разработана новая версия редактора Structure Editor [13] с поддержкой семантической разметки. Она позволяет добавлять узлы в семантическую структуру, стирать их из структуры, проверять соответствие добавленных узлов их описанию в онтологии, а также проводить семантические отношения между узлами структуры и устанавливать или менять соответствия между словами предложения и элементами семантической структуры.

Ниже представлен пользовательский интерфейс редактирования предложения на примере предложения из текста «Маленького принца» *И он очень нуждается в утешении*. В одном и том же окне можно переключаться между синтаксисом и семантикой при помощи кнопок-переключателей слева вверху. Само окно семантической структуры отображает таблицу узлов, в которой представлено а) значение узла, б) слово, которому узел соответствует, и в) некое более подробное семантическое описание. Узлы можно раскрывать, чтобы увидеть, с какими другими узлами связан данный узел. Причём отображаются как исходящие, так и входящие отношения. Узлы помечаются разными значками в зависимости от их типа (а отношения – в зависимости от направления). Если слову не соответствует никакой узел, то в первой колонке вместо значения отображается (no) и белый квадрат в качестве значка. Семантическое описание в третьей колонке представляет собой выражение на языке Эталог [14], включающее в себя сам узел и исходящие отношения до его ближайших узлов, то есть до тех узлов, над которыми данный узел «доминирует».

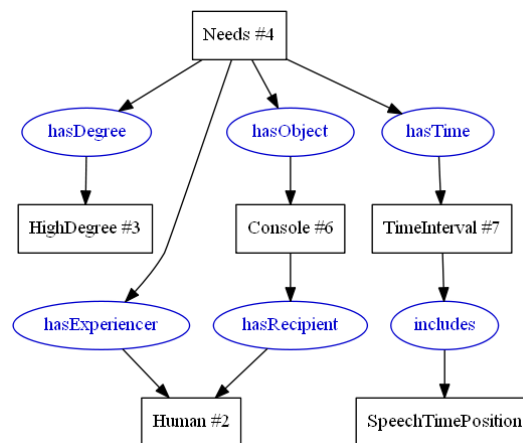


Само редактирование осуществляется при помощи вызова контекстного меню, элементы которого позволяют создавать новые узлы и отношения, редактировать у удалять существующие, а также менять привязку семантического узла к слову. Последний пункт “Follow” позволяет «перейти по стрелке» – он выбирает в таблице узел, на который указывает отношение.

Ниже представлены диалоги создания/редактирования узлов и отношений. При вводе проверяется наличие классов, индивидов и отношений в онтологии, также их можно выбирать из списка. Тип узла определяется автоматически.

А следующий диалог позволяет менять привязку семантического узла к слову. Тут можно выбрать другое слово, к которому пока не привязан никакой семантический узел, либо указать (по), чтобы отвязать узел от слова.

Кнопка “Draw graph” в правом верхнем углу позволяет отобразить семантическую структуру в виде графа. В частности, данная структура будет выглядеть так:



Кнопка “Run ETAP” в режиме Parsing пересоздаёт синтаксическую разметку автоматически, не трогая семантику. А в режиме Translation, наоборот, пересо-

здаёт семантическую разметку, не трогая синтаксис. Таким образом, второй режим можно использовать для добавления семантических структур в уже готовый синтаксически размеченный корпус с исправленными вручную синтаксическими структурами.

4 Заключение

В работе рассмотрены основные принципы и способы построения семантического корпуса в рамках проекта лаборатории компьютерной лингвистики ИППИ РАН. В перспективе корпус может быть использован для машинного обучения, автоматического анализа содержания текстов. Он является важным шагом в направлении построения модели понимания для искусственного интеллекта.

Список литературы

1. Богуславский И.М., Диконов В.Г., Тимошенко С.П. Онтология для поддержки задач извлечения смысла из текста на естественном языке // Информационные технологии и системы (ИТиС'12). Сборник трудов 35-ой конференции молодых ученых и специалистов ИППИ РАН (г. Петрозаводск, 19-25 августа 2012 г.). М.: ИППИ, 2012. С. 152-160.
2. Boguslavsky I.M., Dikonov V.G., Iomdin L.L., Timoshenko S.P. Semantic representation for NL understanding // Компьютерная лингвистика и интеллектуальные технологии. По материалам ежегодной Международной конференции «Диалог» (2013). М.: Изд-во РГГУ, 2013. Вып. 12. Т. 2. С. 124-136.
3. Boguslavsky I.M., Dikonov V.G., Iomdin L.L., Lazursky A.V., Sizov V. G., Timoshenko S.P. Semantic Analysis and Question Answering: a System Under Development // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (г. Москва, 27–30 мая 2015 г.). М.: Изд-во РГГУ, 2015. Вып. 14(21). Т. 1. С. 62-79.
4. Boguslavsky I.M., Frolova T.I., Iomdin L.L., Lazursky A.V., Rygaev I.P., Timoshenko S.P. Knowledge-based approach to Winograd Schema Challenge // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной международной конференции «Диалог» (г. Москва, 29 мая — 1 июня 2019 г.). М.: РГГУ, 2019. Вып. 18(25). С. 86-103.
5. Богуславский И. М. Семантический анализ с опорой на умозаключения в функциональной модели языка. *Вопросы языкознания*, 2021, 1: 29–56.
6. Мельчук И.А. Опыт теории лингвистических моделей «Смысл–Текст». М.: Школа «Языки русской культуры», 1999. 346 с.
7. Апресян Ю.Д. Интегральное описание языка и системная лексикография // Избранные труды. Т. II. М., Школа "Языки русской культуры", 1995, 766 с.
8. Boguslavsky I.M., Frolova T.I., Iomdin L.L., Lazursky A.V., Rygaev I.P., Timoshenko S.P. Semantic Analysis with Inference: High Spots of the Football Match (Семантический анализ с логическим выводом: острые моменты футбольного матча) // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной международной конференции «Диалог» (г. Москва, 30 мая — 2 июня 2018 г.). М.: РГГУ, 2018. Вып. 17(24). С. 124-142.

9. Timoshenko, S., Formal Representation of Temporal Expressions // *EPiC Series in Language and Linguistics*, 4, 2019. С. 84-94.
10. Иншакова Е.С., Иомдин Л. Л., Митюшин Л. Г., Сизов В. Г., Фролова Т. И., Цинман Л. Л. СинТагРус сегодня // *Труды Института русского языка им. В. В. Виноградова*. М., 2019. Вып.21. с.14–40.
11. Маракасова А.А. Автоматическое разрешение анафоры в русском тексте: случай нулевого субъекта // *Информационные технологии и системы 2016 (ИТиС'2016)*. Труды 40-й междисциплинарной школы-конференции ИППИ РАН. СПб., 2016. . С. 431-436.
12. Inshakova E.S. An anaphora resolution system for Russian based on ETAP-4 linguistic processor // *Компьютерная лингвистика и интеллектуальные технологии По материалам ежегодной международной конференции «Диалог» (2019)* Выпуск 18. С. 249-261.
13. Iomdin, L., Sizov, V. Structure Editor: A Powerful Environment for Tagged Corpora. *Research Infrastructure for Digital Lexicography (2009)*
14. Rygaev I. Etalog - a natural-looking knowledge representation formalism // *Труды школы-конференции ИТиС 2018*.