

Разрешение синтаксической местоименной анафоры в системе ЭТАП-3¹

Евгения Иншакова
ИППИ
inshakova@iitp.ru

Резюме

В данной статье идет речь о правилах установления антецедентов для двух классов анафорических выражений: возвратных местоимений и местоимений 3 лица.

1. Введение

Анафора — это отношение между языковыми выражениями (словами или словосочетаниями), состоящее в том, что в смысл одного выражения (анафора) входит отсылка к другому выражению (его антецеденту), которое, как правило, ему предшествует. Например, в предложении *Тридцать лет Люсьен_i жил со своей_i мамой_j, которая_j его_i любила и понимала* анафорические местоимения *своей* и *его* содержат отсылку к антецеденту *Люсьен*, а местоимение *которая* — к антецеденту *мамой*. Класс анафорических отношений частично совпадает с отношениями кореферентности (это отношение между фрагментами текста, которые соотносятся с одним и тем же объектом действительности, т.е. имеют один и тот же референт). Разрешение анафоры и установление кореферентных связей играют важную роль в автоматической обработке естественного языка, особенно после появления таких задач, как извлечение фактов из текста, когда требуется отследить всю цепочку упоминаний об одном и том же объекте, в том числе выраженных местоимениями и синтаксическими нулями. В настоящее время в мире существует большое количество систем разрешения анафоры и установления кореферентности, которые опираются на правила, использующие синтаксическую и семантическую информацию и знания о мире, и/или на методы машинного обучения (см. [1], обзор в [2]). В рамках этой задачи подготавливаются корпуса текстов с размеченными кореферентными отношениями (см., напр., [3], [4]).

С тех пор как в Лаборатории компьютерной лингвистики ИППИ РАН началась работа над проектом семантического анализа русского языка (см. [5]), стала весьма актуальной задача установления анафорических связей в процессе семантического анализа текста. В данной статье пойдет речь о разрешении двух видов местоименной анафоры: с возвратными местоимениями (СЕБЯ, СВОЙ) и местоимениями 3 лица. В настоящее время в системе ЭТАП поиск антецедента возможен только в пределах одного предложения, так что в нашей статье будет говориться только о синтаксической анафоре (действующей внутри предложения). Система разрешения дискурсивной анафоры (программный модуль для установления кореферентных связей между рассматриваемым предложением и одним или двумя предыдущими и соответствующая группа правил) на момент написания данной статьи находится в процессе разработки и не привязана к семантическому модулю системы ЭТАП.

Установление анафорических связей в современных работах, как правило, разбивается на несколько этапов (см., напр., [6], [7]). Сначала в тексте выделяются все именные группы — возможные кандидаты на роль антецедента для каждого местоимения, затем они пропускаются через ряд «фильтров» (правил, использующих семантическую информацию, элементы знаний о мире, расстояние от местоимения до

¹ Данная работа выполнена при поддержке гранта РФФ № 16-18-10422.

его антецедента и другие критерии). Каждый «фильтр» отсеивает неподходящих кандидатов, и в результате каждое местоимение получает один антецедент. Такого же принципа будем придерживаться и мы. В данной статье будут описаны правила, действующие на начальных этапах разрешения анафоры, в которых используется синтаксическая и семантическая информация, доступная в пределах одного или группы из двух-трех предложений. Правила выбора контролера местоимения из нескольких именных групп, одинаково подходящих на эту роль синтаксически и семантически, а также правила, использующие импликации и ожидания, являются предметом дальнейшей работы.

В нашей системе предусмотрены следующие стадии разрешения анафоры:

На этапе постсинтаксической обработки предложения:

- 1) установление антецедентов возвратных местоимений в финитных клаузах и именных группах (для правильного перевода на английский местоимений СЕБЯ и СВОЙ);

На этапе перевода русской поверхностно-синтаксической структуры в семантическую структуру:

- 2) восстановление синтаксических «нулей» — невыраженных субъектов инфинитивных, причастных, деепричастных оборотов и придаточных предложений, установление антецедентов относительных местоимений (см. [8]);
- 3) установление анафорических связей возвратных местоимений с восстановленными синтаксическими «нулями»;
- 4) установление антецедентов и/или кандидатов в антецеденты местоимений 3 лица в пределах одного предложения;
- 5) установление антецедентов и/или кандидатов в антецеденты местоимений 3 лица в пределах нескольких предложений;
- 6) выбор из нескольких возможных антецедентов (на данный момент не реализовано).

Кореферентные связи в системе ЭТАП выглядят следующим образом: все элементы кореферентной цепочки объединяются в гиперузел класса COREF-GROUP с порядковым номером, а (окончательно установленный) антецедент получает признак ANTEC (рис. 1). Если в предложении оказывается несколько кандидатов на роль антецедента, все они получают признак POS-ANTEC, который после прохождения ими правил-«фильтров» должен либо удаляться, либо заменяться на ANTEC.

Фраза	
1.1 Я	S, ИМ, ЕД, МУЖ, ОД, ANTEC, CAP
2.1 НЕ1	PART
3.1 ВИДЕТЬ	V, НЕПРОШ, ЕД, ЛИЧ, ИЗЪЯВ, 1-Л,
4.1 СЕБЯ[ТЧК]	S, РОД, ЕД, МУЖ, ОД
-----Hypernodes:-----	
5.1 FICT-HN (gr: COREF-GROUP-01)	
Групп кореферентности - 1	
Group info	
5 : COREF-GROUP-01	
1.1 Я,	
4.1 СЕБЯ	

Рисунок 1. Представление кореферентности в ЭТАПе (на примере *Я не вижу себя*)

На этапе русско-семантического перевода возвратное местоимение удаляется, а синтаксическая связь, ведущая к нему, либо перенаправляется к его антецеденту (если это местоимение СЕБЯ), либо переводится по словарному правилу в статью

местоимения СВОЙ, зависящего по определительной связи. Данное правило опирается на следующую иерархию синтаксических отношений, на которые может быть заменена эта связь: КВАЗИАГЕНТ или АГЕНТ (*своя победа, свой дом, свое звание генерала, свой анализ проблемы / анализ проблемы Ивановым*) > 1-КОМПЛ, выраженная род. п. (*свое обучение / обучение сотрудников*) > АТРИБ (*свой стол*). Правило содержит информацию о том, что существительное — хозяин местоимения СВОЙ: 1) либо управляет род. п. или тв. п. по 1-й валентности, 2) либо не управляет род. п. и тв. п. по 1-й валентности и управляет род. п. по 2-й валентности, 3) либо не управляет род. п. ни по 1-й, ни по 2-й валентности. Разные семантические структуры, получающиеся при переводе местоимения СВОЙ, показаны на рис. 2 и 3.

Translation

```

hasObject (Negation_1_1, Seeing_1_1)
hasExperiencer (Seeing_1_1, UtteranceSpeaker_1_1)
hasStimulus (Seeing_1_1, Nose_1_1)
isPartOf (Nose_1_1, UtteranceSpeaker_1_1)

```

Рисунок 2. Перевод на семантический язык предложения *Я не вижу свой нос*.

Translation

```

hasGivenName (Human_1_1, "Vasya"^^xsd:string)
hasAboutness (Remembering_1_1, Meeting_1_1)
hasExperiencer (Remembering_1_1, Human_1_1)
hasAgent (Meeting_1_1, Human_1_1)
hasCoagent (Meeting_1_1, Human_1_2)
hasGender (Human_1_2, Female_1_1)
hasGivenName (Human_1_2, "Masha"^^xsd:string)

```

Рисунок 3. Семантический перевод предложения *Вася вспомнил о своей встрече с Машей*.

2. Установление antecedентов возвратных местоимений

Возвратными (рефлексивными) называются такие местоимения, которые могут употребляться анафорически и требуют, чтобы их antecedент:

1) имел перед ними грамматический приоритет (был «выше» их в синтаксической иерархии, а согласно теории связывания — и в синтаксической структуре: например, подлежащее по ряду признаков «выше» прямого дополнения, а прямое дополнение «выше» предложной группы — косвенного дополнения, см. [9]);

2) находился в составе той же локальной области (некоторой синтаксической единицы: простого предложения, инфинитивного, причастного, деепричастного оборота или именной группы), в которую они входят сами (принцип А теории связывания, впервые сформулированный Н. Хомским).

Известно также: 1) что в русском языке рефлексивные местоимения ориентированы на подлежащее, т.е. могут быть кореферентны только подлежащему, а не, например, дополнению (*Петя_i отдал Васе_j свой_{*j/i} рюкзак*), 2) их локальной областью может быть финитная (возглавляемая личной формой глагола) клауза (см. [10]) или именная группа с собственным подлежащим — (квази)агентивной зависимой ИГ (*поступок Андрея, убийство царя террористами*) или притяжательным прилагательным в роли субъекта (*наша жизнь, Танина работа*). Поэтому в предложениях с возвратными местоимениями внутри инфинитивного оборота может возникнуть референциальная неоднозначность: *Он_i не разрешает мне_j [PRO_j производить опыты над собой_{i/j}]*, где местоимение *собой* может относиться к любому субъекту в данном предложении (выраженному *он* или невыраженному PRO), см. также примеры типа: *Официальный*

представитель_i греческой компании, PRO_i просивший [кого_i — не выражено] PRO_j не называть себя_i, сообщил «Известиям»...

Правила, устанавливающие antecedentes возвратных местоимений:

1. Трафаретные правила, применяющиеся на этапе постсинтаксической обработки структуры предложения: RA-REDUCT.14 — для связывания местоимения СЕБЯ с выраженным внутри именной группы antecedентом (*рассказы охотника_i о себе_i*); RA-REDUCT.20 — для связывания местоимения СЕБЯ с выраженным в предложении antecedентом (*Я_i вижу себя_i*); RA-REDUCT.25 — для связывания местоимения СВОЙ с выраженным внутри именной группы antecedентом (*рассказы охотника_i о своих_i подвигах*); RA-REDUCT.29 — для связывания местоимения СВОЙ с выраженным в предложении antecedентом (*Он_i любит свою_i работу*);

2. Общие правила, применяющиеся на этапе построения семантической структуры: SEM-CONV1.A3 — для связывания местоимения СЕБЯ с восстановленным нулевым подлежащим в причастных, деепричастных оборотах и у инфинитивов, субъект которых контролируется субъектом того предиката, от которого зависит инфинитив (*Вероника_i; некоторое время молчала, PRO_i сосредоточенно глядя перед собой_i; Я_i и правда стала PRO_i больше думать о себе_i*); SEM-CONV1.A4 — для связывания местоимения СВОЙ с восстановленным нулевым подлежащим в тех же конструкциях (*Преступник_i; отшвырнул поверженных храбрецов и убежал, PRO_i оставив на асфальте свой_i грабительский трофей; Россия_i не имеет права PRO_i не уделять этому направлению своей_i политики должного внимания*); SEM-CONV1.A5 — для связывания местоимения СЕБЯ с восстановленным нулевым подлежащим в инфинитивных оборотах, субъект которых контролируется дополнением того предиката, от которого зависит инфинитив (*Все реже просим редакции_i; PRO_i сообщить о себе_i; «что-нибудь новенькое»*); SEM-CONV1.A6 — для связывания местоимения СВОЙ с восстановленным нулевым подлежащим в тех же конструкциях, что и в SEM-CONV1.A5 (*Завершая свое выступление, я хотел бы просить вас_i; PRO_i высказать свои_i мнения и предложения*).

Перечисленные здесь правила содержат следующие условия:

- 1) лемму местоимения;
- 2) путь по синтаксической структуре от местоимения к ближайшему предикату (глагольной форме, существительному, находящемуся в вершине предложения прилагательному или наречию, далее — узлу Z), от которого оно непосредственно или опосредованно зависит;
- 3) путь от предиката Z к его существительному — его субъекту (предполагаемому antecedенту местоимения). Он может зависеть от Z по отношению ПРЕДИК или ДАТ-СУБЪЕКТ (если Z — глагол) либо КВАЗИАГЕНТ или АГЕНТ (если Z — существительное), как непосредственно, так и опосредованно, если находится внутри элективной (*один из нас*) или количественной конструкции (*много людей*);
- 4) типы синтаксических узлов, которые не могут встречаться на пути от местоимения к узлу Z: глаголы или существительные, не связанные с Z сочинительной связью, кроме причастий при глаголах-связках (*Преступник был пойман в своей квартире*);
- 5) семантическую категорию antecedента местоимения — дескриптор ‘ЛИЦО’, т.к. известно, что местоимения чаще всего ссылаются на одушевленные объекты (см., напр., [11]);
- 6) в правилах для нефинитных конструкций к этим ограничениям добавлены разные условия для субъектного (SEM-CONV1.A3, A4) и объектного контроля нулевого подлежащего (SEM-CONV1.A5, A6). В первом случае условие

исключает, что узел Z управляется предикатом с объектным контролем (*запрещать, просить*), во втором — требует этого. Это временное решение, принятое для того, чтобы в предложениях с объектным контролем можно было стереть кореферентную связь с подлежащим главной клаузы (*Учитель_i просил учеников_j PRO_j задавать себе_{i/j} вопросы*). Пока мы не можем строить для таких неоднозначных предложений две разные семантические структуры, будем оставлять в них только ближайший к местоимению антецедент.

В условиях правил RA-REDUCT.20, SEM-CONV1.A3 и A5 также содержатся списки конструкций, в которых не должна проводиться антецедентная связь, т.к. в них сочетания местоимений СЕБЯ и СВОЙ и их синтаксических «хозяев» идиоматизировались и местоимения уже не отсылают ни к какому объекту в действительности. Это выражения *представлять₂ себе, представлять₂ собой, вести₃ себя, чувствовать себя [как], приходить в себя, приводить₁ в себя, включать₁ в себя, уходить₁ в себя, выходить₁ из себя, выводить₄ из себя, строить из себя, содержать₁ в себе, нести₁ в себе, представлять₁ из себя, позволять себе, отказывать₁ себе, замыкаться₂ в себе, брать, забирать, нести₁, приносить, уносить, везти₁, привозить, увозить₁ с собой, брать на себя, напускать на себя, находить₁ себя, терять себя, изживать себя, посвящать₁ себя, усилие над собой, так себе, ничего₂ себе, не по себе, не в себе, само собой, сам по себе, себе на уме*, глаголы речи и мыслительной деятельности с сочетанием *про себя* ‘не вслух’, зависимость от глаголов по ограничительному отношению (*А он спит себе*), от прилагательных эстетической оценки (*хорош/нехорош/красив/некрасив собой*). В правила RA-REDUCT.25, 29, SEM-CONV1.A4 и A6 включен список аналогичных конструкций для местоимения СВОЙ: *в свою очередь, в свое время, в своем роде, в своем уме, своего рода, в большинстве своем, со своей стороны*.

3. Установление антецедентов местоимений 3 лица

Главное синтаксическое ограничение на антецедент местоимения 3 лица (принцип В теории связывания) заключается в том, что он не должен находиться в пределах той же локальной области, в которую входит местоимение (см. [9]). Если составить предложения, в которых местоимение 3 лица и его антецедент (не вложенный в другую именную группу) будут внутри такой синтаксической единицы, то их кореферентность будет невозможна (*Саша_i купил ему_{*i} (‘себе’) шапку; Маша_i хочет [PRO_i предложить ее_{*i} (‘свою’) кандидатуру]; Учитель попросил детей_i [PRO_i выбрать им_{*i} (‘себе’) темы для сочинения]; Петин_i рассказ о нем_{*i} (‘о себе’)*).

Однако в русском языке есть исключения из этого правила. Во-первых, местоимение 3 лица в инфинитивном обороте может быть кореферентно нулевому подлежащему инфинитива, подчиненного глаголу с объектным контролем: *Отец_j попросил Ивана_i PRO_i забрать его_{i/j/k} вещи из старой квартиры* (пример из [12]). Во-вторых, оно может быть кореферентно дативному субъекту в своей локальной области: *Ивану_i стало холодно в его_{i/j} машине* ([12]); в-третьих — субъекту событийного существительного в род. п. или тв. п.: *победа Петра_i над его_{i/j} новым соперником, чтение Маяковским_i его_{i/j} новой поэмы* ([12], [13]), в-четвертых — нулевому подлежащему PRO событийного существительного, подчиненного глаголу или существительному с объектным контролем: *Иван_i обвинял Петра_j в PRO_j убийстве его_{i/j/k} жены* ([13]). Это исключение действует, если в цепочке синтаксических связей между предикатом W и местоимением имеется хотя бы одно существительное, но не в тех случаях, когда местоимение подчинено предикату W непосредственно или через предлог (*недовольство Пети_i им_{*i}, победа Пети_i над ним_{*i}, Вася_i попросил Петю_j не говорить о нем_{i/*j/k}*). Нужно заметить, что данное исключение не является регулярным

(Она советует Фатиме_i забыть свое_i / ее*_i чувство и успокоиться; Рональд Филипс_i просит суд разрешить (ему_i) пожертвовать свои_i / его??_i органы нуждающимся матери и сестре). Что влияет на приемлемость или неприемлемость местоимений 3 лица в подобных конструкциях, пока неизвестно, и в наших правилах мы не будем запрещать в них антецедентную связь, но будем считать ее низкоприоритетной по сравнению с другими возможными антецедентными связями данного местоимения.

Известно также, что в русском языке местоимения 3 лица имеют антисубъектную ориентацию: они могут быть кореферентны любому аргументу в их локальной области, находящемуся «выше» и левее их, кроме субъекта (в им. п.): *Иван говорил с Машей_i о ней_i самой_i; Иван представил Машу_i ее_i новым соседям* ([14]). В то же время невозможны примеры типа *Иван показал Маше_i ее??_i саму_i*, и условия допустимости таких анафорических связей также до конца не выяснены.

Перед тем как устанавливать возможных контролеров местоимений, нужно выделить из числа всех именных групп такие нереферентные (реже — референтные) ИГ, к которым местоимения 3 лица вообще не могут отсылать. В литературе приводятся списки таких лексем и конструкций; мы будем отталкиваться от них, дополняя их и добавляя некоторые уточнения. Так, например, в [7] из множества возможных антецедентов исключаются проценты, денежные единицы, количественные значения разных параметров (*10 километров*). Но можно заметить, что при определенных условиях такие именные группы могут контролировать местоимения 3 лица (*2 процента горожан_i считают, что проблема наркомании их_i не касается, проехать на лыжах в общей сложности 982 км_i (из них_i 285 км с малокалиберным карабином за спиной), Прошло примерно 3 часа_i, но как долго они_i тянулись!*). Полный перечень контекстов, в которых количественные выражения нереферентны, сложно формализовать, и в нашем правиле будет упомянута только часть таких конструкций.

Правило, исключаящее определенные классы именных групп из множества возможных антецедентов местоимений 3 лица (SEM-CONV1.C2), приписывает существительным в этих ИГ признак NON-ANTEC. Оно содержит следующие условия: существительное X не имеет признаков NON-ANTEC, ANTEC, POS-ANTEC и оно:

- 1) либо зависит по отношению ПРИСВЯЗ (*Петя был студентом*), АППОЗ (*студент Петя*), ОБ-АППОЗ (*Петя, студент МГУ*), НУМ-АППОЗ (*дом номер восемь*) или КОМПЛ-АППОЗ (*вес (в) сто килограммов*), или само управляет чем-то по предикативному отношению (*Петя студент*), или зависит по второму комплетивному отношению от слова с синтаксическими признаками СОГЛАКТ-3-1, СОГЛАКТ-3-2 или СОГЛАКТ-2-1 (*репутация Петра как человека безукоризненной честности; признание Мадагаскара независимым государством; Иван притворился больным*). Исключение — зависящие по присвязочному отношению местоимения 3 лица в твор. п., которые могут отсылать к существительным с предикатным статусом (*Ивана принимали за директора_i, когда он им_i еще не был*);
- 2) либо относится к классу Event (события) или AbstractObject (в него входят, например, количества и временные интервалы) и при этом зависит от предлога, который является его значением лексической функции _ADV1-UN (*по аналогии, на бегу, в аспекте, с вероятностью*), _ADV2-UN (*по ошибке, под влиянием, под руководством*) или _LOC (*в апреле, на ветру*);
- 3) либо относится к классу Object (материальные объекты), зависит от предлога, который является его значением лексической функции _LOC и зависит по обстоятельственному отношению (*В деревне мы рано вставляли; В институте Ольга изучала право*), при этом существительное X не управляет указательными и относительными местоимениями и элективными

- прилагательными или номером по отношению НУМ-АППОЗ (*в этом институте; в институте, который он окончил; в доме номер восемь*);
- 4) либо относится к классу Event и зависит от предлога, который близок к значению лексической функции _ADV1-UN или _ADV2-UN, но не указан в этом качестве в словарной статье существительного X (список таких предлогов: *без, в1, в2, за1, за2, на1, на2, от, по1, по3, под2, при, с1, с3, в\$виде, в\$качестве, в\$течение, в\$ходе, вне1, во\$избежание, вроде1, наподобие, по\$мере, путем*. Примеры: *при наличии, с благодарностью, в присутствии, на вкус, по привычке, от нетерпения, по вине, без сожаления, с непривычки, в ответ, за неимением, на радость*);
 - 5) либо относится к классу AbstractObject и зависит от предлога, который близок к значению лексической функции _ADV1-UN или _ADV2-UN, но не указан в этом качестве в словарной статье существительного X (список таких предлогов: *в2, за1, на1, на2, по1, по2, с3, в\$течение, в\$ходе*. Примеры: *в количестве, за/на время, с января по май, со временем, на/по скорости, на днях*), при этом существительное X не управляет указательными, притяжательными и относительными местоимениями или существительными по отношению КВАЗИАГЕНТ (*в этом количестве, в вашем присутствии, с его благодарностью, с благодарностью Пети*);
 - 6) либо зависит по отношению ОБСТ (*Осенью она уехала; Они шли дворами*), ДЛИТЕЛЬН (*Весь день он работал*), КРАТНО-ДЛИТЕЛЬН (*Два раза они возвращались*), ДИСТАНЦ (*идти два километра*), ОБ-ОБСТ (*пить водку стаканами*) или АТРИБ в качестве твор. п. способа (*плавать брассом*), кроме случаев, когда непосредственно справа от существительного X — предложная группа *из них*, а справа от нее — числительное (*проехать на лыжах 982 км; из них; 285 км с малокалиберным карабином за спиной*);
 - 7) либо является количественным существительным с признаком ЧИСЛ-ОГР — *десяток, сотня, тысяча* и т.д. (*Призовой фонд равен миллиону; Билет стоит тысячу рублей*) и при этом не зависит от предиката из класса ChangeOfPossession (*Вася дал Пете тысячу_i рублей, и он ее_i потратил*) или не зависит от предиката с дескриптором 'ДЕЙСТВИЕ' и не управляет существительным из класса денежных единиц (*Вася потерял тысячу_i рублей, а Петя нашел их_i*);
 - 8) либо является именем параметра (*высота, скорость* и т.п.) и зависит от его лексической функции по отношению 1-КОМПЛ (*составлять, достигать, равняться, иметь в высоту* и т.д., *развивать скорость*), или ПРИСВЯЗ (*быть в возрасте*) или от существительного по атрибутивной связи (*стена высотой 10 метров, дети в возрасте от пяти лет, книги по цене от 100 рублей*);
 - 9) либо является именем единицы измерения и зависит по отношению 1-КОМПЛ от существительного или глагола со значением параметра (*вес в сто тонн, стоит пятьсот долларов*), предикатов *составлять1, достигать, равняться1, равный, длиться, вмещать*;
 - 10) либо является именем единицы измерения и зависит от предлога из списка: *в1, в2, до1, за1, к, на1, от, около, порядка, по1, под1, при, с1, с2, через, назад1, спустя, тому\$назад²*, который зависит по отношению ОБСТ или АППОЗ (примеры: *в 10 лет, в 2 метрах, до 6 часов, за час, к 5 часам, на 20 лет, порядка 50 %, по 2 штуки, под 90 лет, при 10 градусах, с 2 раз, человек*

² Этот перечень предлогов взят из работы [15].

с 10, 5 дней спустя, через год), или зависит от предлога *в*₁ или *на*₁, который зависит от прилагательного или наречия в сравнительной степени (*на два человека больше, в три раза меньше*), или является единицей измерения времени и зависит по отношению АТРИБ (*человек лет сорока*);

11) либо непосредственно или опосредованно зависит по отношению ВВОДН: *Проблема, как считает Иван, существует; Коля, закройте дверь*).

После завершения работы правила SEM-CONV1.C2 начинает применяться правило для выбора возможных антецедентов местоимения 3 лица (SEM-CONV1.C3). Оно включает следующие условия:

- 1) список лемм местоимений, к которым оно применяется (*он, она, оно, они, него₁, него₂, нее, них*);
- 2) зону поиска существительного Z — кандидата на роль контролера местоимения: слева от него по всему предложению (катафора пока не обрабатывается правилом), и его морфологическую категорию: либо существительное, либо прилагательное с признаком ЭЛЕКТ (*один из мальчиков*), либо притяжательное прилагательное (*знаменитый папин_i рассказ, как маршал Рокоссовский ему_i лично баян подарил*);
- 3) условие, что существительное Z не имеет признаков ANTEC, POS-ANTEC, NON-ANTEC;
- 4) морфологические характеристики местоимения и его потенциального контролера:
 - a. либо они согласуются по роду и числу;
 - b. либо местоимение *они/них* контролируется количественной группой с числительным *полтора / два / три / четыре*, и признаком ед. ч. у существительного (*два мальчика*);
 - c. либо местоимение *они/них* контролируется сочиненной группой (*Саши_i и Маша_j вместе учились, а теперь они_{i+j} вместе работают*), которая по принятым в нашей системе правилам представления множеств на семантическом языке уже должна быть переведена как концепт SetAnd, от которого по отношению hasElement зависят элементы множества;
Контексты, где расщепленный антецедент местоимения — две именные группы, являющиеся разными членами предложения (*Саши_i рассказал Маше_j о встрече с их_{i+j} школьной учительницей*), на данный момент правилом не обрабатываются;
- 5) путь по синтаксической структуре от местоимения к предикату W — глаголу, существительному, прилагательному или наречию, от которого оно зависит (зависимость непосредственная, через предлог и/или элективную конструкцию (*один из них*));
- 6) путь по синтаксической структуре от существительного/прилагательного Z (кандидата на роль контролера местоимения) к предикату U — глаголу, существительному, прилагательному или наречию, от которого оно зависит (зависимость непосредственная, через предлог и/или количественную конструкцию (*много людей, у десятка человек*)), а также редкие контексты, в которых Z зависит от того же предиката W, что и местоимение, которое в таких случаях управляет интенсификатором *сам* (*Иван говорил с Машей_i о ней_i самой*); То, что предикаты U и W, за редкими исключениями, не совпадают, обеспечивает соблюдение принципа В теории связывания в большинстве случаев (отсутствие кореферентной связи в контекстах типа *Саши_i видит / хочет видеть его_j*_i, рассказ Саши_i о нем_j*_i*).

- 7) проверку семантического согласования по модели управления для предиката W и потенциального антецедента местоимения (Z). Например, если местоимение зависит от W по предикативной связи, то нужно проверить, что Z тоже может занимать эту позицию, т.е. что в 1-м столбце модели управления предиката W содержится хотя бы один из тех семантических дескрипторов или онтологических концептов, которые имеются у Z в зоне дескрипторов. После того, как в онтологии *OntoEtap*, над которой ведется работа в нашей лаборатории (см. о ней [16]), достаточно большому количеству концептов (онтокоррелятов русских слов) будут приписаны их свойства (например, у концепта *Walking* ‘ходьба’ есть следующие свойства: *hasAgent some Animal*, *hasInstrument some Leg*, *hasStartingPoint some Region*, *hasTerminalPoint some Region*, *hasSpeed some FiniteQuantity*, *hasLocation some Region* и др.), появится возможность включить в данное правило проверку сочетаемости онтокоррелятов предиката W и контролера местоимения;
- 8) запрет на анафорическую связь в конструкции, где местоимение зависит от существительного W , его контролер (Z) является подлежащим клаузы или субъекта ИГ и при этом W и Z зависят от одного и того же глагольного или именного предиката U (непосредственно, через предлог и/или элективную конструкцию) (*Петя_i рассказал[U] некоторым из его_{j/*i} знакомых[W] о его_{j/*i} работе[W]; Петин_i разговор[U] с его_{j/*i} начальником[W]*). Это ограничение на еще не обработанные правилом конструкции, в которых действует принцип B ;
- 9) запрет на контроль местоимения одним из элементов сочиненной группы (*Петя_i и Вася_j гуляли с его_{k/*i/*j} собакой*);
- 10) запрет на то, чтобы предикат U управлял существительным W и при этом был значением его лексической функции *_OPER1*, *_INCEPOPER1*, *_FINOPER1*, *_OPER1-NS*, *_S0_OPER1*, *_OPER2*, *_OPER2-NS*, *_S0_OPER2*, *_REAL1-M*, *_INCEPREAL1-M*, *_ANTIREAL1-M*, *_MANIF*, т.е. функции с субъектным контролем существительного W (*Петя_i обратил_{OPER1} на него_{j/*i} внимание / был_{OPER1} от него_{j/*i} в восторге / получил_{OPER2} от него_{j/*i} выговор*).

Данное правило приписывает существительному, для которого выполняются все эти условия, признак POS-АНТЕС и связывает с ним местоимение кореферентной связью. Для антецедентов — сочиненных цепочек правило проводит кореферентную связь с концептом *SetAnd*.

Правило SEM-CONV1.C3 должно применяться циклически к одному и тому же местоимению, пока не будут отобраны все возможные кандидаты на роль его антецедента.

Правило, аналогичное SEM-CONV1.C3, но рассматривающее существительные из 1–2 предшествующих предложений, включает те же группы условий, кроме запретов на локальный контекст для местоимений (пункты 8 и 10 в описании правила SEM-CONV1.C3), т.к. данные запреты действуют только в пределах предложения.

5. Заключение

Работа над разрешением анафоры в системе ЭТАП находится, по сути, в начальной стадии. В частности, пока не решены следующие задачи:

- внесение анафорической разметки в корпус *SynTagRus*;
- разрешение дискурсивной анафоры и более сложных случаев синтаксической анафоры, чем описаны здесь;
- восстановление эллипсиса (например, для предложений типа *Вася_j может поднять свой_i вес, а Петя_i не может поднять свой_i вес*);

- установление антецедентной связи, если антецедент местоимения ОНИ/НИХ расщеплен (представляет собой несколько именных групп) или является кванторной группой (*Медвежонок_i прыгнул к Ёжику_j, они_{i+j} обхватили друг друга и так несколько мгновений стояли, трясясь; Понятно каждое слово в отдельности, но непонятно, что они означают в комбинации*).

Литература

- [1] Branco, A., McEnery, T. and Mitkov, R. (2005). Anaphora Processing: Linguistic, Cognitive And Computational Modelling: Selected Papers From Daarc 2002.
- [2] Mitkov, R. (1999): Anaphora resolution: the state of the art, Working paper, (Based on the COLING'98/ACL'98 tutorial on anaphora resolution), University of Wolverhampton, Wolverhampton.
- [3] Toldova, S.Ju. et al. Evaluating Anaphora and Coreference Resolution for Russian // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 4-8 июня 2014 г.). Вып. 13(20). М. : Изд-во РГГУ, 2014. Р. 681-695.
- [4] Schäfer, U., Spurk, Ch., Steffe, J.A. Fully Coreference-annotated Corpus of Scholarly Papers from the ACL Anthology. In: Proceedings of COLING 2012: Posters, pages 1059—1070.
- [5] Boguslavsky, I. Semantic Analysis based on linguistic and ontological resources In Proceedings of the 5th International Conference on Meaning-Text Theory (MTT'2011). Barcelona, September 8—9, 2011. P. 25-36.
- [6] Bogdanov, A. V. et al. Anaphora analysis based on ABBYY Comprendo linguistic technologies // Диалог-2014, т. 2.
- [7] Lee, H. et al. Stanford's Multi-Pass Sieve Coreference Resolution System at the CoNLL-2011 Shared Task. In Proceedings of the CoNLL-2011 Shared Task, 2011.
- [8] Маракасова А.А. Автоматическое разрешение анафоры в русском тексте: случай нулевого субъекта // Наст. сб.
- [9] Тестелец Я.Г. Введение в общий синтаксис. М., 2001.
- [10] Rappaport, G. On Anaphor Binding in Russian. In Natural Language and Linguistic Theory 4,1 (1986), 97-120.
- [11] Мальковский М. Г., Старостин А. С., Шилов И. А. Метод разрешения местоименной анафоры в процессе синтаксического анализа // Сборник научных трудов SWorld по материалам международной научно-практической конференции. 2013. Т. 11, № 4. С. 41—49.
- [12] Kazenin, K. I. 2000. Infinitives and Constraints on Pronominals in Russian. In *FASL: The Philadelphia Meeting 1999*, eds. T.H. King and I.A. Sekerina, 189–213.
- [13] Podobryaev, A. On the possibilities of anaphora in non-finite clauses: Evidence from Russian. In *LENCA III*, Tomsk State Pedagogical University, 2006.
- [14] Asarina, A. 2005. Russian binding theory: Two improved movement approaches. Ms., Massachusetts Institute of Technology.
- [15] Иншакова Е.С. Употребление аппроксимативной инверсии (по данным Национального корпуса русского языка): Дипломная работа. РГГУ, 2015.
- [16] Богуславский И.М. и др. Семантический анализ и ответы на вопросы: система в стадии разработки // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог». Вып. 14 (21): В 2 т. Т. 1: Основная программа конференции. — М.: Изд-во РГГУ, 2015, с. 62-79.