

Руководство по разметке кореференции в корпусе ru:corner

Анастасия Никифорова steysie@gmail.com	Сергей Терновых fostroll@gmail.com
---	---------------------------------------

Денис Киреев dkireev.71@gmail.com	Константин Ремизов mr.enslin@mail.ru
--------------------------------------	---

Декабрь 2020

Содержание

Глоссарий	3
1 Обзор корпуса	1
2 Виды упоминаний	1
2.1 Именные группы	2
2.2 Глагольные группы	3
2.3 Предикативы	3
2.4 Нереперентные именные группы	4
2.4.1 Денотация	4
2.4.2 Пустые множества	5
2.4.3 Определители	6
2.4.4 Нереперентные местоимения	6
2.4.5 Эксплетивные местоимения	7
2.4.6 Относительные прилагательные	7
2.4.7 Указательные местоименные наречия	8
3 Виды кереперентных связей	8
3.1 ANAPH/CATAPH	10
3.1.1 Тексты без именных упоминаний	11
3.1.2 Прямая речь	12
3.1.3 Притяжательные местоимения	12
3.1.4 Общие понятия	12
3.1.5 Местоимения <i>сам, сама, само</i>	13
3.2 IDENT	13
3.2.1 Собирательные именные группы	14
3.2.2 Копулярные конструкции	15
3.2.3 Связь копулярных конструкций с другими упоминаниями	15
3.3 APPOS	16
3.3.1 Выбор HEAD в аппозитивных конструкциях	17
3.3.2 Выбор между APPOS и CATAPH	18
3.3.3 Связывание аппозитивных цепочек с другими референтами	19
3.3.4 Денежные эквиваленты	20
3.3.5 Неаппозитивные конструкции	20

4	Отдельные случаи	20
4.1	Объединение именных групп в одно упоминание	20
4.2	Аббревиатуры	21
4.3	Даты, временные отсылки	21
4.4	Денежные суммы	22
4.5	Именованные сущности	22
4.6	Множества и подмножества	22
4.7	GPE и правительства	23
4.8	Количественные выражения, квантификаторы	24
4.9	Связь количественных выражений с другими упоминаниями . . .	25
4.10	Парцелляция и короткие предложения	26
5	Резюме	27
	Приложение А Направленные связи ANAPH/CATAPH	28
	Приложение В Ненаправленные связи IDENT	28
	Приложение С Направленные связи APPOS	28

Глоссарий

Ана́фора – лингвистическое явление, зависимость интерпретации некоторого выражения от другого выражения, обычно ранее встречавшегося в тексте. Это определяющее выражение называется **антецедент** (если оно появляется до анафорического выражения) или **постцедент** (если появляется после, в этом случае наблюдается катафора).

Ана́фор – упоминание, указывающее на другое упоминание в тексте и связанное с ним анафорической связью.

Глагольная группа – глагол с зависимыми словами.

Именная группа – имя существительное с зависимыми словами.

Именованная сущность (англ. *named entity*) – реальный объект, напр. люди, местоположения, организации, продукты и т. д., который может быть обозначен именем собственным (*город Москва, королева Великобритании Елизавета II, Google*).

Ката́фора – слово или выражение в тексте, более точное представление о котором можно получить только по мере дальнейшего чтения данного текста. Направление связи в катафоре противоположно анафоре.

Категория состояния (слова категории состояния) – класс слов, которые обозначают независимый признак, состояние (душевное, физическое или эмоциональное состояние человека, окружающей среды и природы) и не имеют форм словоизменения (склонения и спряжения), но могут с помощью глагола-связки выражать значение времени. Они обычно отвечают на вопрос *каков?* и используются, как правило, в качестве главного члена односоставных предложений (предикативного члена составного именного сказуемого).

Квантификатор (англ. *quantifier*) – логический определитель, например, *все, какой-то, никто, много, несколько, много* и др. (не конкретные цифры), который указывает на часть множества. Обычно фраза с квантификатором выглядит как “X (из) Y”, где X – квантификатор, а Y – именная группа в родительном падеже (*кого? чего?*).

Копула – специальное слово для связи подлежащего и сказуемого, выраженного не глаголом. Часто выражено формами глагола “*быть*” (который в настоящем времени выражен нулевой связкой или тире). Копулярными также могут быть глаголы *казаться, чувствовать, выглядеть, оставаться, становиться, оказаться* и т.п.

Кореференция (англ. *coreference / co-reference*) – вид текстовой или синтаксической связности, при которой два или более упоминания называют один и тот же объект (референт).

Предикативы – слова, обозначающие статическое состояние и выступающие в функции сказуемого (предиката) безличного предложения.

Референт – объект внеязыковой действительности, подразумеваемый автором конкретного речевого отрезка; предмет **референции**. Референт некоторого речевого отрезка может принадлежать не только реальному миру, но и воображаемому, например, быть персонажем художественного произведения.

Референция – отношение включенного в речь или текст упоминания к объекту действительности.

Синглтоны (англ. *singletons*) – упоминания, которые имеют референт, но они не кореферентны с другими упоминаниями в рамках одного контекста.

Упоминание (англ. *mention*) – именованная сущность, именная или другая группа (подробнее про виды упоминаний см. раздел 2), которая ссылается на референт, либо (анафор) на другое упоминание в тексте, имеющее референт.

1 Обзор корпуса

ru:corner – корпус русскоязычных текстов с разметкой именованных сущностей и кореференции. В этом руководстве приведены виды кореферентных связей, виды упоминаний (англ. *mentions*) и правила разметки кореференции в корпусе.

Корпус с разметкой именованных сущностей и кореференции позволит обучить модели для автоматического распознавания всех упоминаний одного и того же объекта в тексте.

Примеры кореферентной разметки в ru:corner:

- (1) Приехал к ним в гости [миллионер Константин Смирнов]_X и [[его]_{X-ANAPH} сын]_Y, [который]_{Y-ANAPH} в 24 года стал нобелевским лауреатом.
Референт X: миллионер Константин Смирнов, его.
Референт Y: его сын, который.
- (2) Но, вот, если говорить о персоналиях, например, то, вот, скажем, тот же [Александр Лукашенко]_X, [который]_{X-ANAPH} как-то периодически о [себе]_{X-ANAPH} напоминает, и [СМИ]_Y, на основании [которых]_{Y-ANAPH} [Интерфакс]_Z делает [свои]_{Z-ANAPH} рейтинги, периодически начинают тоже о [нем]_{X-ANAPH} вспоминать и активно комментировать [его]_{X-ANAPH} слова и действия.
Референт X: Александр Лукашенко, который, себе, нем, его.
Референт Y: СМИ, которых.
Референт Z: Интерфакс, свои.
- (3) Спасибо большое [Александру Куликовскому]_{X-HEAD}, [члену Общественного совета при ГУВД Москвы]_{X-ATTRIB}.
Референт X: Александру Куликовскому, члену Общественного совета при ГУВД Москвы

2 Виды упоминаний

Упоминания, называющие один и тот же объект действительности (референт), образуют кореферентную группу. Обычно упоминание представляет собой именную группу, однако время от времени в таком качестве может выступать глагольная группа, а также – в совсем уже редких случаях – предикативная часть составного сказуемого.

2.1 Именные группы

Именная (или номинативная) группа — это словосочетание, состоящее из имени существительного или местоимения и, если таковые есть, зависимых слов. Именная группа является *референтной*, если она отсылает к конкретному, реальному объекту, событию, действию, времени и т.п. Именные группы, имеющие один и тот же референт, объединяются кореферентной связью.

- (4) [В конце июня месяца]_x, [когда]_x [мы]_y делали в рамках [мониторинга] [первый доклад о готовности страны к введению в действие новой реформы]_z, [мы]_y в [докладе]_z отметили, что, к сожалению, [правительство] нарушило [закон].

В примере (4) видно, что не все именные группы участвуют в кореферентной разметке. Так, именные группы “мониторинг”, “правительство” и “закон” — *синглтоны*. В разметке `gi:corefer` синглтоны игнорируются.

Именные группы могут быть вложенными. В примере (4) именная группа [первый доклад о [готовности [страны] к [введению в действие [новой реформы]]]] включает в себя несколько вложенных именных групп.

В большинстве случаев, в кореферентной связи участвует внешняя именная группа, как в примере (4). Однако, в некоторых случаях, вложенные именные группы могут быть кореферентны с другими упоминаниями.

- (5) [Книги [Хэмингуэя]_x] интереснее всего читать, когда [он]_{x-ANAPH} описывает охоту.
- (6) [Кремль]_x всегда притягивал толпы туристов как [всемирно известная достопримечательность [Москвы]_y]_x. Тем не менее, в [городе]_{y-ANAPH} много и других знаменитых мест.

В примере (5) в кореферентную группу входит только вложенная именная группа, тогда как в примере (6) и внешняя, и вложенная именные группы связаны кореферентной связью с разными упоминаниями.

В `gi:corefer` учитывается не только внешняя именная группа¹, но и вложенные именные группы.

¹В отличие от, например, корпусов `OntoNotes` или `RuCor`, в которых вложенные именные группы игнорируются.

2.2 Глагольные группы

В качестве упоминаний могут быть размечены глагольные группы (глагол с зависимыми словами), если они кореферентны с именной группой в рамках одного текста.

- (7) Поставки труб из [Китая]_х в Россию [выросли на 790%]_у. Эксперты считают, что [рост]_{у-ANAPH} обусловлен государственной политикой [КНР]_х.
- (8) [В конце марта закрылись]_х музеи, театры, рестораны и прочие общественные места. [Закрытие]_х было связано с ростом эпидемии коронавируса.
- (9) Я решил начать [бегать трусцой по утрам]_х, потому что [это]_{х-ANAPH} поможет мне улучшить свое здоровье.
- (10) Я [решил начать бегать трусцой по утрам]_х, и рассказал об [этом]_{х-ANAPH} друзьям.

В примере (7) глагольная группа “*выросли на 790%*” и существительное “*рост*” кореферентны, поскольку оба называют одно и то же реальное, конкретное событие — рост поставок труб из Китая в Россию на 790%.

В примере (8) глагольная группа “*В конце марта закрылись*” и существительное “*закрытие*” кореферентны, они обозначают одно и то же конкретное событие — закрытие общественных мест в конце марта в связи с ростом эпидемии коронавируса.

Как и именные группы, глагольные группы могут быть вложенными. В примерах (9) и (10) сказуемое [решил [начать [бегать трусцой по утрам]]] состоит из трех глагольных групп. В кореферентной связи может участвовать любая из них, в зависимости от значения контекста. Например, в примере (9) местоимение “*это*” связано с глагольной группой “*бегать трусцой по утрам*” (потому что это полезно для здоровья), а в примере (10) — с глагольной группой “*решил начать бегать трусцой по утрам*” (и решил рассказать об этом решении друзьям).

2.3 Предикативы

Упоминаниями также могут быть предикативные части составных сказуемых², если они связаны с другими упоминаниями в тексте. При этом глагол

²составные сказуемые состоят из глагола-связки, или копулы, (часто это формы глагола “*быть*”) и предикатива с зависимыми словами

связка, или копула, в упоминание не входит. Ввиду особенностей русского языка, глагол-связка “*быть*” в настоящем времени опускается (для сравнения: “*she is pretty*” – “она (*есть*) красивая” vs. “*she was pretty*” – “она была красивая”).

- (11) Вчера было [очень холодно]_х, из-за [этого]_{х-ANAPH} сегодня на дорогах гололед.
- (12) Она была [самая красивая и умная из всех]_х и благодаря [этим качествам]_{х-ANAPH} победила в Мисс Мира.

2.4 Нереперентные именные группы

В текстах могут встречаться нереперентные именные группы, не участвующие в кореферентных связях с другими упоминаниями.

2.4.1 Денотация

Денотация — это словарное значение. Упоминания, носящие денотационный характер, обычно не используются в кореферентной группе, поскольку они обозначают не конкретный объект, событие, действие и т.п., а, как правило, общий класс объектов.

В приведенном ниже примере (13) оба упоминания “хирургии катаракты” просто обозначают тип операции, а не относятся к какому-либо конкретному случаю такой процедуры, и поэтому не могут быть связаны.

- (13) В «Три-З» [хирургия катаракты] возможна в один день с диагностикой: сама процедура длится в среднем всего около 7 минут, благодаря капельной анестезии пациент не испытывает неприятных ощущений, но находится в полном сознании. На следующий день – обязательный прием у врача, и на этом [хирургию катаракты] можно считать завершенной.

В примере (14) первое упоминание “львов” не может быть связано со вторым упоминанием, потому что оба относятся к львам как к типу существ, а не к каким-либо или даже всем конкретным экземплярам этого вида.

- (14) [Львы] свирепые хищники. Однако [львы] обычно не охотятся на людей.

Nota Bene! Исключением являются случаи, когда в контексте присутствует местоимение, которое указывает на общее понятие:

- (15) [Львы]_х свирепые хищники. Однако [они]_{х-ANAPH} обычно не охотятся на людей.

2.4.2 Пустые множества

Упоминания, обозначающие пустые или нулевые множества людей или объектов, не могут использоваться в кореференции. Чаще всего, это отрицательные местоимения. В приведенных ниже примерах (16), (17) кореферентная группа не может быть создана, поскольку “*никто*” обозначает пустое множество, а “*они*” относится к неопределенному набору людей, которые не были явно упомянуты.

- (16) [Никого] нет на лекции. [Они] все проспали.

- (17) Нас было пятеро, больше из наших [никто] не было, но я думаю, это потому что Катя [их] не пригласила.

Однако если упоминание состоит из неопределенного местоимения, которое служит отрицательным квантификатором, и именной группы, обозначающей множество лиц или объектов, — именная группа может участвовать в кореференции.

- (18) [Никто из [студентов]_х] не пришел на лекцию. [Они]_{х-ANAPH} все проспали.

- (19) Хотя [владельцы автобусов и маршрутных такси]_х исчисляются сотнями, самостоятельно не работает [никто из [них]_{х-ANAPH}].

В примере (18) множество “*студенты*” было явно упомянуто и, таким образом, оно может быть связано с “*они*”. Аналогичным образом, в примере (19) кореферентны именная группа “*владельцы автобусов и маршрутных такси*” и местоимение “*них*”. При этом целиком упоминания “*ни один из студентов*” и “*никто из них*” не участвуют в кореферентных связях, поскольку обозначают пустые множества.

Точно так же упоминания, которые управляются отрицательными глаголами или отрицательным словом *нет*, не могут использоваться в кореференции:

- (20) У моей сестры нет [машины].

- (21) Я не стал покупать [торт] на свой день рождения.

В обоих примерах, упоминания нереферентны, так как обозначают общий класс предметов. Однако если имеется в виду конкретный предмет из общего класса, он может быть кореферентен с другими упоминаниями:

- (22) Я не купил [тот торт из соседней кондитерской]_x, [который]_{x-ANAPH} так хотел попробовать.

2.4.3 Определители

Определители, такие как “каждый” и “любой”, обозначают неявный набор объектов. В приведенном ниже примере (23) “любая компания” не может быть кореферентна с местоимением “они”, поскольку первое упоминание может относиться к любой компании из некоторого ограниченного множества.

- (23) [Любая компания] может выиграть тендер. У [них] равные шансы.

- (24) [Каждый участник] получит утешительный приз, несмотря на то, победит [он] или проиграет.

Однако если выражение включает квантификатор (подробнее см. раздел 4.8), становится возможной кореференция:

- (25) [Любая из [компаний]_x] может выиграть Олимпиаду. У [них]_{x-ANAPH} равные шансы.

2.4.4 Нереферентные местоимения

Местоимения *ты, вы, мы* в обобщенно-личных предложениях не выделяются как упоминания. В таких предложениях подлежащее, выраженное личным местоимением, употребляется в значении обобщенного лица. Такой вид предложения распространен в устойчивых выражениях, поговорках, художественных описаниях или в гипотетических высказываниях. Например:

- (26) Проверила на себе золотое правило: «Сначала [ты] работаешь на зачетку, потом она на [тебя]».

- (27) Охотно [мы] дарим, что [нам] не надобно [самим].

- (28) [Вы] входите в лес. [Вас] сразу охватывает прохлада. [Вы] медленно идете по опушке. [Вы] бросаетесь на землю, [вы] напились, но [вам] лень пошевелинуться. [Вы] в тени, [вы] дышите пахучей сыростью; [вам] хорошо, а против [вас] кусты раскаляются и словно желтеют на солнце.

- (29) Допустим, [вы] воспользовались услугами рейсового автобуса или маршрутного такси, и у [вас] возникли какие-то проблемы с перевозчиком.

Во всех этих случаях дается описание состояния и действий вообще, любого человека, без отнесенности к конкретному лицу.

2.4.5 Эксплетивные местоимения

Эксплетивные местоимения не имеют значения в рамках контекста и используются для заполнения в предложении какой-либо синтаксической позиции или акцентирования фраз:

- (30) Подмосковье - [это] одна из самых красивых частей света, просто не изученных, на мой взгляд.
- (31) Дмитрий пишет, что в тюрьме и так сидеть не сладко, а [тут] еще моральные и физические угрозы.
- (32) Ну как же, да вспомните любую программу, рекламу [там], не только пива, но и сигарет.

2.4.6 Относительные прилагательные

Относительные прилагательные — это слова, которые обозначают постоянный признак через отношение к другому предмету (*стеклянная чашка — чашка из стекла, осенний день — день осени*). Относительные прилагательные не выделяются как упоминания, несмотря на неявную отсылку к референту.

В том числе **не являются упоминаниями и не входят в кореферентные группы** относительные прилагательные, отсылающие к именованным сущностям:

- (33) Во многих произведениях [Пушкина] встречаются диалоги на французском. Не сомневаюсь и имею доказательства, что пушкинская проза дает яркие отражения французского языкового строя.
- (34) Провал расчетов австралийской реакции, связанных с референдумом, означает, что большинство народа [Австралии] выступает против политики войны.

2.4.7 Указательные местоименные наречия

Указательные местоименные наречия (*там, туда, оттуда, сюда, здесь, так, тогда, потому, поэтому, затем*) выделяются как упоминания, если указывают на конкретное упоминание в тексте.

(35) Я вырос в [Новосибирске]_X, но уже с детства [там]_{X-ANAPH} не был.

(36) Он напрямик пошёл в [спальню]_X. [Оттуда]_{X-ANAPH} раздалось бормотание.

(37) Многие [здесь]_{X-CATAPH}, в [эфире Эха Москвы]_X, говорят, что...

(38) Это [биологический препарат]_X, [безопасный]_Y, [поэтому]_{Y-ANAPH} [его]_{X-ANAPH} можно применять с любого возраста.

3 Виды кореферентных связей

В ru:corner используется три вида кореферентных связей: анафора/катафора (ANAPH/CATAPH), идентичные (IDENT, кореференция) и аппозитивные (ARPOS, приложения). В рамках одного текста могут встречаться и пересекаться разные виды связей.

Местоимения, указывающие на какое-либо упоминание в тексте, отмечаются с этим упоминанием связью ANAPH (если упоминание-антецедент употребляется до местоимения) или CATAPH (если местоимение употребляется до упоминания-постцедента). Связь направлена от местоимения на антецедент/постцедент. В **приложении А** изображены связи ANAPH/CATAPH в виде направленных стрелок для соответствующих примеров.

Связь ANAPH всегда направлена влево, CATAPH – вправо.

(39) Она выдвинула [хорошее предложение]_X, [его]_{X-ANAPH} все поддержали.
ANAPH X: хорошее предложение $\xleftarrow{\text{ANAPH}}$ его

(40) За несколько недель до [его]_{X-CATAPH} смерти, [отец]_X завещал мне [свой]_{X-ANAPH} дом.
CATAPH X: его $\xrightarrow{\text{CATAPH}}$ отец
ANAPH X: отец $\xleftarrow{\text{ANAPH}}$ свой

(41) [ФБР]_X на запросы никак не отвечало. То есть [они]_{X-ANAPH} изначально были готовы и ориентированы именно на арест и на [публичное дело]_Y, [которое]_{Y-ANAPH} [они]_{X-ANAPH} подгоняли к

[саммиту]_Z, [который]_{Z-ANAPH} проходил в [Хельсинки]_W - [там]_{W-ANAPH}
 Трамп и Путин встретились в первый раз. То есть [им]_{X-ANAPH} не
 нужны были доказательства. [Им]_{X-ANAPH} нужно было [публичное
 дело]_Y.

ANAPH X: ФБР $\xleftarrow{\text{ANAPH}}$ они, они, им, им

IDENT Y: публичное дело, публичное дело

ANAPH Y: публичное дело $\xleftarrow{\text{ANAPH}}$ которое

ANAPH Z: саммиту $\xleftarrow{\text{ANAPH}}$ который

ANAPH W: Хельсинки $\xleftarrow{\text{ANAPH}}$ там

Именные сущности, именные группы и местоименные упоминания одного и того же объекта, концепта или события связываются кореферентной связью **IDENT**. Нет никаких ограничений на то, какие семантические классы могут рассматриваться для этого вида кореференции. Связь IDENT ненаправленная, все члены цепочки равноправны.

(42) Во вторник 27 ноября у [Романа Ротенберга]_X была назначена очень важная встреча в [Кремле]_Y. [Там]_{Y-ANAPH} [главе штаба национальной сборной России по хоккею]_X вручили орден Дружбы.

IDENT X: Романа Ротенберга, главе штаба национальной сборной России по хоккею

ANAPH Y: Кремле $\xleftarrow{\text{ANAPH}}$ там

Аппозитивные конструкции (**APPOS**) состоят из двух или более соседних именных групп, разделенных только запятой, двоеточием, тире или круглыми скобками. Одна из именных групп – это референт (HEAD), который указывает на конкретный объект, а соседние именные группы, приложения, определяют один или несколько атрибутов (ATTRIB) этого референта, таким образом переименовывая или дополнительно определяя основной референт. Связь APPOS направлена от ATTRIB к HEAD.

Вся аппозитивная конструкция целиком является отдельным упоминанием и может входить в IDENT цепочки или быть антецедентом или постцедентом в направленных связях ANAPH/CATAPH.

(43) [[Джон]_{X-HEAD}, [мой коллега]_{X-ATTRIB}]_X, приедет к нам на ужин.

[Он]_{X-ANAPH} работает юристом. [Джон]_X давно не был в гостях.

APPOS X: Джон $\xleftarrow{\text{APPOS}}$ мой коллега

ANAPH X: Джон, мой коллега $\xleftarrow{\text{ANAPH}}$ он

IDENT X: Джон, мой коллега; Джон

3.1 ANAPH/CATAPH

Местоимения или именные группы, указывающие на другие упоминания в тексте, связываются с ними связью ANAPH (если упоминание-антецедент употребляется до местоимения) или CATAPH (если местоимение употребляется до упоминания-постцедента). Связь ANAPH всегда направлена влево, CATAPH – вправо.

Важно отметить, что в приоритете анафор связывается с **ближайшим** предшествующим упоминанием связью ANAPH. Если в предшествующем контексте нет упоминания, на которое указывает анафор, он связывается с ближайшим последующим упоминанием связью CATAPH. Как правило, катафора встречается в текстах редко.

- (44) [Он]_{X-CATAPH} один из самых влиятельных людей 2020 и не нуждается в представлении. [Джефф Безос]_X основал [Amazon]_Y в 1994, и теперь [[его]_{X-ANAPH} компания]_Y является одной из крупнейших в мире.

CATAPH X: он $\xrightarrow{\text{CATAPH}}$ Джефф Безос

ANAPH X: Джефф Безос $\xleftarrow{\text{ANAPH}}$ его

IDENT Y: Amazon, его компания

- (45) [Барбра Стрейзанд]_X привыкла откровенно общаться с публикой. Вот и вчера во время своего шоу [она]_{X-ANAPH} поделилась интересными фактами биографии. В частности, [актриса и певица]_{X-ANAPH} припомнила слухи о том, что у [нее]_{X-ANAPH} был роман с принцем Чарльзом.

ANAPH X: Барбра Стрейзанд $\xleftarrow{\text{ANAPH}}$ она, актриса и певица, нее

Обычно в качестве анафоров выступают местоимения, однако анафором также может быть именная группа, которая имеет конкретный референт только если связана с референтной именной группой, как “*актриса и певица*” в примере (45).

Однако если упоминание состоит из местоимения-анафора и именной группы, как в примере (45), считается, что такое упоминание имеет конкретный референт: “*его компания*” \Rightarrow “*компания Джеффа Безоса*”. Такое упоминание (“*его компания*”) в свою очередь может быть частью кореферентной группы IDENT или быть антецедентом/постцедентом.

Заметим, однако, что это верно только если существует лишь одна общеизвестная компания, принадлежащая Джеффу Безосу. Если бы таких компаний было две или более (то есть нет конкретного референта), упоминание “*его компания*” было бы анафором.

3.1.1 Тексты без именных упоминаний

В некоторых текстах в качестве упоминаний могут встречаться только местоимения³, без конкретных референтов.

Если в тексте употребляются личные местоимения в начальной форме (*я, ты, он и т.д.*), и их конкретный референт в контексте не указан, такое местоимение рассматривается как референтное упоминание и может входить в IDENT, ANAPH/CATAPH или APPOS связи с другими упоминаниями:

- (46) Эта скамья [моя]_{X-CATAPH}. [Я]_X сижу здесь каждый день.

CATAPH X: моя $\xrightarrow{\text{CATAPH}}$ Я

- (47) [Ты]_X даже представить не можешь, как я рад [тебя]_{X-ANAPH} видеть!

ANAPH X: Ты $\xleftarrow{\text{ANAPH}}$ тебя

- (48) [Он]_X спал глубоко и спокойно, но ровно через 20 минут [он]_X проснулся.

IDENT X: Он, он

Если в тексте нет личных местоимений в начальной форме и нет референта, все последующие местоимения следует связать как ANAPH с первым местоимением, указывающим на тот же объект действительности. При этом в интерфейсе разметки подобные безреферентные цепочки упоминаний следует обозначить специальным атрибутом NoRef. Такое может происходить, в частности, если предыдущий контекст обрезан.

- (49) Родственники лично выразили [ему]_{*X} признательность за искрометный и обаятельный образ, оценив [его]_{*X-ANAPH} актерский дар, и не придирались к цвету [его]_{*X-ANAPH} формы и нашивкам!

ANAPH *X: ему $\xleftarrow{\text{ANAPH}}$ его, его

- (50) А потому и не могут, что [им]_{*X} по большому счету начхать на [фирму [босса]_{*Z}]_{*Y}, [она]_{*Y-ANAPH} же [его]_{*Z-ANAPH} а не [их]_{*X-ANAPH}...

ANAPH *X: им $\xleftarrow{\text{ANAPH}}$ их

ANAPH *Y: фирму босса $\xleftarrow{\text{ANAPH}}$ она

ANAPH *Z: босса $\xleftarrow{\text{ANAPH}}$ его

³Все местоимения – дейктические, т.е. имеют референт только в рамках контекста.

3.1.2 Прямая речь

Все местоимения должны быть связаны со своими референтами, даже если они встречаются в прямой речи, как в примерах (51) и (52).

- (51) [Борис Джонсон]_X представил [[свой]_{X-ANAPH} план “зеленой промышленной революции”]_Y. «[[Мой]_{X-ANAPH} план]_Y поможет создать тысячи “зеленых” рабочих мест и поможет прийти к нулевым выбросам в 2050 году», – заявил [британский премьер-министр]_X. [Он]_{X-ANAPH} отметил, что [амбициозный план]_Y должен «изменить жизнь в Великобритании».
- IDENT X: Борис Джонсон, британский премьер-министр
ANAPH X: Борис Джонсон $\xleftarrow{\text{ANAPH}}$ свой
ANAPH X: британский премьер-министр $\xleftarrow{\text{ANAPH}}$ он

- (52) Вот однажды, не помню кому, [товарищ Сталин]_X сказал: «У [меня]_{X-ANAPH} нет для вас другого народа».
- ANAPH X: товарищ Сталин $\xleftarrow{\text{ANAPH}}$ меня

3.1.3 Притяжательные местоимения

Притяжательные местоимения (ваш, мой, наш, твой, свой, его и т. д.) должны быть связаны с владельцем, даже если они также содержат неявную ссылку на предмет.

- (53) На вопрос, как появление ребенка изменило [его]_{X-SATAPH} жизнь, [Купер]_X ответил, что, безусловно, сильно.
- SATAPH X: его $\xrightarrow{\text{SATAPH}}$ Купер

3.1.4 Общие понятия

Как отмечено в разделе 2.4.1, общие или собирательные понятия, не имеющие конкретного референта, могут участвовать в кореферентной связи только если в тексте есть местоимение, которое на него указывает.

- (54) Если [родители]_X хотели отдать ребёнка в школу с шести лет, [они]_{X-ANAPH} выбирали программу «1-4», [родители] же семилетних первоклассников – программу «1-3».
- (55) Ещё в XVII и XVIII вв. известные мастера-мебельщики реставрировали [старую мебель]_X, нередко при этом переделывая [её]_{X-ANAPH} в соответствии с требованиями моды своего времени.

В примере (54) существительное “*родители*” не обозначает конкретных людей, на него указывает местоимение “*они*”. Стоит отметить, что второе упоминание родителей не входит в кореферентную группу, так как обозначает другое множество родителей.

Аналогичным образом в примере (55) собирательное именная группа “*старая мебель*” обозначает мебель в целом, не конкретный объект этого класса. Однако в контексте есть местоимение “*её*”, которое на него указывает, поэтому они объединяются связью ANAPH.

3.1.5 Местоимения *сам, сама, само*

Местоимения *сам, сама, само* соединяется связью ANAPH/CATAPH с упоминанием, на которое указывает:

(56) [Он]_x сделал это [сам]_{x-ANAPH}

(57) [Она]_x [сама]_{x-ANAPH} не поняла, в чем суть

Исключением являются фразы “*сам за себя, само по себе, сам про себя, само собой*” и аналогичные им фразы, а также “*сам*” в значении “*тот самый, всем известный*”, как в примере (62). Они не выделяются как отдельные упоминания, даже если они разорваны, как в примере (61).

(58) Само по себе оружие не более как кусок металла.

(59) В этом мире ничто не происходит само собой.

(60) Этот факт говорит сам за себя.

(61) Нужно самим принимать решения за себя.

(62) Ди Каприо противостоит в роли агента ФБР сам Том Хэнкс.

3.2 IDENT

Тип связи IDENT используется для разметки кореференции между упоминаниями конкретных референтов, взаимозаменяемых без потери смысла.

(63) Второй день пребывания в Марокко у [Меган Маркл и принца Гарри]_x начался с поездки в школы в городе Асни. [Пара]_{x-ANAPH} познакомилась с воспитанниками школ и их учителями, которые

провели [герцогу и герцогине Сассекским]_X экскурсию по школе и классам.

ANAPH X: Меган Маркл и принца Гарри $\xleftarrow{\text{ANAPH}}$ Пара

IDENT X: Меган Маркл и принца Гарри, герцогу и герцогине Сассекским

В примере (63) упоминания “Меган Маркл и принц Гарри” и “герцог и герцогиня Сассекские” обозначают один и тот же конкретный референт, тогда как существительное “пара” вне контекста может обозначать любую пару объектов, а в контексте относится к упоминанию “Меган Маркл и принц Гарри”, поэтому является его анафором.

Кореференция IDENT не включает общие или собирательные понятия и нереферентные упоминания (см. раздел 2.4).

3.2.1 Собирательные именные группы

Важно, что повторения собирательных именных групп в рамках одного контекста не могут быть включены в одну кореферентную группу IDENT. Поэтому каждый раз, когда в тексте встречается собирательная именная группа, на которую указывает местоимение, начинается новая цепочка ANAPH:

- (64) [Родители]_X должны участвовать в обучении [своих]_{X-ANAPH} детей дома, а не в школе. [Они]_{X-ANAPH} должны следить за тем, чтобы [их]_{X-ANAPH} дети не прогуливали занятия; [они]_{X-ANAPH} должны следить за тем, чтобы дети уделяли достаточно времени домашнему заданию; [им]_{X-ANAPH} следует внимательно изучить таблицу успеваемости. [Родители]_Y слишком склонны обвинять школы в ограниченном образовании [своих]_{Y-ANAPH} детей. Если [родители]_Z недовольны школой, у [них]_{Z-ANAPH} должна быть возможность перейти в другую.
- ANAPH X: Родители $\xleftarrow{\text{ANAPH}}$ своих, Они, их, они, им
- ANAPH Y: Родители $\xleftarrow{\text{ANAPH}}$ своих
- ANAPH Z: родители $\xleftarrow{\text{ANAPH}}$ них

В примере (64) ни одно из многократных упоминаний “детей” не образует кореферентных групп IDENT, поскольку, во-первых, это обобщающее упоминание, а во-вторых, нет местоимений, которые указывали бы на одно из этих упоминаний.

3.2.2 Копулярные конструкции

Копула — связка между подлежащим (субъектом) и сказуемым (предикатом). В русском языке, как правило, это тире или разные формы глагола *быть*. Кроме того, копулярными являются глаголы *казаться*, *чувствовать*, *выглядеть*, *оставаться*, *становиться*, *оказаться* и другие логически схожие глаголы. Кроме того, связка может быть нулевой. Например, в “*Пушкин писатель*”.

В таких конструкциях именные группы (субъект и предикат), связываемые копулой, не объединяются в кореферентную группу IDENT, как в примерах (65)-(70).

(65) [Петя] (-) [школьник]

(66) [Петя] был [школьником]

(67) [Джон] стал [начальником отдела]

(68) [Джон] чувствовал себя [идиотом]

(69) [Фред] считает, что [Джон] (-) [идиот]

(70) [Проход] оказался [тупиком]

3.2.3 Связь копулярных конструкций с другими упоминаниями

Как правило, с последующими упоминаниями связывается наиболее конкретное (см. шкала конкретности в 3.3.1) или наиболее логически подходящее упоминание (субъект или референт) – чаще всего это первый элемент, однако иногда это может быть и второй элемент копулярной конструкции:

(71) [Джон]_X [лингвист]. [Люди]_Y обычно бешутся, когда [он]_{X-ANAPH} [их]_{Y-ANAPH} поправляет.

(72) [Джон]_X [лингвист]_Y, [поэтому]_{Y-ANAPH} [люди]_Z обычно бешутся, когда [он]_{X-ANAPH} [их]_{Z-ANAPH} поправляет.

(73) Одна из крупнейших фармацевтических компаний России – [Фармасинтез]_X. [Она]_{X-ANAPH} заключила договор со Сбербанком.

(74) Директор там – [Иван Петров]_X. [Он]_{X-ANAPH} очень любит свою работу.

(75) [Директор]_X там – Иван Петров. [Эта должность]_{X-ANAPH} подразумевает много ответственности.

Второй элемент копулярной конструкции может быть в свою очередь связан с каким-либо упоминанием в тексте. Так, в примере (76) упоминание “*поэтому*” является анафором для упоминания “*школьник*”, и в примере (77) оба упоминания, связанные копулой, являются антецедентами:

(76) [Петя]_X [школьник]_Y - [поэтому]_{Y-ANAPH} [его]_{X-ANAPH} не пустили в бар.

(77) [Директор]_X там - [Иван Петров]_Y. На [этой должности]_{X-ANAPH} [он]_{Y-ANAPH} работает уже три года.

3.3 APPOS

Аппозитивная конструкция содержит именную группу, которая дополнительно уточняется одной или несколькими непосредственно соседними именными группами, разделенными тире, запятой, двоеточием или круглыми скобками. Цепочки APPOS состоят из основной именной группы HEAD, референта и одного или нескольких атрибутов ATTRIB, которые переименовывают или дополнительно определяют этот референт. В устной речи такие конструкции обычно выделяются паузами и отдельной интонацией.

(78) [Джон Смит]_{X-HEAD}, [СЕО нашей компании]_{X-ATTRIB}, приедет к нам на ужин.

APPOS X: Джон Смит $\xleftarrow{\text{APPOS}}$ СЕО нашей компании

Nota Bene! Тест на определение аппозитивов – любая часть аппозитивной конструкции (HEAD или ATTRIB) может использоваться с остальной частью предложения без искажения смысла:

- [Джон Смит] приедет к нам на ужин.
- [СЕО нашей компании] приедет к нам на ужин.

При разметке все упоминания следует присоединять к HEAD, а не ко всей аппозитивной цепочке. В финальной разметке все ATTRIB будут присоединены автоматически.

3.3.1 Выбор HEAD в аппозитивных конструкциях

В каждой аппозитивной конструкции наиболее конкретный элемент помещается как HEAD.

Шкала конкретности, на которую можно ориентироваться при разметке:

КОНКРЕТНЫЕ	—————>	НАИМЕНЕЕ КОНКРЕТНЫЕ
имя собственное > местоимение > референтная ИГ > нереферентная ИГ		
<i>Джон Смит > он > мой друг > лингвист</i>		

В приведенных ниже примерах подчеркнутый элемент является наиболее конкретным и поэтому отмечен как HEAD.

- (79) [Джон Смит]_{X-HEAD}, [известный лингвист]_{X-ATTRIB}, прибыл на конференцию.
- (80) [Президент нашего лингвистического клуба]_{X-ATTRIB}, [Дж. Смит]_{X-HEAD}, выступил с докладом.
- (81) Сбербанк и [одна из крупнейших фармацевтических компаний России]_{X-ATTRIB} - [«Фармасинтез»]_{X-HEAD} - заключили договор о сотрудничестве.

В случаях, когда два или более членов цепочки APPOS эквивалентны по конкретности, крайний левый член аппозитивной конструкции отмечается как HEAD:

- (82) [Друг детства]_{X-HEAD}, [генерал-майор]_{X-ATTRIB}, позвал меня на рыбалку.
- (83) [Генерал-майор]_{X-HEAD}, [друг детства]_{X-ATTRIB}, позвал меня на рыбалку.

Конкретные названия болезней, технологий и т.п. классифицируются как имена собственные, независимо от того, пишутся они с заглавной буквы или нет.

- (84) Ранее в июле британские ученые выразили обеспокоенность тем, что [супермикробы]_{X-HEAD} - [патогены, выработавшие устойчивость к широко используемым антибиотикам]_{X-ATTRIB} - представляют особую опасность для пациентов с онкологическими заболеваниями.

- (85) Спектр ССЗ достаточно широк - к самым распространенным из них относят [нарушения мозгового кровообращения]_{X-ATTRIB} ([инсульт]_{X-HEAD}), [ишемическую болезнь сердца]_{Y-ATTRIB} ([инфаркт]_{Y-HEAD}), [гипертонию]_{Z-HEAD} ([высокое кровяное давление]_{Z-ATTRIB}), фибрилляцию предсердий, врожденные пороки сердца и [кардиомиопатии]_{W-HEAD} ([заболевания миокарда, при которых происходят структурные и функциональные изменения в сердечной мышце]_{W-ATTRIB}).

В предложениях с ссылками на нормативно-правовые акты упоминание статьи отмечается как HEAD:

- (86) Вскоре задержанных отпустили из полиции с протоколами по [статье 20.6.1 КоАП]_{X-HEAD} ([невыполнение правил поведения при чрезвычайной ситуации или угрозе ее возникновения]_{X-ATTRIB}).
- (87) Плющева признали виновным в [нарушении правил пикета]_{X-ATTRIB} ([часть 5 статьи 20.2 КоАП]_{X-HEAD}).

Аппозитивные конструкции, состоящие из более двух подряд идущих именных групп содержат один HEAD, все остальные именные группы размечаются как ATTRIB. Более того, внутри ATTRIB может быть вложена еще одна APPOS цепочка, как APPOS Y в примере (88).

- (88) [Илон Маск]_{X-HEAD}, [49 лет]_{X-ATTRIB}, [генеральный директор и главный инженер компании SpaceX]_{X-ATTRIB}, [генеральный директор и [главный идейный вдохновитель]_{Y-HEAD} ([Chief Product Architect]_{Y-ATTRIB}) компании Tesla]_{X-ATTRIB}, стал вторым богатейшим человеком в мире.

APPOS X: Илон Маск $\xleftarrow{\text{APPOS}}$ 49 лет; генеральный директор и главный инженер компании SpaceX; генеральный директор и главный идейный вдохновитель (Chief Product Architect) компании Tesla
 APPOS Y: главный идейный вдохновитель $\xleftarrow{\text{APPOS}}$ Chief Product Architect

3.3.2 Выбор между APPOS и САТАРН

В текстах с аппозитивными конструкциями, в которых одним из членов является местоимение, нужно выбрать либо APPOS, либо САТАРН связь. При выборе связи следует опираться на шкалу конкретности из раздела 3.3.1.

Если аппозитив – собирательное или общее понятие без конкретного референта, как “*выпускник Гарварда*” и “*дети*” в примерах (89) и (90), местоимение будет HEAD, а аппозитив – ATTRIB. Если, наоборот, аппозитив является конкретным наименованием или именем, как “*Иван Петров*” в примере (91), то местоимение будет катафорой CATAPH для имени-постцедента.

(89) [Он]_{X-HEAD}, [выпускник Гарварда]_{X-ATTRIB}, стал президентом.

APPOS X: Он $\xleftarrow{\text{APPOS}}$ выпускник Гарварда

(90) Мать ехала с отцом со станции Сиверской, а [мы]_{X-HEAD},

[дети]_{X-ATTRIB}, выехали им навстречу.

APPOS X: мы $\xleftarrow{\text{APPOS}}$ дети

(91) [Я]_{X-CATAPH}, [Иван Петров]_X, находясь в трезвом уме и твёрдой памяти, хочу сделать чистосердечное признание.

CATAPH X: Я $\xrightarrow{\text{CATAPH}}$ Иван Петров

3.3.3 Связывание аппозитивных цепочек с другими референтами

Цепочки APPOS могут в свою очередь входить в цепочки IDENT. При этом в IDENT входит вся цепочка APPOS, HEAD и ATTRIB неразрывно, как IDENT X в следующем примере.

Nota Bene! При разметке все упоминания, кореферентные с аппозитивной конструкцией, связываются с HEAD. В финальной разметке все ATTRIB будут присоединены автоматически.

(92) [[Илон Маск]_{X-HEAD}, [49 лет]_{X-ATTRIB}, [генеральный директор и главный инженер компании SpaceX]_{X-ATTRIB}, [генеральный директор и [главный идейный вдохновитель]_{Y-HEAD} ([Chief Product Architect]_{Y-ATTRIB}) компании Tesla]_{X-ATTRIB}]_X, стал вторым богатейшим человеком в мире. В июне [Маск]_X заявил, что [его]_{X-ANAPH} не волнует размер [его]_{X-ANAPH} капитала.

IDENT X: [Илон Маск, 49 лет, генеральный директор и главный инженер компании SpaceX, генеральный директор и главный идейный вдохновитель (Chief Product Architect) компании Tesla], Маск

APPOS X: Илон Маск $\xleftarrow{\text{APPOS}}$ 49 лет; генеральный директор и главный инженер компании SpaceX; генеральный директор и главный идейный вдохновитель (Chief Product Architect) компании Tesla

APPOS Y: главный идейный вдохновитель $\xleftarrow{\text{APPOS}}$ Chief Product

Architect

ANAPH X: Маск $\xleftarrow{\text{ANAPH}}$ его, его

3.3.4 Денежные эквиваленты

Смежные именные группы, обозначающие эквиваленты денежных сумм в разной валюте, размечаются как аппозитивная конструкция.

- (93) Они сэкономили еще дополнительно [9 млн. долларов]_{X-HEAD} ([680 млн. рублей]_{X-ATTRIB}).

3.3.5 Неаппозитивные конструкции

От аппозитивных конструкций следует отличать причастные и деепричастные обороты.

- (94) На вокзале мы встречаем родственников, приехавших к нам.
- (95) Мебель, привезенную с дачи, мы поставили в коридоре.
- (96) Человек шёл, оглядываясь по сторонам.
- (97) Сборная США, традиционно не блещущая результатами на чемпионатах мира, одолела сходную по стилю игры сборную Италии.

4 Отдельные случаи

4.1 Объединение именных групп в одно упоминание

Именные группы, перечисляемые через запятую или с помощью союза “и”, могут быть объединены в одно упоминание, как в примерах (98)-(101).

- (98) В этом году у него [две громкие премьеры]_{X-APPOS} - [‘Веном’ и ‘Фонзо’]_{X-HEAD}.
- (99) [Война, солдаты, штурм и всё прочее]_X, - [всё это]_{X-ANAPH} мне приснилось
- (100) [Вася, Маша, Петя и Саша]_X - [все]_{X-ANAPH} были там.
- (101) [Все мы]_{X-APPOS} - [родители и учителя]_{X-HEAD} - заинтересованы в успеваемости детей.

4.2 Аббревиатуры

Аббревиатуры, называющие конкретный референт (компанию, продукт, технологию и т.п.) могут быть связаны с другими упоминаниями.

- (102) [ЮНЕСКО]_х считает образование своим приоритетным направлением деятельности с момента основания [организации]_{х-ANAPH}. После войны [организация]_{х-ANAPH} занималась восстановлением разрушенных школ и налаживанием прерванных контактов в сфере образования в Западной Европе.
- (103) [Российская академия наук]_х – колоссальная научно-исследовательская база. И именно [РАН]_х во многом определяет стратегию и перспективы развития отдельных областей знания и направления исследований.

4.3 Даты, временные отсылки

Даты, временные промежутки или другие временные упоминания (включая дейктические выражения⁴, такие как *сейчас*, *потом*, *сегодня*, *завтра*, *вчера* и т. д.) могут входить в кореферентные группы.

- (104) Джон провел [три года]_х в тюрьме. За [это время]_{х-ANAPH} мир кардинально изменился.
- (105) А произошло это [10 лет назад]_х в [1994 году]_х, [когда]_{х-ANAPH} журналист погиб от срабатывания взрывного устройства прямо на рабочем месте.
- (106) Ну, мы приобрели собственность в [2006 году]_х. И в [это время]_{х-ANAPH} на опытном заводе уже происходили конверсионные программы.
- (107) Слушай, пошли [завтра]_х гулять? – Нет, [завтра]_х точно не получится.

Временные выражения с несколькими датами (например, месяц, день, год) считаются атомарными (целостными, неразрывными), а вложенные даты не связаны с другими упоминаниями этой даты. В примере (108) “*ноябрь*” или “*2020 г.*” не могут входить в кореферентную группу с “*3 ноября 2020 г.*”.

⁴От англ. *deixis*, *deictic*, их референт изменяется в зависимости от говорящего и контекста.

- (108) Выборы президента США всегда проводятся в [ноябре]. Выборы [2020 г.] состоялись [3 ноября 2020 г.]_х. Однако в [тот день]_{х-ANAPH} победитель не был объявлен.
 ANAPH X: 3 ноября 2020 г. $\xleftarrow{\text{ANAPH}}$ тот день

4.4 Денежные суммы

Денежные суммы могут быть кореферентны с другими упоминаниями в тексте, если относятся к тому же референту.

- (109) Они дешевле в производстве ([рекомендованная цена]_х - [60 евро]_х, [дилерская стоимость]_у - [до \$80]_у), несколько более устойчивы при легких ударах и абсолютно не подвержены коррозии.
- (110) [Объем торгов]_х в российских АДР это [порядка 200-250 млн долларов в день]_х.
- (111) [Инвестиции]_х за эти годы в модернизацию и создание новых производств достигли [2 млрд. долларов]_х, [эти средства]_х, [финансовые ресурсы]_х были взяты из финансового рынка.

4.5 Именованные сущности

Именованные сущности рассматриваются как атомарные упоминания. Если именованная сущность включает в себя вложенные наименования, как в примерах (112) и (113), рассматривается только внешняя именованная сущность.

- (112) Чего Вы ждёте от начала работы [Дома Москвы]_х? Как в целом относитесь к [его]_{х-ANAPH} открытию?
- (113) А в 1981-м «Оскар» получил фильм [«Москва слезам не верит»]_х. [Кинолента]_{х-ANAPH} была признана культовой.

4.6 Множества и подмножества

Кореферентными и анафорическими связями могут объединяться только равнозначные по количеству/объему/охвату множества и подмножества. В примере (114) равнозначными являются только упоминания “5221 из них” и “них”, при этом второе упоминание является местоимением-анафором. Упоминания 408802 людей, одних, других и третьих обозначают разные по количеству подмножества, поэтому не кореферентны.

- (114) В ходе исследования ученые наблюдали за 408 802 людьми в возрасте от 37 до 73 лет, при этом у [5221 из них]_X была обнаружена сердечная недостаточность. Оказалось, что у [них]_{X-ANAPH} были проблемы со сном: одни мучались от бессонницы, другие часто просыпались по ночам из-за собственного храпа, у третьих продолжительность сна была слишком короткой.

Также, между упоминанием организации (или любой другой группировки людей) и подмножества ее членов нет кореферентных связей:

- (115) 26 октября забастовка началась на [«Гродно-Азот»], [одна из смен] остановила работу цеха. В 7 утра у проходной собралось [порядка 200 рабочих]. Сотрудники милиции задержали [32 человека].

4.7 GPE и правительства

GPE⁵ связаны с упоминаниями их правительства, включая метонимические упоминания столицы (например, когда “Москва” или “Кремль” употребляется в значении “Россия” или “Правительство России”).

- (116) Необходимо дать понять [Москве]_X, что США и Запад не приемлют вмешательства [России]_X во внутренние дела бывших республик СССР.

IDENT X: Москве, России

- (117) Христианские законодатели настаивают на выводе сирийских войск из [Ливана]_X, прежде чем согласиться на политические изменения, дающие мусульманам [страны]_{X-ANAPH} большую роль в правительстве [Бейрута]_X.

ANAPH X: Ливана $\xleftarrow{\text{ANAPH}}$ страны

IDENT X: Ливана, Бейрута

В примере (116) упоминания *Москва* и *Россия* имеют один и тот же референт – правительство РФ. В примере (117) под упоминаниями *Ливан* и *Бейрут* имеется в виду Ливан, как страна. Кроме того, упоминание *страна* является анафором для антецедента *Ливан*.

Nota Bene #1! Упоминания страны и жителей страны не являются кореферентными, например:

⁵GPE (англ. Geopolitical entity) – геополитические сущности

- (118) За 13 лет правления Милошевича [народ [Югославии]_Y]_X увидел, как
 [[его]_{X-ANAPH} страна]_Y разорвана на части.
 ANAPH X: народ Югославии $\xleftarrow{\text{ANAPH}}$ его
 IDENT X: Югославии, его страна

Нужно также отметить, что в примере (118) “*его*” – это анафора к “*народу Югославии*”. Отсюда получается, что “*его страна*” \Rightarrow “*страна народа Югославии*”. Таким образом, это упоминание является не анафором, а референтной именной группой и входит в цепочку IDENT.

Nota Bene #2! Упоминания названия организации/правительства и названия здания, в котором оно находится, не кореферентны. Например, *Кремль* в значении “*правительство России*” и *Кремль* в значении “*историческая крепость в центре Москвы*” не будут кореферентны. Аналогично, в примере (119) выделяется две отдельные цепочки для упоминания здания ООН (цепочка X) и ООН как организации (цепочка Y).

- (119) Об этом сообщает корреспондент Брек Ардери из [штаб-квартиры ООН]_X. Сегодня с коротким выступлением выступил главный палестинский наблюдатель при [ООН]_Y Насир Аль-Кидва, за ним последовал посол Израиля в [ООН]_Y Иегуда Ланкри. Брек Ардери, VOA News, репортаж из [Организации Объединенных Наций]_X.

4.8 Количественные выражения, квантификаторы

В текстах могут встречаться предложения с квантификаторами. Квантификаторы – это выражения, которые определяют количество сущностей или упоминаний. Обычно фраза с квантификатором выглядит как “X (из) Y”, где X – квантификатор, а Y – именная группа в родительном падеже (*кого? чего?*). Квантификаторы включают в себя количественные числительные (например, *три, сто пятьдесят девять, миллион*), партитивы⁶ (например, *несколько, часть, половина*), измерения (например, *килограмм, дюжина, пригоршня, десяток*) и коллективные существительные (например, *отряд, стадо, группа*).

- [Трое из [них]]_X
- [Много [чепухи]]_X
- [Отрезок [дороги]]_X

⁶Партитивы – это квантификаторы, которые обозначают подмножество (или часть) большего набора сущностей.

- [Стая [птиц]]_X

В таких случаях внутренняя именная группа (*их, чепуха, дорога, птицы*) не кореферентна с внешней. Однако внешняя группа целиком может связываться кореферентными и анафорическими связями с другими упоминаниями в тексте.

Исключением являются выражения с *все* и *оба*. В таких случаях внешняя и внутренняя группа обозначает одинаковое количество объектов, поэтому внешняя и внутренняя группы соединяются связью IDENT:

- (120) [Все [участники конференции]_X]_X выступили с докладами.

IDENT X: Все участники конференции, участники конференции

- (121) Это подарок от [нас [обоих]_X]_X.

IDENT X: нас обоих, нас

4.9 Связь количественных выражений с другими упоминаниями

Количественное выражение (квантификатор + именная группа в родительном падеже) может быть кореферентно с другими упоминаниями в тексте:

- (122) Я заказал [кружку [кофе]]_X, но [ее]_{X-ANAPH} так и не принесли.

ANAPH X: кружку кофе $\xleftarrow{\text{ANAPH}}$ ее

Однако, если анафор или другое упоминание согласуется по роду/числу с именной группой внутри количественного выражения, а не с квантификатором, связь ANAPH будет направлена на эту именную группу:

- (123) Я заказал [кружку [кофе]]_X, но [его]_{X-ANAPH} так и не принесли.

ANAPH X: кофе $\xleftarrow{\text{ANAPH}}$ его

Внутренняя именная группа в количественном выражении может участвовать в кореферентной или анафорической связи, если обозначает равнозначное с другим упоминанием количество объектов, как ANAPH Y в примере (124).

- (124) [Пятеро врачей]_X представили свои исследования, а затем [трое из [них]_{X-ANAPH}]_Y изложили [свое]_{Y-ANAPH} мнение по проблеме.

ANAPH X: пятеро врачей $\xleftarrow{\text{ANAPH}}$ них

ANAPH Y: трое их них $\xleftarrow{\text{ANAPH}}$ свое

Однако внутренняя именная группа не участвует в кореферентных или анафорических связях с общими или собирательными понятиями:

- (125) [Трое докторов] высказали свое мнение, однако всем известно, что доктора часто расходятся во мнениях.

4.10 Парцелляция и короткие предложения

Парцелляция – конструкция экспрессивного синтаксиса, представляющая собой намеренное расчленение связного текста на несколько пунктуационно и интонационно самостоятельных отрезков (*Википедия*).

В некоторых текстах одна логическая именная группа быть разбита по двум предложениям. В таком случае, в кореферентных связях будет участвовать только существительное без зависимых слов:

- (126) Я в сельмаге купил [книжку]_х. Про трактор. [Она]_{х-ANAPH} показалась довольно интересной.

Если в тексте встречаются короткие предложения, которые напоминают аппозитивы, однако являются отдельными предложениями, – они включаются в IDENT или ANAPH/CATAPH связи:

- (127) [Цена]_х? [\$300]_х. [Очень много]_х по нынешним меркам.

5 Резюме

В *ru:corner* в качестве упоминаний выделяются именные группы, местоимения, глагольные группы и предикативы, которые связаны кореферентной связью с другими упоминаниями в тексте. Упоминаниями могут быть слова и фразы разных семантических категорий: люди, организации, объекты, события, явления, даты, денежные суммы, концепты и т.д.

Упоминания могут быть вложенными, при этом в кореферентной связи участвует то упоминание, которое логически связано с другими упоминаниями. Внешнее и вложенные упоминания могут входить в разные кореферентные группы.

В *ru:corner* есть три вида кореферентных связей:



1. **ANAPH/COREF** – направленная связь от упоминаний-анафоров (местоимений и нереферентных именных групп) к референтным упоминаниям в тексте. В отрыве от контекста анафоры не имеют конкретный референт. Связь **ANAPH** направлена влево, **CATAPH** – вправо. Анафоры связываются с ближайшим предшествующим, либо (если в предшествующем контексте нет упоминания, на которое указывает анафор) с ближайшим последующим упоминанием.
2. **IDENT** – ненаправленная связь между идентичными референтными упоминаниями. Такие упоминания взаимозаменяемы без потери смысла.
3. **APPOS** – направленная связь между упоминаниями и дополнительно описывающими или переименовывающими их одним или более аппозитивами. Аппозитивы, как правило, обособлены запятыми, тире или скобками. Связь направлена от аппозитивов к основному упоминанию.

В рамках одного текста могут встречаться и пересекаться несколько видов кореферентных связей.



Нереферентные упоминания (например, общие или собирательные понятия) выделяются как референтная группа, только если в тексте есть анафор, который на них указывает. Несколько повторяющихся нереферентных упоминаний не образуют **IDENT** групп, каждое из таких упоминаний образует свою цепочку **ANAPH/CATAPH** с указывающими на него анафорами.

В программе разметки при связывании упоминаний с цепочкой **APPOS**, связи следует направлять к основному упоминанию, не к аппозитиву.


А Направленные связи ANAPH/CATAPH

- Она выдвинула хорошее предложение, его все поддержали .

- За несколько недель до его смерти , отец завещал мне свой дом .


В Ненаправленные связи IDENT

- А произошло это 10 лет назад в 1994 году .

- Дэвид Бекхэм собирается запустить свое кулинарное шоу . Кулинарией Бекхэм увлекся во время карантина .


С Направленные связи APPOS

- Джон Смит , СЕО нашей компании , придет к нам на ужин .

- Президент нашего лингвистического клуба , Дж. Смит , выступил с докладом .
