

Инструкция по разметке ошибок в корпусе REALEC

I. Общие рекомендации

1. В начале работы введите логин и пароль во всплывающем окне справа вверху.
2. Если вы заметили ошибку, кликните дважды на слово или несколько слов, которые вы считаете областью этой ошибки (вопросы о подходах к выбору области ошибки обсуждаются в приложении к конкретным случаям ниже). На экране появится окно аннотации¹ с областью ошибки, с полем классификации ярлыков² и с полем **Notes** для ввода исправленного варианта.

Индикатор удаления (если область ошибки нужно полностью убрать из текста)

Добавление фрагментации (не использовать!)

Удаление аннотации

Изменение области ошибки

Сохранение аннотации

Отмена изменений

3. Если вы хотите просто привлечь внимание к тому, что в данной области имеется ошибка, без выбора тэга для данной ошибки, вы можете выбрать тэг «**Исправление**», поставив точку в кружке рядом с этим ярлыком, и при этом ничего не вставляя в поле **Notes** (однако позже этот тег рекомендуется заменить на тэг конкретной ошибки). При необходимости оставить свою подпись следует применить тэг «**Комментарий**» и указать свою фамилию в поле **Notes**.

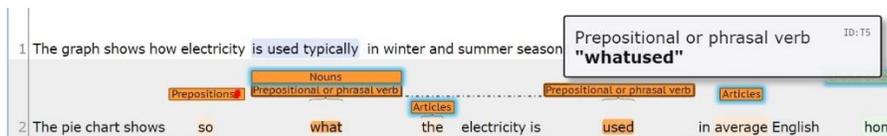
¹ В данной инструкции термины «аннотация» и «разметка», равно как и глаголы «размечать» и «аннотировать» являются полными синонимами.

² В тексте данной инструкции термины «ярлык» и «тэг» являются полными синонимами.

4. Если же вы хотите указать, какого рода ошибку вы видите в выделенной области, то кликните один раз на плюс рядом с ярлыком «исправление», плюс превратится в минус, а под ярлыком «исправление» вы увидите первый, самый общий, уровень классификации тэгов – “**Punctuation**”, “**Spelling**”, “**Capitalisation**”, “**Grammar**”, “**Vocabulary**” и “**Discourse**”. Знак плюс рядом с ярлыком означает, что у ярлыка имеется дальнейшее деление на уровни – вы увидите плюсы перед последними тремя ярлыками. Таким образом, для ошибок в пунктуации и правописании вы просто кликаете на ярлык “**Punctuation**” или “**Capitalisation**” или “**Spelling**” (см. Раздел II). Для ошибок в остальных трех областях нажимаете на плюс и смотрите на подразделения. Рекомендации для разметки ошибок в этих областях см. в разделах III, IV и V, соответственно.
5. В поле **Notes** нужно привести правильный вариант вместо ошибочного. Отдельные слова или все выражение из области ошибки в окне аннотации можно скопировать в поле **Notes**, если выделить его мышкой и перетащить его вниз в поле **Notes** до тех пор, пока рядом с курсором не появится прямоугольник (пока рядом с курсором кружок, возможность копирования отсутствует).
6. Ярлык для ошибок в большинстве случаев выбирается не по типу ошибки, а по той форме, которая должна быть на месте ошибочной – например, в случае ошибки в модели управления глагола надо выбирать ярлык той модели, которая для данного глагола в том его значении, в котором он употреблен в данном контексте, записана для него в словаре. Ниже случаи, когда ярлык указывает на тип ошибки, а не на требуемую форму, будут оговорены.
7. Для каждой ошибки нужно стараться выбрать как можно более конкретный ярлык, т.е. вместо верхних уровней подходящий подуровень, если таковой имеется. Исключением является тот случай, когда у ярлыка верхнего уровня есть деление на некоторые подуровни, но это деление не охватывает все случаи ошибок, относящихся к ярлыку верхнего уровня. Для случая ошибки, не охваченного подуровнями, выбирают ярлык верхнего уровня.
8. Если вам попала ошибка в слове, входящем в состав словосочетания, необходимо проанализировать, является ли областью этой ошибки само ошибочное слово или все словосочетание. Так, например, ошибка в предлоге после глагола может оказаться **либо** словарной ошибкой выбора части предложного глагола (*put off with his egoism* > *put up with his egoism* “**Choice of a part of lexical item**” – область ошибки здесь фразовый глагол *put off*), **либо** ошибкой в модели управления глагола (*see on us* > *see us* “**Transitive verb**” – область ошибки *see on*), **либо** ошибкой в предлоге (*seated on the armchair* > *seated in the armchair* “**Preposition**” – область ошибки *on*).

9. Если исправление ошибки заключается в том, что слово или слова, отмеченные как область ошибки, нужно убрать из текста, после нахождения нужного ярлыка ошибки при разметке следует поставить галочку в клетке **Delete** поля **Entity Attributes**, расположенного непосредственно под полем с классификацией тэгов. В документе это будет отражаться красной галочкой на ярлыке с именем ошибки.
10. Если в процессе нахождения нужного ярлыка вы понимаете, что первоначально выбранная область ошибки должна быть короче или длиннее, или просто на других словах, не обязательно удалять готовую аннотацию кнопкой **Delete** и открывать новую - можно после выбора нужного ярлыка (или после открытия этой аннотации) нажать кнопку **Move** вместо кнопки **OK**, а когда окошко аннотации закроется, выделить ту область ошибки, которая вам нужна.
11. Если изменение в области ошибки влечёт за собой изменение следующего слова, следует включить это слово в область ошибки и добавить его правильный вариант к исправлению в поле **Notes**. Указывать отдельную ошибку на этом следующем слове не следует. Если же слово, которое следует изменить в результате исправления ошибки, находится на некотором отдалении от исправленной ошибки, то его можно выделить ярлыком **Исправление** и протянуть от ярлыка исправленной ошибки к этому слову стрелку **Dependent change** (о стрелках см. Раздел VI).
12. На одной и той же области ошибки могут стоять несколько тэгов - например, **“Spelling”** и **«Words often confused»** для случая, когда написано *loose* вместо *lose*, или **“Noun number”** и **“Articles”** для замены *the employee* на *employees*. Очень важно, чтобы исправления в поле **Notes** всех аннотаций данной области были совершенно одинаковые.
13. Когда выбор нужного ярлыка закончен, нажмите кнопку **OK**. Нажатие кнопки **OK** вернет вас в текст документа, а над областью ошибки, с которой вы работали, возникнет ярлык определенного цвета, и подведение курсора к этому ярлыку покажет правильный вариант, если он был введен. Если вы хотите что-то исправить в вашей аннотации, произведите двойной клик по этому ярлыку, и вы получите окно, с которым вы уже работали. Если вы хотите удалить аннотацию, после двойного клика в полученном окне, с которым вы уже работали, нажмите кнопку **Delete**. Вверху появится окно с запросом на подтверждение вашего желания удалить аннотацию, что производится нажатием кнопки **OK**.
14. Внизу окна аннотации также имеется кнопка **Add Frag**. Изначально она предназначена для разделения аннотации на две части (к примеру, в случае если область

ошибки разделена клаузой или вспомогательным словом), однако текущая платформа корпуса пока что не может корректно взаимодействовать с фрагментацией, и её применение может вызвать ошибки, поэтому пользоваться этой функцией не рекомендуется. В случае, если фрагментация появилась в вашем эссе, хотя вы не добавляли её, попробуйте перенести область ошибки самостоятельно или сообщите об этом администратору. Распознать фрагментацию можно по пунктирной линии между тегами:



15. В некоторых папках по умолчанию отображаются ярлыки частеречной разметки. Если эти ярлыки мешают при аннотации, их можно скрыть, открыв меню **Options** (кнопка для открытия находится рядом с кнопкой Logout) и выбрать в поле **POS-tagging** переключатель **Off**. При необходимости возможно провести обратную операцию, выбрав переключатель **On**.
16. В конце работы не забудьте нажать **Logout** во всплывающем окне справа вверху.

II. Разметка ошибок классов “Punctuation”, “Spelling”, "Capitalisation"

В этих классах нет подклассов. Областью ошибки следует выбирать тот кусок текста, на котором должно быть соблюдено некоторое правило пунктуации или правописания, а именно:

- все вводное выражение, ошибочно не отделенное запятой от остальной части предложения - в поле **Notes** это выражение копируется и после него ставится запятая;
- союз, перед которым либо должна быть запятая, а она отсутствует, либо не принято ставить запятую (например, в сравнительной конструкции перед союзом *than*), а она стоит - в поле **Notes** этот союз копируется, соответственно, либо после запятой, либо без предшествующей запятой;
- вся конструкция (например, приложение или парентетическое выражение в середине предложения, ошибочно не выделенные запятыми в тексте эссе, или конструкция с эмфатическим выделением в начале предложения, ошибочно отделенная запятыми - **но не конструкции класса Relative clauses**, о которых см. следующий абзац) - в поле

Notes эта конструкция копируется с запятыми перед и после нее или, соответственно, без запятыx в начале и в конце конструкции).

Среди ошибок типа отсутствия нужной запятой (запятыx) и наличия лишней запятой (запятыx) имеются часто встречающиеся в студенческих работах случаи неправильного выбора типа *relative clause* - *defining*, *non-defining* или *coordinate*. Их следует размечать соответствующим ярлыком того типа, который нужен в данном контексте - “**Defining relative clause**” или “**Non-defining relative clause**” или “**Coordinate relative clause**” (синие синтаксические ярлыки в подразделе “**Relative clauses**” раздела “**Attributes**”), а не ярлыком “**Punctuation**”. При этом если стоит лишняя запятая, или необходимая запятая отсутствует только **перед** *relative clause*, то областью ошибки считается союзное слово, соответственно, с запятой или без нее, а если такая же ситуация с запятой повторяется в конце конструкции, то областью ошибки считается весь текст *relative clause*, копируемый в поле **Notes**, соответственно, с устрaнением или добавлением запятыx.

III. Разметка ошибок типа Grammar

В классификации грамматических ошибок наглядно видно, что они делятся на морфологические, оранжевого цвета, и синтаксические, синего цвета, хотя ярлыков с именами Morphology и Syntax в классификации нет.

III-a. Разметка морфологических ошибок

Деление морфологических ошибок произведено по части речи, в которой допущена ошибка – “**Determiners**”, “**Verbs**”, “**Nouns**”, “**Prepositions**”, “**Conjunctions**”, “**Adjectives**”, “**Adverbs**”, “**Numerals**”, “**Pronouns**”. Во всех ярлыках этого уровня, кроме предлогов и союзов, есть дальнейшее подразделение, причем в некоторых оно очень детальное – например, в глаголах. Первый ярлык подгруппы “**Verbs**” – тэг “**Tense**” – содержит подразделение на ошибки в выборе данной категории или в построении формы данной категории: “**Choice of tense**” & “**Tense form.**” Так, неправильный выбор времени при правильно построенной форме времени следует размечать тэгом “**Choice of tense**”, а такая форма, как, например, *have took* должна быть размечена ярлыком “**Tense form**”.

Самая подробная группа ярлыков представлена для ошибок в глагольных моделях управления, “**Verb patterns**”. Требуемая модель управления указывается в месте ошибки не только для финитных (видовременных) форм глаголов, но и для всех глагольных форм – инфинитива, инфинитива без частицы *to*, причастия настоящего времени, причастия прошедшего времени, герундия. К типам моделей управления в схеме тэгов не относятся ошибки в сочетаниях глагола-связки с прилагательными (*be proud of*, *be confident of*, etc.) -

так, например, неправильный предлог в таких сочетаниях должен быть размечен на области прилагательное-предлог ярлыком “**Prepositional adjective.**”

Для объяснения использования ярлыков в разделе “**Verb patterns**” разберем три примера из корпуса, а далее примеры будут подобраны по тэгам данного раздела.

(1) *But from my point of view, when we do this kind of collecting data, we just work on **understanding of** > **understanding** another person’s opinion.*

В предложении (1) имеется лишний предлог после глагола, стоящего в форме герундия, и этот глагол имеет модель управления прямым дополнением, т. е. является переходным глаголом, что в нашей классификации передает тэг “**Transitive Verb.**”

(2) *It is clear that approximately 70% of young people between 20 and 29 have ideal weight in both charts; however, situation for those who are in 60-69 age group is completely different, with more than a half in 2010 **suffering with** > **suffering from** obesity.*

В примере (2) автор допустил ошибку в предлоге после глагола *suffer*, стоящего в форме причастия настоящего времени. Модель управления этого глагола – *suffer from something*. Разметка данной ошибки заключается в выборе тэга “**Prepositional Verb.**”

Необходимо уметь различать случаи, когда предлог входит в модель управления глагола, а когда он образует отдельное обстоятельство.

(3) *They shake with fear, their hands begin to sweat, and hearts try to jump **out from** > **out of** their chests.*

В примере (3) ошибка именно в неправильном выборе предлога (тэг “**Prepositions**”), а не в модели управления глагола *to jump*. Иными словами, предлог *out of* (правильный вариант) вводит обстоятельственный комплекс, не являющийся обязательным аргументом глагола *to jump*.

II-b. Краткое описание разных моделей управления глаголов с характерными примерами и разбором нетривиальных случаев.

“**Intransitive Verb**” – к этому классу относятся ошибки в моделях управления всех непереходных глаголов типа *sleep, go, shine*. Важно не путать их с “**Prepositional Verb**”: у непереходных глаголов выражение какого-либо обстоятельства при глаголе факультативно (например, можно просто сказать «*I’m sleeping*», «*The sun is shining*» без указания времени или места, а можно сказать “*I slept for two hours*” и “*The sun is shining in the sky*”), в то время

как у “**Prepositional Verb**” дополнение, вводимое предлогом, является обязательным и входит в модель управления глагола – *I am looking for my keys*.

“**Transitive Verb**” – к этому классу относятся все переходные глаголы типа *see, show, write*. Важно отметить, что некоторые из них (*eat, understand, write*) могут употребляться как с прямым дополнением, так и без него, в зависимости от контекста. Тэг следует ставить тот, который требует контекст:

(4) *I'm writing about > writing now and cannot draw a picture in my essay*. – “**Intransitive Verb**”

Если бы из контекста не было ясно, что автор эссе говорит о том, чем он занят в настоящий момент, то аннотатору следовало бы предположить, что автор просто не дописал дополнение, которое вводит предлог, для разметки такой ошибки понадобился бы дискурсный тэг “**Absence of a component in clause or sentence**”.

(5) *Usually I write by > write poems*. – “**Transitive Verb**”

Ошибка скорее всего сделана под влиянием русского словосочетания «пишу стихами» - аннотатору это следует отметить выбором строки “**L1 Interference**” в разделе **Causes**.

“**Reflexive Verb**” – к этому классу относятся ошибки в переходных глаголах, управляющих возвратными местоимениями:

(6) *Please introduce > introduce yourself*.

(7) *If they want to wash > wash themselves in privacy, they have to wait for hours*.

NB: Важно помнить, что глаголы *afford* и *feel* в английском не имеют возвратного местоимения в своей модели (в отличие от их русских эквивалентов – *позволять себе* и *чувствовать себя*):

(8) *When I do not feel myself > feel well, I cannot travel across the city*.

“**Verbs with as**” – к этому классу относятся ошибки в глаголах типа *regard as, consider as*.

(9) *I have always regarded him to be > regarded him as my friend*.

(10) *I am considering him like > considering him as a candidate for Head of Department position*.

Отметим, что глагол *consider* имеет еще и другую модель, в которой его значение несколько другое:

(11) *I consider him to be very clever*.

В этой модели глагол-связка может опускаться:

(12) *I consider him very clever*.

Поскольку предложения 11 и 12 считаются полными синонимами, то в качестве тэга для ошибки в модели того же глагола в предложении 13 следует выбирать “**Verb+Infinitive**” :

(13) *I have always **considered the role of parents as** > **considered the role of parents to be** the most important factor in prevention of delinquency.*

В предложении(14) ошибка в модели управления глагола *consider* осложнена вовлечением отрицания, но для ее разметки следует использовать тот же тэг “**Verb+Infinitive**”:

(14) *The experts, however, **consider it's not** > **do not consider it to be** dangerous.*

“**Ambitransitive Verb**” – этим тэгом размечаются ошибки в глаголах, которые могут употребляться как в качестве переходных, так и в качестве непереходных: *open, begin, seat, drop, burn, finish, stop* и некоторые другие. Некоторые из таких глаголов могут также употребляться и в пассиве. Вот варианты из словарей для трех глаголов этого типа:

She dropped her bag. / The bag dropped on the floor. /The bag is dropped, and its handle gets torn.

Have you finished? / Have you finished your class? / Are you finished?

The waiter will seat you. / The theatre seats a thousand people. /We were seated in the corner.

Однако, на употребление пассивных форм ряда других глаголов имеются ограничения, с которыми изучающие язык могут не быть знакомы - как, например, в следующем случае:

(15) *She **was dropped** > **dropped** out of university after the first set of exams.*

“**Prepositional or phrasal verb**” – для ошибок в глаголах, которые управляют некоторым дополнением с помощью предлога – *look at, depend on, etc.* К этому классу относятся также глаголы с двумя предлогами для управления одним дополнением и с двумя дополнениями, прямым и предложным, в модели управления (*replace smth with smth, transform smth into smth, complain to smb about smth, look forward to smth, go on to smth, etc.*)

Среди глаголов этого класса есть такие, которые имеют несколько моделей управления и могут вести себя по-разному - ср. глагол *distinguish*:

(16) *This leads to some difficulties in the distinction between the two concepts and this is where the main problem lies: how to **distinguish among** > **distinguish between** homonymy and polysemy?*

Глагол *distinguish* имеет 3 модели управления:

- to distinguish smth/smb from smth/smb
- to distinguish between smth/smb and smth/smb

- to distinguish smth/smb among smth/smb

Первые две модели семантически ведут себя довольно похоже, в них речь идет о различении двух объектов, в то время как в третьей акцент делается на выделении чего-то одного из целой области. При исправлении ошибки в предложении (16) можно выбрать любую из первых двух моделей: если мы выбираем первую, то исправляем данную область ошибки на «*how to distinguish homonymy from polysemy*»; а если выбираем вторую модель, то, соответственно, исправляем на «*how to distinguish between homonymy and polysemy*.» В данном контексте третья модель управления невозможна, поскольку речь идет о двух классах, а не о большем числе классов.

Еще один нетривиальный случай вариативности в модели управления глагола:

(17) *Don't believe the information you find on the packaging of a herbal product, you should always **consult with** > **consult** a specialist.*

Среди примеров с этим глаголом находятся четыре разных модели управления:

- *I consult my doctor when I am not well.*
- *My doctor consults me when I am not well.*
- *I consult with my doctor about the best treatment.*
- *My doctor consults me about the best treatment.*

На сайте <https://forum.wordreference.com/threads/consult-and-consult-with.2262139/> представлено следующее объяснение различий между первым и третьим употреблениями:

If you *consult with* someone then you each consult the other, on an equal footing. For example:

- [$\ast=1$]"What time do you and your wife want to visit us this evening?"[$\ast=1$]"I don't know---I need to **consult with** her first"[$\ast=1$]

The two of you then consult with each other to agree what answer to give. If you *consult* someone, they have some information or an opinion, and you go to them in order to get it from them. It only happens in one direction. For example:

- [$\ast=1$]"If you feel ill, you should **consult** a doctor."
[$\ast=1$]"To check the meaning of a word, I **consult** my dictionary."

В предложении (17) кажется более вероятным именно второй случай – важно получить какую-то информацию от специалиста, значит нам нужна модель управления прямым дополнением – “**Transitive Verb.**”

“Dative verb with alternation” – глаголы, при которых адресат может выступать в двух позициях: после прямого дополнения присоединяться с помощью предлога *to*, или перед прямым дополнением без предлога - *give smth to smb/give smb smth*

Следует отличать глаголы данного класса от случаев, когда адресат всегда выступает только с предлогом *to* независимо от позиции. Эти глаголы относятся не к дативным, а к классу **“Prepositional or phrasal verb”** например, *to explain smth to smb.* или в следующем примере:

(18) *Then she **described me** > **described to me** the days spent in this wonderful city.*

Составные глагольные сказуемые в английском языке представлены в классификации глагольных моделей управления тэгами **“Verb+Infinitive”** или **“Verb+Gerund”** или **“Verb+Infinitive OR Gerund”** - они все собраны под тэгом **“Two verbal forms in the predicate.”** Приведем пример ошибки, которую разметили в корпусе одним из этих тэгов.

(19) *If one day the majority of people **start uses** >**start using** sun energy for vehicles it will be great success.*

В примере (19) финитная форма глагола *USE* невозможна при глаголе *START*; он допускает употребление после себя или инфинитива, или герундия. Соответственно, эту ошибку можно исправить двумя способами – *start to use* или *start using*, поставив при этом тэг **“Verb+Infinitive OR Gerund”** (NB: в поле **Notes** исправление должно быть только одно, без указания вариантов!)

“Verb+that+Subjunctive Clause” – к этому классу относятся случаи, когда глагол требует употребление после себя так называемой клаузы в сослагательном наклонении, присоединяемой через союз *that*. Приблизительный список глаголов, которые требуют после себя Subjunctive clause: *to ask, to demand, to insist, to propose, to recommend, to request, to suggest, to urge* - по существу, это все те глаголы, русские эквиваленты которых управляют придаточным предложением, которое вводится союзом *чтобы* - за исключением глагола *WANT* (он управляет только инфинитивом).

(20) *Will Gardner, a founder of a law firm, informed his employee Alicia about this case and **demanded to** > **demanded that she** deal with it.*

В предложении (20) на область ошибки **demanded to** следует выбрать из тэгов модели управления глагола тэг **“Verb+that+Subjunctive Clause.”**

Таким образом, мы описали аннотирование нескольких сложных ошибок в выборе модели управления английских глаголов - области, которая требует особо внимательной проверки от аннотатора.

II – с. Разметка синтаксических ошибок

Ошибки в синтаксических построениях классифицируются в основном по типу конструкций, в которых они встретились. Имеются следующие ярлыки: **“Agreement”** (для ошибок в согласовании по числу подлежащего и сказуемого или существительного и его *determiner* типа указательного местоимения); **“Word order”** с делением на тэги **“Standard word order”**, **“Emphatic shift”**, **“Cleft sentence”** и **“Interrogative word order”** - из них, памятуя, что тэг должен классифицировать ту конструкцию, которая требуется, а не ошибочную, встретившуюся в тексте учащегося, чаще всего используется **“Standard word order.”** Тэг **“Incomplete sentence”** используется для случаев, когда в тексте требуется назывное предложение, и у него есть три нижестоящих тэга - **“Exclamation”**, **“Title structure”**, **“Note structure.”** В примере (21) ошибка допущена в предложении, не относящимся к трем подтипам неполных предложений, поэтому выбираем тэг **“Incomplete sentence”** для разметки этой ошибки:

(21) *I know who can do it: **parents or teachers can** > **parents or teachers.***

Далее следуют тэги **“Attributes”**, подразделяющийся на **“Relative clauses”** и **“Attributive participial construction”**; **“Parallel constructions”** для ошибок, заключающихся в выборе разных форм в однородных членах предложения и в перечислении; **“Negation”** для разметки случаев ошибочного двойного отрицания или отсутствия отрицательных форм там, где они нужны; **“Comparative constructions”** с одним подклассом - **“Numerical comparison”** для ошибок такого рода:

(22) *Twice more > **twice as many people** in this age group use LinkedIn.*

III. Разметка лексических ошибок

В подавляющем большинстве лексические ошибки, допускаемые изучающими английский язык (зеленые тэги класса **“Vocabulary”**), делятся на два класса – ошибки в выборе слова или слов и ошибки в словообразовании. Соответственно, имеются два ярлыка следующего за верхним уровня – **“Word choice”** и **“Word formation”**. У первого дальнейшее деление предлагает различать выбор лексемы (**“Choice of lexical item”**) - а в нем еще отдельно выделяются **«Words often confused”**, т.е. ошибки в использовании слов, которые часто путают (стоит заметить, что этот тег предназначен для слов с похожим звучанием, таких как *rise/raise, lie/lay, lose/loose, then/than*), а не случаев неправильного выбора из слов с похожим значением, таких как *talk/tell/speak/say*), и выбор части идиоматически связанного словосочетания (**“Choice of a part of lexical item”**) с отдельными ярлыками для ошибочно отсутствующего слова (или слов) - **“Absence of a part of collocation”** и лишней части - **“Redundant word(s)”**. Примером, в котором ошибку надо разметить ярлыком **“Absence of a part of collocation”**, является следующее предложение:

(23) *This esolution is very common for both Europe and the USA, where the percentage of people having **own** > **their own** cars is higher, so for Russia it's a bit less effective buta nevertheless it may help a lot.*

Ярлыком из группы ошибок в словообразовании (“**Word formation**”) размечаются, во-первых, случаи неправильного или отсутствующего префикса или суффикса в употребленном слове («**Formational prefix**” и «**Formational suffix**”) и, во-вторых, случаи выбора из словарной семьи не той части речи, которая требуется в контексте. Этот последний ярлык – “**Confusion of categories**” – является единственным ярлыком в этой группе, который описывает природу ошибки, а не тип того, что должно быть в исправлении области ошибки.

В примере (24) имеются две ошибки, размеченные тэгом “**Confusion of categories**”:

(24) *According to the graph, the vadt majority of young **adulting** > **adults** tend to use Facebook or Instagram on a daily basis, with 87 and 53 percent respectively, **contrasting** > **in contrast** to the usage of LinkedIn at only about one fifth of this age group.*

IV. Разметка дискурсных ошибок

Под ярлыком “**Discourse**” собраны тэги, передающие ошибки в логической структуре текста, в построении его частей и в связях между частями. Основные группы тэгов здесь - “**Coherence**” с нижестоящим тэгом “**Linking device**” и отдельные тэги: “**Referential device**” (используется для обозначения ошибок в выборе референциального средства, отсутствия оно, а также его присутствия в случае, когда оно не нужно); “**Inappropriate register**” (для обозначения выражений, которые запрещены в регистре академического письма или неэтичны); “**Absence of component in a clause or sentence**” (для конструкций с отсутствующим элементом, которые в результате этого отсутствия оказываются синтаксически дефектными или бессмысленными); “**Redundant component in a clause or sentence**” (для элемента, присутствие которого вызывает синтаксический дефект или мешает пониманию текста); “**Absence of the necessary explanation or detail**” - для синтаксически правильных и понятных кусков текста, тем не менее требующих уточнения (например, в первом предложении сказано, что график отражает исследование в трех странах, после чего во втором анализируются цифровые данные без указания, к какой стране они относятся).

Примеры применения дискурсного тэга “**Referential device**”.

(25) *It is important to mention that the amount of female graduating students was always higher **than of** > **than that of** male graduates.*

(26) *We have computers, smartphones, laptops, and **it** > **they** influence the way we live, study and work.*

(27) *There is a statement that great Goethe, when he was doing different types of work, **he** > **he** used spectacles with glass of different-colors.*

В предложении (27) после подлежащего вставлено придаточное предложение, после которого следует сказуемое, так что нужды в референции не возникает. Ошибочное употребление референтного личного местоимения привело к тому, что у сказуемого “used” два подлежащих: “Goethe” и “he”, второе из них было удалено. Возможно, это ошибка возникла под влиянием местоимения ‘he’ в клаузе ‘when he was creating different types of work’.

Примеры применения тэга “Inappropriate register”.

Эссе в корпусе REALEC должны быть написаны с соблюдением правил академического письма, поэтому слова с пониженным регистром, жаргонизмы в текстах должны быть заменены на стилистически нейтральные, например: *babies* > *young children*, *very old people* > *the elderly*, *teens* > *teenagers* (кроме выражения ‘in their teens’), и др.

(28) *You can say that business relationship is framed by rules. **That's it.** > **That is true.***

(29) *It is like a basis for] today's literature, which is characterized by simple vocabulary and **pretty** > **easily understandable** plot.*

(30) *Other groups show less distinct change: 70% among the **youngsters** > **young** and...*

Еще один случай применения данного тэга - использование местоимений *he*, *his*, *him* в качестве гендерно-нейтральных:

(31) *If a person is hungry and cold, **he** > **he or she** will not feel much admiration towards the beautiful streets of **his** > **his or her** city.*

Данная ошибка допущена под влиянием русского языка, в котором выбор рода - мужского, среднего или женского – для референтного местоимения производится по грамматическому роду существительного или местоимения, к которому референт относится. Современная письменная традиция в английском языке требует в соотнесении со словом типа «человек» использовать оба местоимения – женского и мужского рода.

Тэг “**Coherence**” имеет один нижестоящий тэг - “**Linking device**” для случаев, когда предложенный союз или вводное слово нарушает логику или когда клаузы или предложения не соединены никакой связкой, а она в данном контексте нужна. Примеры разметки тэгом “**Linking device**”:

(32) *The situation is better with birds and reptiles, as well as with mammals and aquatic mammals (-6% for owls **and** >**but**+23% for bats).*

(33) *Working in the office obliges everyone to work at the same time and with the same rhythm, which can lead to the decrease in the efficiency of working. Secondly, not coming to the office saves a lot of time that people usually spend on the way to work. **Nevertheless** > **Besides**, if more people started working from home, there would be fewer traffic jams, and this would be also better for ecology.*

(34) *Nowadays, with constantly developing computer technologies, new opportunities appear and working conditions change, >**and** the whole world becomes a world of mobility and rapid adaptability.*

Тэг “**Absence of a component in clause or sentence**” используется для случаев, когда отсутствующий элемент вызывает нарушение синтаксической конструкции.

(35) *The only thing we have to notice is that the number of people who were overweight was a little smaller, while **appeared** > **there appeared** more people who were obese.*

(36) *The last group is people 65 years old and above **use** > **who use** social networks on a daily basis.*

В примере (36) имеются две возможности исправления синтаксически неправильного текста - и, соответственно, два разных выбора тэга для разметки. Кроме предложенного выше, можно заменить финитную форму глагола USE на его причастие - *using*. В этом случае аннотатор выберет тэг “**Attributive participial construction**” из синтаксической группы “**Attributes**.”

Тэг “**Absence of necessary explanation or detail**” отличается от предыдущего тэга тем, что он используется для областей ошибки без нарушения синтаксической или семантической правильности высказывания, однако, в них отсутствует элемент, оговоренный в

предшествующей части эссе или требуемый какими-то другими логическими соображениями.

(37) *As we can see, there were no obese or even overweight respondents among young adults in 1950. The percentage of **people** > **people in this group** whose weight can be named ideal is about 70%.*

В примере (37) было необходимо уточнить, что процент людей с «идеальным весом» 70% именно среди юношей и девушек, а не среди всех возрастных групп.

Тэг “**Redundant component in clause or sentence**” чаще всего используется для случаев ненужного повторения или для «лишних» слов, которые нарушают синтаксическую или семантическую правильность, не неся при этом никакого важного смысла:

(38) *To summarize all above, it can be said that in case of celebrities the media has to cover their private life mainly because there is a great demand for it from **both opposite** > **both sides** - the stars and the public.*

(39) *Many children receive little attention from their parents **today** > ~~today~~ because both parents work full time nowadays.*

В примере (39) аннотатор может выбрать, что убрать - *today* или *nowadays*, но не следует оставлять оба.

В примере (40) можно увидеть двойную дискурсную ошибку примерно в одной и той же области:

(40) *First of all, it can be observed from the graph that the highest rate of using Instagram and Facebook among adults at the age of 18-29 **by** > ~~by~~ [**“Redundant component in clause or sentence”**] 87 and 53 percent > was 87 and 53 percent [**“Absence of component in clause or sentence”**] respectively.*

Замена предлога *by* на глагол *was* противоречила бы языковой интуиции аннотатора, а удаление лишнего предлога и добавление отсутствующего сказуемого двумя разными тэгами, наоборот, вполне соответствует интуитивному пониманию природы ошибок в данном предложении.

Напоминание о пункте 8 раздела I «Общие рекомендации»: если область ошибки удаляется целиком (как в примерах (39) и первая ошибка в (40)), необходимо поставить галочку в маленьком окошке **Delete** справа внизу в окне аннотации, и тогда на ярлыке появится красный флажок. В поле **Notes** в этом окне аннотации ничего вписывать не надо.

VI. Отношения между тэгами ошибок

После изменений в области некоторой ошибки может возникнуть необходимость изменить какую-то еще область в данном предложении или в следующих предложениях данного текста. В корпусе REALEC предусмотрены два вида отношений между тэгами: **Dependent change** и **Parallel constructions**. Первое отношение возникает, когда, например, замена числа местоимения, являющегося подлежащим, может вызывать замену числа в глагольной форме сказуемого:

(41) *Everybody > They all (“Referential device”) knows > know it.*

Второе отношение требуется для разметки случаев, когда ошибка повторяется во втором (и дальше) однородном члене.

(42) *Many a true word are > is spoken in jest.*

В предложении (41) можно вторую область ошибки (**knows**) разметить тэгом “**Agreement**” или просто тэгом “**Исправление.**” В любом случае от первого тэга “**Referential device**” следует кликом по ярлыку провести ко второму тэгу стрелку, причем система откроет специальное окно отношения и спросит, хотите ли вы установить отношение зависимости (“**Dependent change**”) или отношение однородности изменений, присущее однородным членам предложения (“**Parallel constructions**”). Для примера (42) следует выбрать первое отношение. А в следующем примере требуется установить второе отношение между первым и вторым тэгом:

(43) *History shows that while men are > were (“Choice of tense”) involved in mathematics and technical disciplines, women are > were (“Choice of tense”) mostly involved in arts.*