

# REALEC Tagging Manual

## I. General principles

1. The best browser to access REALEC is Google Chrome.
2. To start annotating a text in the corpus, type in your login name and password in the upper right-hand pop-up window.
3. When you have spotted an error, think which word or words have to be chosen as the error span (see some recommendations for specific errors below) and double click on this word or words. A new annotation window will open on your screen with the error span in the section above, the classification scheme with all error tags in the central area, and with the section **Notes** for the correction of the error span. You can also choose the tag for a comment - «**Исправление**», and in this case you do not have to fill anything in **Notes** (although you can if you need).
4. To choose an appropriate tag for the error you are annotating, click on a plus in front of the first tag («**Исправление**»), and you will see the following tags of the highest level: “**Punctuation**”, “**Spelling**”, “**Capitalisation**”, “**Grammar**”, “**Vocabulary**” and “**Discourse**”. A plus in front of any tag means there is a further division of this tag, and. You will see a plus in front of the last three tags on the list, and recommendations on how to use specific tags in these groups are given in sections **III**, **IV** and **V**, correspondingly, On the contrary, for errors belonging to the first three tags, just use one of these tags - “**Punctuation**” or “**Capitalisation**” or “**Spelling**” (see section **II**).
5. A correction of the error span has to be given in **Notes** section. The word or the expression from the error span can be clicked on and dragged with a mouse into **Notes** until a circle turns into a rectangular.
6. A tag is selected by the required form or construction, and not by the type of error. A few tags named by the type of error are pinpointed below.
7. Annotators have to select the most specific tag possible. One exception to this convention is when a tag is subdivided into a few sublevels, but they do not cover all classes, so a case representing a class for which no specific tag is available will be annotated by a more general tag.
8. If you annotate an erroneous word in a word combination, you have to analyse whether it is the erroneous word itself, or a part of a collocation, or a word that shows the syntactic relation that is the error. The wrong preposition, for example, can be a “**Choice of a part of lexical item**” type of error, as in *put off with his egoism*>*put up with his egoism* with *put off* as an error span, **or** “**Transitive verb**” type of error, as in *see on us* > *see us* with *see on* as an error span, **or** “**Preposition**” type of error, as in *seated on the armchair*>*seated in the armchair* with error span just the preposition - *on*.
9. If in order to correct the error one needs to delete the error span, section **Delete** of the **Entity Attributes** area has to be ticked, and it will show as a red flag on the tag label.

10. Pressing **Move** button of the existing annotation instead of **OK** gives one a way to change the error span.
11. More than one tag can be assigned to an error span. A good example is *loose* written instead of *lose*, for which both “**Spelling**” and “**Words often confused**” are appropriate tags, or - another example: both “**Noun number**” и “**Article**” are required for correcting *the employee* for *employees*. It is important that the correction in **Notes** be the same in all tags for one and the same error span.
12. When all the errors that you have been able to spot have been annotated, press **OK** button to get back to the essay. You will see labels of different colours, and placing a curser on the label you can see the correction suggested for it. A double click on the label will re-open the annotation window you have already been working with.
13. When you open the existing annotation window, you can delete it by pressing **Delete** button next to buttons **OK** and **Move**. The system will want a confirmation of the need to delete the annotation, which you can give by pressing **OK** one more time. N.
14. At the end of your work in the corpus press **Logout** in the upper right-hand pop-up window.

## II. Annotation of errors with “Punctuation”, “Spelling”, and “Capitalisation” tags

These categories have no subdivision.

The error span for “**Punctuation**” tag will be:

- the introductory or parenthetical expression that has not been separated with a comma at the end or with two commas, at the beginning and at the end - it has to be copied in **Notes** area with the comma after it or with commas before and after it;
- the conjunction or the conjunction with a comma, if a comma is missing before a conjunction or if there is a redundant comma before a conjunction, and the correction in **Notes** will have the same conjunction, correspondingly, with or without a comma.

See section IV about NOT applying “**Punctuation**” tag for mistakes in Relative clauses.

## III. Annotation with tags in the group “Grammar”

There are orange tags in this group, which correspond to the area of morphology, and blue tags for the errors in syntax.

Depending on the part of speech of the word with an error, an annotator has to look up tags in the following groups – “**Determiners**”, “**Verbs**”, “**Nouns**”, “**Prepositions**”, “**Conjunctions**”, “**Adjectives**”, “**Adverbs**”, “**Numerals**”, “**Pronouns**”. All except “**Prepositions**” and “**Conjunctions**” have subdivisions of tags, and in some the subdivision is quite detailed. The first tag in “**Verbs**” is “**Tense**,” which is further divided into “**Choice of tense**” & “**Tense form**.” For example, a correctly constructed tense form inappropriate for the context will be

tagged with “**Choice of tense**” tag, while a form like *have took* will be marked with “**Tense form.**”

The most detailed division is in the group “**Verb patterns**”. Tags in this group show the pattern that should have been used instead of the one that has been used in the essay, and not only for the verbal forms in the predicate, but also for errors in non-finite verbs – infinitives, bare infinitives, gerunds and participles. If an error has been made, say, in the preposition that follows the adjective with the auxiliary verb BE (*be proud of, be confident of, etc.*), the error is NOT tagged as one of the “**Verb patterns**” cases but rather as “**Prepositional adjective**” tag. The first three examples show when we do use “**Verb patterns**” tags:

(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.*

To show that a gerund has the pattern of attaching the object without a preposition, we apply “**Transitive Verb**” tag to the error span of the gerund with the preposition.

(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.*

To indicate the error in the preposition following the participle of the verb *SUFFER*, we apply “**Prepositional Verb**” tag on the error span of the participle with the preposition.

It is important to distinguish prepositions that are a part of the verb from prepositional groups that form a separate adverbials:

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

The wrong preposition in (3) is not related to the pattern of the verb *JUMP* as it does not introduce an object as an obligatory element of the verb - it is a part of the adverbial and has to be tagged with “**Prepositions**”.

There are cases when - possibly, under the influence of a Russian construction - a student uses a preposition to introduce an object, and then the tag will show the pattern that is more appropriate, as in example (4):

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

In other cases, a learner may use a verb that requires a reflexive pronoun as an object without an object, because in its Russian equivalent’s pattern a special ending, and not an object, is required. “**Reflexive Verb**” tag is to be assigned then:

(5) *Please **introduce** > **introduce yourself**.*

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

Often Russian learners make the mistake of adding a redundant reflexive pronoun after verbs that are not used with them in English, but their Russian equivalents do have reflexive pronouns. Such verbs as in examples (7) and (8) can be used as transitive or intransitive, and their pattern is called “**Ambitransitive verbs**” in our scheme:

(7) *Most people cannot **afford themselves** > **afford** to go abroad for a holiday.*

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

Russian students often confuse the patterns of verbs like CONSIDER/THINK/BELIEVE... SOMEBODY TO BE with verbs REGARD SOMEBODY AS and CONSIDER SOMEBODY AS. In cases like those in examples 9 and 10 annotators have to use “**Verbs with as**” tag:

(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.*

In (11) and (12), on the contrary, an annotator needs to correct the verb CONSIDER for the pattern “**Verb+Infinitive**” :

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

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

In example (13) the verb in the predicate is used in the pattern similar to the verb GIVE (like *She gave me a bag*, which is covered by “**Dative verb with alternation**” tag). However, DESCRIBE is a transitive verb in which the addressee is not obligatory, but if it is given, it is introduced with preposition *to*, so instead of a tag for a specific verb pattern we need “**Preposition**” tag for the pronoun *me*:

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

### III – c. Annotating syntactic mistakes

Mistakes in syntax are usually classified by the type of construction where they are in. The tags are blue in our scheme, they also belong to “**Grammar**” group of tags, and they are the following:

- “**Agreement**” (for mistakes in number agreement between subject and predicate or noun and its demonstrative pronoun determiner)
- “**Word order**” (further subdivided into four more precise tags: “**Standard word order**”, “**Emphatic shift**”, “**Cleft sentence**” and “**Interrogative word order**”). As a tag describes the construction that should have been used and not the one that was used in the text, “**Standard word order**” is used more frequently than the others in this group; “**Incomplete sentence**” tag is used when the text needs a sentence without a predicate, it can be further subdivided into three tags: “**Exclamation**”, “**Title structure**”, “**Note structure**.” In example (14) the mistake was not made in either of those three types of incomplete sentences. That is why “**Incomplete sentence**” tag is chosen, without being further specified:

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

- “**Attributes**”, which is subdivided into “**Relative clauses**” (three types of them make up three tags below this one) and “**Attributive participial construction**”

- “**Parallel constructions**” for mistakes arising from incoherent choice of forms for coordinated parts of the sentence and enumeration
- “**Negation**” for annotating cases double negation unacceptable in English, or the lack of necessary negation forms
- “**Comparative constructions**” with a subtag - “**Numerical comparison**” for the mistakes of the following kind:

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

#### IV. Annotating lexical mistakes

The overwhelming majority of lexical mistakes made by the learners of English (green coloured tags in the group named “**Vocabulary**”) can be divided into two types - mistakes in word choice and mistakes in word formation. Correspondingly, there are two categories of tags under “**Vocabulary**”: “**Word choice**” and “**Word formation**”. The former is further subdivided into “**Choice of lexical item,**” with “**Words often confused**” for commonly confused words like *rise/raise, lie/lay, lose/loose, then/than*; and “**Choice of a part of lexical item**” for errors in a part of an idiomatic expression, with more specific tags “**Absence of a part of collocation**” and “**Redundant word(s)**”. (16) serves as an example of a mistake that should be annotated with “**Absence of a part of collocation**” tag:

(16) *This solution 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 but nevertheless it may help a lot.*

Word formation group of tags is used for mistakes, firstly, in choosing suffixes and prefixes or in the lack of them (“**Formational prefix**” and “**Formational suffix**”), and secondly, for cases of choosing the wrong part of speech from a word family. This last tag – “**Confusion of categories**” – is the only tag in the group that describes the nature of the mistake and not that of the correction.

In example (17) there are two mistakes which were assigned “**Confusion of categories**” tag:

(17) *According to the graph, the vast 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.*

#### V. Annotating discourse mistakes

“**Discourse**” group of tags helps annotate mistakes affecting the logical structure of the text, mistakes in construction of its parts and connections between them. The principal subgroups of tags here are - “**Coherence**” with a subtag “**Linking tool**“, separate tags “**Referential device**”

(used for mistakes in referential devices, their absence or superfluousness); “**Inappropriate register**” (for annotating expressions inadequate for academic writing, and sexism in language like using *he, him, his* when referring to people of both sexes); “**Absence of component in a clause or sentence**” (for constructions lacking a certain element and, thus, difficult or impossible to understand); “**Redundant component in a clause or sentence**” (for an element whose presence causes a syntactic defect or gets in the way of understanding the text); “**Absence of the necessary explanation or detail**” (used for syntactically correct and legible pieces of text that call for further explanation, for example, in the first sentence the author tells us that the graph reflects the results of the study conducted in three countries, yet in the second sentence he or she proceeds to analyse the data without specifying which country it applies to).

Examples of using discourse tag “**Referential device**”:

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

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

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

In sentence (20) the subject is followed by a subordinate clause, which is followed by the predicate, so no reference is necessary. Erroneous usage of referential personal pronoun leads to the predicate having two subjects at once: ‘Goethe’ and ‘he’, so the latter was deleted. It is possible that the mistake was caused by the influence of the pronoun ‘he’ in clause ‘when he was creating different types of work’.

Essays in REALEC corpora must be written with respect to the rules of academic writing. That is why words of the low register and jargonisms are substituted with stylistically neutral equivalents, for example: *babies > young children, very old people > the elderly, teens > teenagers* (except for ‘in their teens’), etc.

Examples of using “**Inappropriate register**” tag:

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

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

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

One more example of assigning this tag - a case of sexism:

(24) *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.*

That mistake was made because of Russian language interference, where the choice of gender for a referential pronoun - masculine, feminine or neutral - depends on grammatical gender of the noun or pronoun to which the reference is made. Modern written tradition of English requires to assign both pronouns of feminine and masculine gender as a reference to words like ‘a person’.

”Coherence” has “Linking tool” as its subtag for cases when the linking device or parenthesis used in the text disrupts the logical structure, or when clauses or sentences lack a necessary link between them.

Examples of using “Linking tool” tag:

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

(26) *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.*

(27) *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.*

“Absence of a component in clause or sentence” tag is used when the element missing makes the syntactic construction incorrect.

(28) *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.*

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

In example (29) there are two possibilities for correcting the syntactically erroneous text, which is why either of the two tags are possible to be assigned. Besides the way suggested in (29), there’s another possible correction - to substitute the verb *USE* for its participle - *using*. In this case the annotator would choose “Attributive participial construction” tag from “Attributes” group of syntax tags.

“Absence of necessary explanation or detail” tag differs from the previous tag in that it is used for error spans that are syntactically and semantically correct but lack an element that was discussed in the previous sentences or is necessary for some other reasons.

(30) *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%.*

In example (30) it was necessary to clarify that the percentage of people with “ideal weight” is given among young men and women and not among all age groups.

“Redundant component in clause or sentence” tag is more commonly used for the cases of unnecessary repetition or for redundant words that disrupt the syntactic or semantic structure of the clause or sentence without conveying any important information:

(31) *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.*

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

In example (32) the annotator can choose which word is to be deleted - *today* or *nowadays*, but both of them cannot remain in the sentence.

In example (33) there are two discourse mistakes in close proximity:

(33) *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.*

Substituting preposition *by* with verb *was* would be against the language intuition of the annotator, while deleting the redundant preposition and adding the lacking predicate *is*, on the contrary, quite logical considering the nature of the mistakes in this sentence.

**Reminder:** If the entire error span is deleted (like in examples (32) and the first error span in (33)), it is necessary to tick **Delete** section at the bottom of the annotation window. Do not write anything in the **Notes** section in this case.

## VI. Relations between tags

Sometimes after the changes have been made in one error span, the following element(s) may also need to be corrected. In REALEC corpora there are two types of relations between tags: **Dependent change** and **Parallel constructions**. The former is assigned when, for example, the correction of the pronoun in the subject from Singular to Plural, or the other way round, calls for correction of the verb form of the predicate:

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

The second relation is needed for annotating the cases when a mistake occurs once again in the following homogeneous part(s) of the sentence.

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

In sentence (34) the second error span will be annotated with **“Agreement”** tag, and a relation has to be established between this error span and that of **“Referential device”** tag. This can be done by clicking on the first tag and dragging an arrow to the second one. In the new window the annotator should choose which relation he or she wants to use: **“Dependent change”** relation for changes that trigger other changes, or **“Parallel constructions”** relation for identity of changes between coordinated parts of the sentence. In example (35) the former relation should be chosen. In the next sentence the latter relation should be established between the first and the second tag:

(36) *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.*