



International Conference
The Future of Education

USE OF LEARNER CORPUS IN GENERAL ENGLISH AND ACADEMIC ENGLISH COURSES AT THE HIGHER SCHOOL OF ECONOMICS

Olga Vinogradova, olgavinogr@gmail.com

**School of Linguistics, National Research University Higher
School of Economics**

Moscow / Russia

The study was implemented in the framework of the Basic Research Program at the National Research University Higher School of Economics (HSE) in 2015-2016, and the author is a member of the team that has won a Research Team Project Competition in 2016



REALEC

- **Russian** – texts written by speakers of Russian
- **Error-Annotated** – the mistakes in the texts annotated manually
- **Learner** – texts written by learners=HSE students
- **English** – texts written in English
- **Corpus**



International Conference

The Future of Education

In the open access, available at <http://realec.org>

Set up by members of the faculty at the School of Linguistics (Higher School of Economics)

By now almost 3400 pieces of HSE students' writing with about 838,000 word tokens

About 4000 more student essays are ready to be uploaded and annotated



International Conference The Future of Education

Olga

REALEC: Russian Error-Annot x

realec.org



Russian Error-Annotated Learner English Corpus



[Browse](#)



[Download](#)



[About](#)



[Contacts](#)



Russian Error-Annotated Learner English Corpus by REALEC Team is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](#). Permissions beyond the scope of this license may be available [here](#).

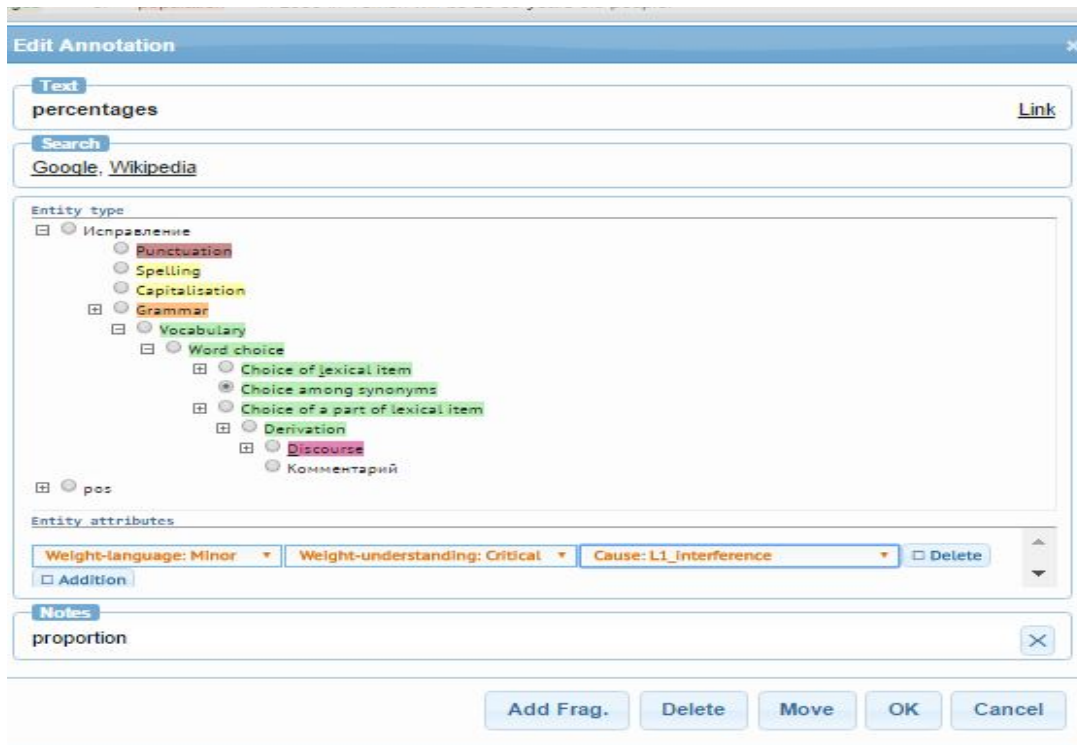


HIGHER SCHOOL OF ECONOMICS
NATIONAL RESEARCH UNIVERSITY



A well developed system of hierarchical tags to mark the errors (tags of different colours) and Tagging Manual (http://realec.org/tagging_instructions.pdf).

Annotation window with error tags scheme:





International Conference The Future of Education



- Исправление
 - Punctuation
 - Spelling
- Grammar
 - Determiners
 - Articles
 - Choice of articles
 - Form of articles
 - Choice of determiners
 - Form of determiners
 - Quantifiers
 - Choice of quantifiers
 - Form of quantifiers
 - Verbs
 - Nouns
 - Prepositions
 - Conjunctions
 - Adjectives
 - Adverbs
 - Numerals
 - Pronouns
 - Agreement
 - Word order
 - Incomplete sentence
 - Conditionals
 - Attributes
 - Parallel constructions
 - Negation
 - Comparative construction
- Vocabulary
- Discourse
- Комментарий
- pos



International Conference The Future of Education



- ☐ ○ Agreement
 - Agreement - Animacy
 - Agreement - Number
 - Agreement - Person
- ☐ ○ Word order
 - Standard word order
 - Emphatic shift
 - Cleft sentence
 - Interrogative word order
- ☐ ○ Incomplete sentence
 - Exclamation
 - Title structure
 - Note structure
- ☐ ○ Conditionals
 - Choice of conditionals
 - Form of conditionals
- ☐ ○ Attributes
 - ☐ ○ Relative clause
 - Defining relative clause
 - Non-defining relative clause
 - Coordinate relative clause
 - Attributive participial construction
- Parallel constructions
- Negation
- ☐ ○ Comparative construction
 - Numerical comparison



International Conference The Future of Education



- ☐ ● Vocabulary
 - ☐ ● Word choice
 - ☐ ● Choice of lexical item
 - Words often confused
 - Choice among synonyms
 - ☐ ● Choice of a part of lexical item
 - Absence of certain components of a collocation
 - Redundant word(s)
 - ☐ ● Derivation
 - Conversion
 - ☐ ● Formational affixes
 - Formational suffix
 - Formational prefix
 - Confusion of categories
- ☐ ● Discourse
 - ☐ ● Referential device
 - Lack of referential device
 - Dangling reference
 - Redundant reference
 - Wrong choice of referential device
 - ☐ ● Coherence
 - Incoherent articles
 - ☐ ● Incoherent tenses
 - Incoherent tenses in conditionals
 - Incoherent pronouns
 - ☐ ● Linking device
 - Incoherent conjunction
 - Incoherent introductory unit
 - Lack of connective tool
 - Inappropriate register
 - Absence of a component in clause or sentence
 - Redundant component in clause or sentence
 - Tautology
 - Absence of necessary explanation or detail



- **A sentence from a student essay with tags for errors spotted by the annotator – Example 1**

6 At the South coast there is a pier large enough to two boats fit in.

Annotations:

- Prepositions (At)
- Confusion of categories (South)
- Capitalisation (South)
- Standard word order (to two boats fit)
- Redundant component in clause or sentence (in.)

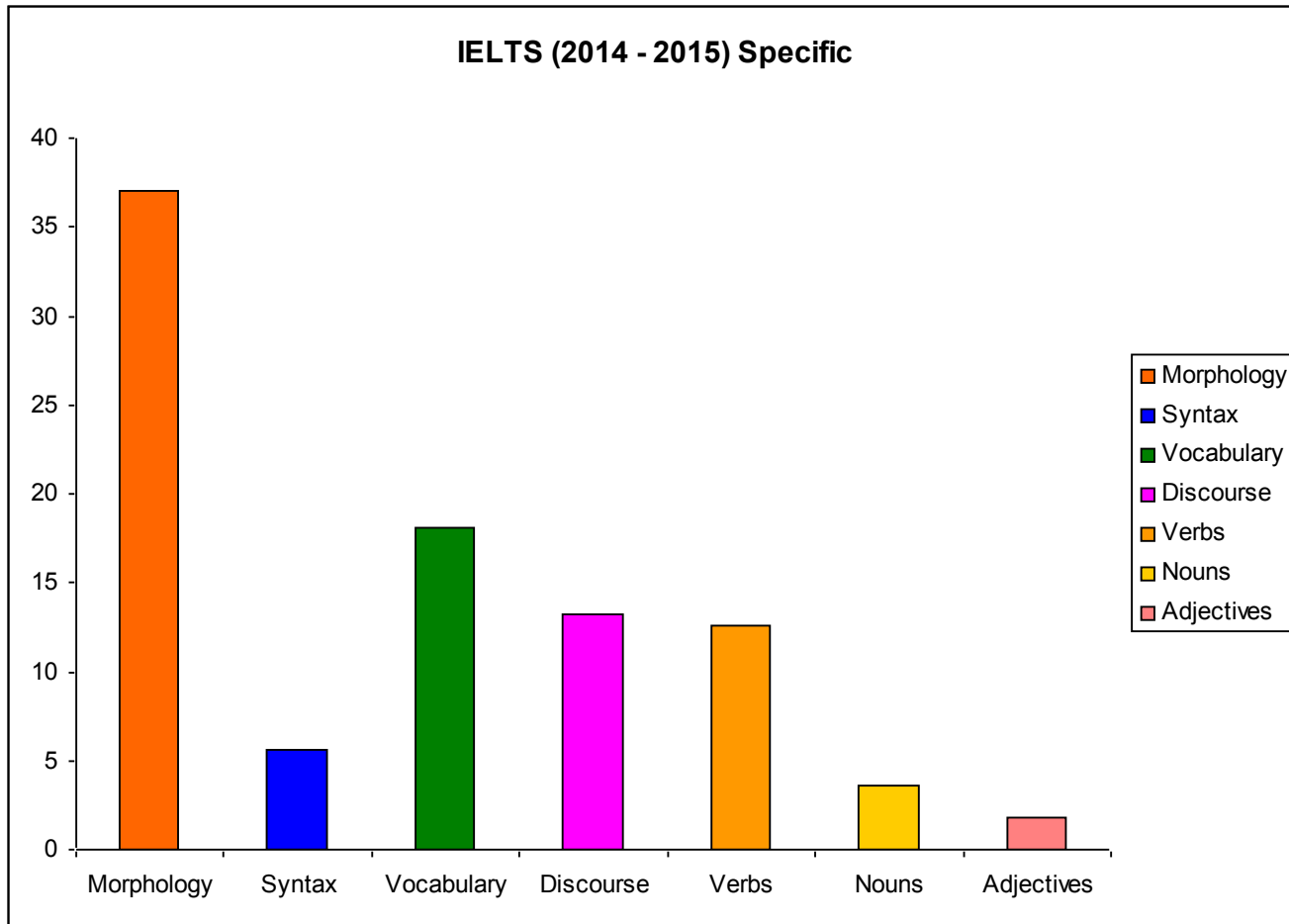


Example 2

- 1 The charts illustrate the changes between the ages of the population in two countries: Italy and Yemen, in two period of time: 2000 and 2050.
 - Agreement - Number
- 2 We can see that children and teenagers (people from 0 to 14) in both country lose their position in proportion of population.
 - Agreement - Number
 - Choice of tense
 - Dependent change
 - Choice of articles
 - Absence of necessary explanation or detail
 - Choice of articles
- 3 And elderly people have a bigger proportion in 2050 than in 2000 in Italy and Yemen.
 - Choice of tense
 - Lack of referential device
 - Comparative construction
- 4 The differences between the Italy and Yemen is that in Yemen the number of people from 15 to 59 years old is risen and in Italy is declined.
 - Choice of articles
 - Tense form
 - Choice of tense
 - Choice of tense
 - Absence of a component in clause or sentence
- 5 Also, the biggest percentages of population in 2050 in Yemen will be 15-59 years old people.
 - Noun number
 - Choice among synonyms
 - Choice of articles
- 6 In fact, in Italy we can see the similar situation in the future, but the proportion will be different.
 - Choice of articles
- 7 To sum up, the charts show that the ages of the populations can be different in past, present and future time.
 - Absence of a component in clause or sentence
 - Absence of necessary explanation or detail
 - Redundant component in clause or sentence
- 8 In both country we can see the changes between 2000 and 2050 in the number of the populations (depends on ages).
 - Agreement - Number
 - Word choice
 - Punctuation
 - Absence of necessary explanation or detail



International Conference The Future of Education





- ***twice lucky*** > ***twice as lucky*** (text 11) - the same correction, different tags:
- “Absence of certain component” (a vocabulary tag) – 1 annotator
- “Numerical comparison” – 2 annotators
- “Comparative degree of adverbs” – 2 annotators - wrong tag!
- “Prepositions” – 1 annotator - wrong tag!
- “Absence of a component in clause or sentence” (a discourse tag) _ 1 annotator
- ***twice lucky*** > ***double lucky*** (text 11) - different corrections, different tags (“Vocabulary” – 1 annotator)



- **Experiment 1:**

10 annotators

30 essays (150-300 words)

Preliminary error spans marking

Total 2128 error tags assigned

KA = 0.57

- **Experiment 2:**

12 annotators

1 essay (600 words)

Error spans not marked

Total 156 error tags assigned



12 annotators

the same text about 350 words long

The total number of error spans marked in this text - 156.

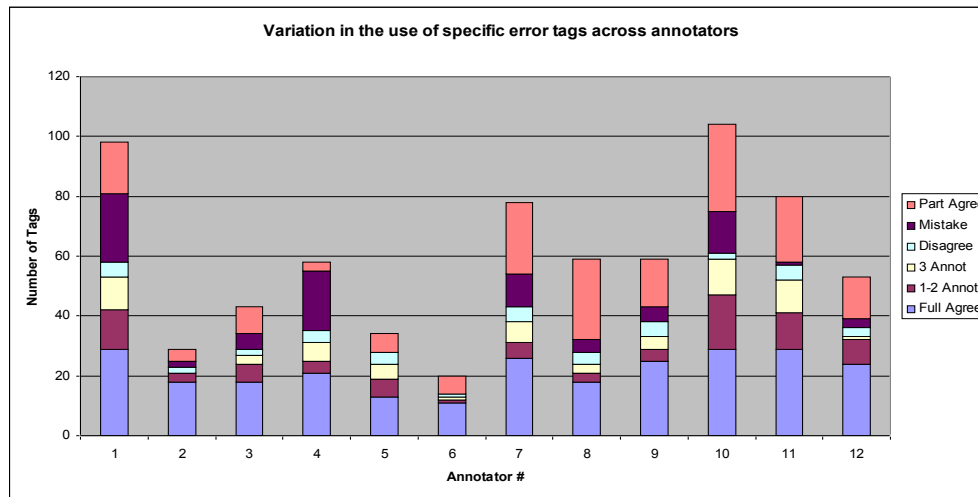
57 were spotted by no more than 2 annotators

23 were spotted by only 3 annotators

30 errors were marked by at least 10 annotators of the 12 participants with the same tag

6 errors spotted by at least 10 annotators were marked with different tags

40 tags noticed by 4 to 9 annotators, in 19 the annotators agreed in their choice of tags



Variation in the use of specific error tags by annotators in Experiment 2



- **RETM** – **REALEC** English **T**est-**M**aker

In-built transition to a more sophisticated question after the success, and to lower-level complexity in case of a failure.

- placement test
- custom-made progress tests
- lexical trainers
- training exercises and games for new annotators



International Conference The Future of Education



http://web-corpora.net/realec/course/index.php

по нису - olgavinogr... REALEC: All courses

View Favorites Tools Help

REALEC

Russian Error-Annotated Learner English Corpus

[Home](#) ▶ [Courses](#)

NAVIGATION

[Home](#)

▶ [Courses](#)

 [Test for linguists on May 30](#)

Teacher: [Random Teacher](#)

 [Test for linguists](#)

Teacher: [Random Teacher](#)

 [Current version of test](#)

Teacher: [Random Teacher](#)

 [Simple test sample](#)

Teacher: [Random Teacher](#)



International Conference The Future of Education



Children, teenagers and even adults spend **huge amount** of time trying to succeed academically, because education is associated with prestige and success.

Answer:



Correct!

The correct answer is: a huge amount

adpq_2

The bigger part is in the **east part** of this tourist village, and the smaller part is situated on the other side of the reception.

Answer:



The correct answer is: east

1

Some parents are so busy that they do not have any time to explain to their children, for instance, how to cook **meal**.

Answer:





International Conference

The Future of Education

- Automatic generation implies that a test-taker will have to correct what (s)he sees as an error, and his/her correction will be compared with the one given by an expert in the annotation: if they coincide, then the test-taker has won a score assigned to the question.



1. Choice of what to test.

- Which tags constitute relevant material for the test?
- What if a span includes multiple errors in one span?
- Type of test - highlighting the error span, giving the sentence without any highlighting, or giving the sentence as a multiple-choice question?
- No mistakes that learners make very rarely, nor accidental slips.
- No errors too difficult to spot
- No errors with multiple options of correction



2. Preparation of the selected sentences

according to the level of difficulty it poses for a learner. At present the system allows to assign any question one of the three levels – the lowest (1 point), middle-level (2 points), and the highest (3 points). If for some reason it is necessary, the number of levels can be increased or decreased.



3. Test administration

- questions randomly chosen from the pool
- all test-takers get the same number of questions
- start at the lowest level of difficulty
- success > the next question is taken from the pool of middle-level difficulty; failure > the next question is also of the lowest level.



4. Analysis of the testing statistics.

- a test-taker gets the number of correct answers, the number of correctly spotted error spans with the wrong correction suggested, and all the wrong answers are presented along with the expected answers in a way of feedback
- an instructor gets the statistics for the whole group in the form of the list from the best to the worst
- in a placement test, the system offers to add other criteria to sort out the division of students into the necessary number of groups
- in a progress test, a test-taker with the low score can be urged to take the test more times until (s)he reaches the decent level (the same questions are excluded)



International Conference

The Future of Education

Problem – the need for more annotators

- Crowdsourcing?
- International cooperation?
- Automated annotation?
- Automated evaluation?



International Conference

The Future of Education

Olga Vinogradova

olgavinogr@gmail.com

School of Linguistics

National Research University Higher School of Economics

Moscow, Russia

THANK YOU!