

Design of test-making tools for the learner corpus¹

Olga Vinogradova and Ekaterina Gerasimenko (National Research University Higher School of Economics, Russia)

Russian Error-Annotated Learner English Corpus (further REALEC) is an electronic collection of texts written in English by Russian university students studying English at an intermediate – upper-intermediate level. The corpus is easily available at <http://realec.org/#/realec/> and includes over 1200 texts (about 360 thousand word tokens), mostly essays written in preparation for IELTS-type English examination taken at the end of the second-year of Bachelor studies, and the newer, growing sub-corpus of 2nd-year Bachelor students' essays written in the English examination at http://realec.org/#/hse/data_4_staff/IELTS/ consisting by the beginning of 2017 of about 3000 essays comprising almost 700 thousand word tokens. The value of a learner corpus is in its annotation practice [Granger, S. (2003), Leech, G. (2015)], so experts at the Higher School of Economics (teachers and students familiar with annotation approaches adopted in REALEC) annotate errors in the essays with the help of the error classification scheme established at the start of setting up the corpus [Kutuzov, A. and Kuzmenko, E. (2014)]. The error annotation scheme consists of 152 categories organized into a tree-like structure. Annotators are instructed to choose a specific tag for the error they have spotted, and unified approaches to annotation have been worked out [Kutuzov, A., Kuzmenko, E., and Vinogradova, O. (2015)].

REALEC is not only a ground for erratological statistics and analyses with results applicable in language acquisition [Vinogradova, O. (2016)], but also a great resource of sentences that can be used in tests for English learners, as annotated sentences in the corpus have all the necessary elements for generating testing questions fully or semi-automatically, namely, error spans are outlined, errors are labeled, and the corrected option is offered by the annotator. The current paper presents RETM – REALEC English Test Maker, the system that works as a tool to automatically generate tests for students on the basis of the errors that experts have marked in student works submitted to REALEC. With the help of the scripts written in Python (source code can be found at <https://github.com/ElizavetaKuzmenko/realec-exercises>). RETM extracts the necessary testing questions from Brat and transforms them into the XML file required for the format of the user interface.

At first the script parses the annotation files (initially stored separately for each text) and marks annotations in the texts using different tags for relevant error types and for errors that have to be corrected. Afterwards, the script takes annotated texts, corrects all mistakes and creates questions for relevant mistakes if there are any. The annotations marked in the texts contain the right answer (suggested by the expert annotator in the corpus) and indices that are used to insert words in an open cloze or to

¹ This article is an output of a research project implemented as part of the Basic Research Program at the National Research University Higher School of Economics (HSE).

highlight the mistake in the question text. Subsequently, questions and right answers are passed to the XML formatter and then uploaded manually to the user interface.

For the sake of the convenience of working with the tests, the interface was set up in Moodle (Modular Object-Oriented Dynamic Learning Environment), the open source learning platform which has progressively been found well applicable in teaching processes (<https://moodle.org/>). This platform provides a user-friendly environment for instructors to prepare and administer the test they want and get the statistics on their students' performance after the test. The convenience works equally well for testees as throughout the test they get as much feedback and support as their instructor chooses for them. The tests that have already been applied can be seen at <http://web-corpora.net/realec>, the space where both instructors and their students undergo the necessary authorization.

After the first introduction of RETM in 2016 [Kustova, M. (2016), Vinogradova, O. (2016-2)], the group of linguists updated the system and expanded its functions by adding new types of tests and giving some new components of the automated generation test-making process a try. At the moment there are about 30 different tests with about 1200 testing questions altogether. These recent achievements are in the focus of the present talk.

The experimental ground for the research was formed with two tests in Moodle. The questions for these tests were selected from the pool of sentences generated at the first stage from the sentences of student essays by correcting all error spans except for one for the option given as the correction by the expert annotator. At the second stage the testing instructor looks at each testing question in order to make sure that:

- there is no more than one error left in the sentence;
- the error span can be clearly identified;
- the context of one, two, or three sentences will be enough for a testee to be able to spot and correct the error.

At this stage a test-maker can edit a question, delete a part of the question, delete a sentence from the pool altogether, and make a decision what kind of test will best fit the testing purposes for this error. In the reported research we have applied the following two test types, with two subtypes in either of them, namely:

- correcting errors (short answer): spot and correct the error, or correct the highlighted error.
- filling gaps (open cloze): give the appropriate form of a word in the parentheses, or fill the gap with the appropriate word or words.

Thus, the overall process of creating and testing questions is the following:

1) extracting sentences containing the error. The annotations are visualised in the user interface:

1	Nowadays	the problem of public health	become	sagnificantly	important worldwide.
2	Many people all over the world forget about their physical development and health because of	whole new set	of opportunities	which modern time has.	

In this example, the question is created from the error marked with the tag 'Choice of tense'.

2) creating questions automatically by processing the relevant error and correcting all other errors. There are two formats of output – XML (for uploading questions to Moodle) and plain text.

Nowadays, the problem of public health {1:SHORTANSWER:=has become} incredibly important worldwide. Many people all over the world forget about their physical development and health because of a whole new set of opportunities we have today.

3) editing manually:

Nowadays, the problem of public health {1:SHORTANSWER:=has become ~is becoming}(become) incredibly important worldwide. Many people all over the world forget about their physical development and health because of a whole new set of leisure opportunities they have today.

Question 16
Not yet answered
Marked out of 1.00
Flag question
Edit question

Nowadays, the problem of public health (become) incredibly important worldwide. Many people all over the world forget about their physical development and health because of a whole new set of leisure opportunities they have today.

4) the testees answer the question and get the score (three options of answers are presented):

Question 22
Correct
Mark 1.00 out of 1.00
Edit question

Nowadays, the problem of public health ✓ (become) incredibly important worldwide. Many people all over the world forget about their physical development and health because of a whole new set of leisure opportunities they have today.

Question 11
Correct
Mark 1.00 out of 1.00

Nowadays, the problem of public health **is becoming** ✓ (become) incredibly important worldwide.

Many people all over the world forget about their physical development and health because of a whole new set of leisure opportunities they have today.

🚩
⚙️ Edit question

Question 13
Incorrect
Mark 0.00 out of 1.00

Nowadays, the problem of public health **have become** ✗ (become) incredibly important worldwide.

Many people all over the world forget about their physical development and health because of a whole new set of leisure opportunities they have today.

Incorrect
The correct answer is:
has become
Mark 0.00 out of 1.00

🚩
⚙️ Edit question

In the process of editing, the testing instructor may also apply one of the additional codes to signify the level of complexity of each question (up to 3), so that the test can be administered as an adaptive quiz: it starts with the easiest question, and if the answer is correct, the next question is more difficult, but if the answer is incorrect, the next question stays at the lowest level of complexity. This is repeated at each step: the success in answer leads to the increase in the complexity of the next question, and the failure, to the lower level of complexity. Correspondingly, the correct answers have different weights depending on the level of the question.

The following data can show the efficiency of RETM-produced tests administered to two groups of first-year students of the Higher School of Economics:

- ❖ Number of tests - 2
- ❖ Number of questions in each test: 1 - 50, 2 - 60;
- ❖ Number of students - 54
- ❖ Testing areas - tense choice, tense form, voice choice, infinitive construction, gerund construction, participial construction
- ❖ Range of correct answers: 46% to 81% with the average success of 66%
- ❖ Increase in the success rate in diagnostic samples from pre-testing to post-testing in 67% of students, the same rate of success in 13% of students, and decrease in 20%.
- ❖ Number of questions deleted at the editing stage - 12 of 122.
- ❖ Time for editing 122 questions - about 50 minutes.

With the help of such semi-automatically generated tests English instructors can make sure that their students should not repeat the mistakes that they themselves have previously made in their written works. Moreover, the tests can be administered as a specially designed course of tests from simpler to gradually more sophisticated. Finally,

the courses of tests can be a part of work in class or of independent learning trajectory, while the instructor gets the necessary statistics in any mode of administration.

The tests produced by RETM allow English instructors to save much time required for the process of collecting and organizing testing materials and grading the results of the tests, even though instructors do have to carry out some preliminary evaluation in the semi-automatic regime at the stage of finalising the testing questions. In addition to being a useful tool for the teaching-learning process, RETM tests also provide the area for more research into language acquisition and make the teaching process much more custom-made.

References

- Granger, S. (2003) The international corpus of learner English: a new resource for foreign language learning and teaching and second language acquisition research. *Tesol Quarterly*, 2003, pp.538-546.
- Kustova, M. (2016) Design of corpus-generated EFL placement and progress tests for university students. *Conference Proceedings. The Future of Education. – libreriauniversitaria. 2016*, pp. 148-152.
- Kutuzov, A., Kuzmenko, E., and Vinogradova, O. (2015) Evaluating inter-rater reliability for hierarchical error annotation in learner corpora, in *Corpus Linguistics 2015: Abstract Book*, pp. 211-213.
- Kutuzov, A. and Kuzmenko, E. (2014) Russian Error-Annotated Learner English Corpus: A tool for computer-assisted language learning. *NEALT Proceedings Series 2014, Vol. 22*, p. 87.
- Leech, G. (2015) Adding linguistic annotation. *Developing linguistic corpora: a guide to good practice*. Oxbow Books, Oxford, 2015, pp. 17-29.
- Vinogradova, O. (2016) The role and applications of expert error annotation in a corpus of English learner texts. *Computational Linguistics and Intellectual Technologies Issue 15 (22) International Conference "Dialogue 2016" Proceedings*, pp. 740-751.
- Vinogradova, O. (2016-2) Use of learner corpus in General English and Academic English courses at the Higher School of Economics *Conference Proceedings. The Future of Education, edition 6: libreriauniversitaria, 2016*, pp. 310-314.