Draft of the corpus presentation at CL2017 conference

Olga Vinogradova, Ekaterina Gerasimenko

12.06.2017

REALEC (<http://realec.org>) – Russian Error-Annotated Learner Corpus – consists of texts written in English by Russian university students studying English at an intermediate – upper-intermediate level. The corpus provides data for erratological research, for essay comparison and for various instruments. We have implemented RETM - REALEC English Test-Maker, which generates questions for tests of different types. Tags to be used as basis for questions should be selected carefully. Moreover, the overall test process should be organized neatly and in a way that is understandable for students.

Test creation consists of several steps:

1) preprocess initial files (paying attention to details concerning annotation overlaps), store texts with annotations in temporary files;

2) correct all other mistakes that are not to be used further;

3) extract the mistake and its correction;

4) edit the sentence with mistake to create a question of a particular type;

5) create an XML file with questions and import it to Moodle;

6) for teacher: edit questions manually.

Moodle (<https://moodle.org/>), which is used for testing students, has a straightforward interface, which makes editing easily. Editing process usually involves correcting mistakes that were not originally annotated and adding more variants of the answer. Moodle also provides several statistical measures that help to evaluate student’s performance and overall progress.