Russian Error-Annotated Learner English Corpus: a Tool for Computer-Assisted Language Learning

E. Kuzmenko, A. Kutuzov
National Research University Higher School of Economics
REALEC
(Russian Error-Annotated Learner English Corpus)

- http://realec.org
- Composed of English essays written by native Russian speakers.
- Errors encountered in essays are annotated.
- Annotation deploys 4 tiers of information about a particular error.
- Supplied with morphological annotation
Distribution of error types in REALEC
Novelty and timeliness

- Few learner corpora with Russian speakers’ data.
- None of such corpora available online.
- The necessity of tuning language teaching to the needs of learners

- REALEC’s aim:
  - to integrate data-driven methods into language teaching;
  - to create opportunity to study Russian-English interlanguage.
Applications and implications

Educational:
- Knowledge about problematic grammar patterns among Russian learners.
- Corpus-based grammar and grammar exercises.
- Automatic assessment of students’ works.

Theoretical:
- Comparing Russian-English interlanguage to interlanguages with other L1 backgrounds.
- Attempt to define the notion of error in learner’s language.
Thank you for your attention!

Please visit our poster.