Russian Learner Corpus: Towards Error-Cause Annotation for L2 Russian

D. Kosakin[1], **S. Obiedkov**[2], **E. Rakhilina**[3], **I. Smirnov**[3], **A. Vyrenkova**[3], **E. Zalivina**[3]
[1]Faculty of Computer Science, HSE University, Moscow, Russia
[2]Faculty of Computer Science / cfaed / ScaDS.AI, TU Dresden, Germany
[3]School of Linguistics, HSE University, Moscow, Russia
daniil.kosakin@gmail.com, sergei.obiedkov@tu-dresden.de, rakhilina@gmail.com,
smirnof.van@gmail.com, avyrenkova@hse.ru, zalivina01@mail.ru

Abstract

Russian Learner Corpus (RLC) is a large collection of learner texts in Russian written by native speakers of over forty languages. Learner errors in part of the corpus are manually corrected and annotated. Diverging from conventional error classifications, which typically focus on isolated lexical and grammatical features, the RLC error classification intends to highlight learners' strategies employed in the process of text production, such as derivational patterns and syntactic relations (including agreement and government). In this paper, we present two open datasets derived from RLC: a manually annotated full-text dataset and a dataset with crowdsourced corrections for individual sentences. In addition, we introduce an automatic error annotation tool that, given an original sentence and its correction, locates and labels errors according to a simplified version of the RLC error-type system. We evaluate the performance of the tool on manually annotated data from RLC.

Keywords: L2, learner corpus, grammatical error correction, error classification, automatic error annotation, crowdsourced corrections

1. Introduction

In recent decades, the role of learner corpora in applied linguistics has evolved increasingly. They are widely used to support the teaching process and design teaching materials, such as frequency-based dictionaries, exercises, etc. However, learner corpora provide further benefits beyond the acquisition and teaching of second languages. They offer valuable statistics about errors made by non-native speakers, which makes them useful for NLP and machine learning. Comprehensive markup of a learner corpus includes metadata about learners (their dominant language, age, gender, language acquisition conditions, language proficiency level, etc.), as well as a classification of errors. Using representative corpora and reliable markup ensures the validity of corpus data and the high quality of models trained on them. For this reason, large annotated datasets of learners' texts are highly valued in the field of grammatical error correction (GEC).

GEC tasks for L2 writing are currently being solved for many languages. At the same time, representative learner data are mainly collected for English as a foreign language, and there is a significant shortage of such for other languages. Error markup is especially problematic, since, at present, it is mostly done manually, which is time-consuming and labor-intensive. Of all currently known learner corpora, more than half are resources dedicated to English as a foreign language (Dahlmeier et al., 2013; Tajiri et al., 2012; Granger, 1998). There are learner corpora available for German as a foreign language (Boyd et al., 2014), Czech (Rosen, 2016), Japanese (Mizumoto et al., 2011), etc., but most of them are relatively small. The same applies to the Russian language.

Another substantial problem for learner corpora is relatively poor quality of error markup. The errors of L2 students vary significantly from those made by native speakers, both quantitatively and qualitatively. Such errors stem mainly from significant limitations in L2 input and a heavier processing load on a foreign language learner when they produce linguistic structures in their target language. This causes multiple errors that can be attributed to the patterns of second language acquisition and use by L2 learners, and annotated learner corpora offer a deeper insight into these patterns (Kisselev, 2021). Annotation schemes of datasets currently available for the Russian language are mainly based on lexical and grammatical features of individual words (Rozovskaya, 2021; Trinh and Rozovskaya, 2021; Katinskaia et al., 2022). We think that an approach with a stronger focus on causes of errors can be worthwhile.

In this paper, we pursue two goals. First, we present two open datasets related to GEC tasks that are derived from a large learner corpus collected over the last ten years by linguists and tutors of Russian as a second language from around the world. The corpus consists of free production samples produced by L2 students of Russian dominant in more than 40 languages, and it uses markup by error types that is aimed

LREC-COLING 2024, pages 14240–14258
20-25 May, 2024.

at highlighting grammatical and lexical issues of greater concern for them. The second goal of the paper is to introduce a tool for automatic error extraction and classification expected to streamline the process of error annotation for the permanently growing dataset. The tool is similar to the well-known ERRANT developed for English (Felice et al., 2016; Bryant et al., 2017).

The paper is structured as follows. In Section 2, we give a brief overview of related work on L2 datasets and error annotation for Russian. In Section 3, we describe the Russian Learner Corpus (RLC) together with its error-type system and introduce two open datasets based on it: a manually annotated full-text dataset and a dataset with crowdsourced corrections for individual sentences. Each dataset contains over 30,000 sentences. In Section 4, we present an error-annotation tool for the RLC error-type system and evaluate it on a small subset of RLC.

2. Related Work

Several datasets of Russian as a foreign language have become available recently. The most cited dataset used for GEC purposes is RULEC-GEC (Rozovskaya and Roth, 2019; Rozovskaya, 2021). This dataset includes texts written by students (heritage speakers and L2 learners) who studied Russian at the University of Oregon and whose proficiency level is intermediate or higher. RULEC-GEC is manually annotated for errors and divided into training, development and test subsets. It is formatted in congruence with English GEC (Ng et al., 2014) and contains more than 206,000 tokens (12,480 sentences).

Later, a larger RU-Lang8 dataset including 633,124 tokens (51,575 sentences) was proposed (Trinh and Rozovskaya, 2021). Though the corpus is much larger than RULEC-GEC, the reliably corrected and annotated part is smaller, amounting to 54,000 tokens (4,412 sentences) .

Both RULEC-GEC and RU-Lang8 are datasets consisting of sentences that learners produced entirely on their own (so called free production). The ReLCo dataset (Katinskaia and Yangarber, 2021; Katinskaia et al., 2022) offers a different kind of data. It was collected automatically with the help of Revita L2 learning platform, where learners of Russian had to complete exercises of various types (filling in the missing word, multiple choice). The size of the dataset is 375,453 tokens (22,370 sentences).

To facilitate the tagging procedure for large L2 datasets, Bryant et al. (2017) introduced a rule-based annotation toolkit ERRANT that, taking as input an original and corrected sentences, extracts edits and determines their types. The toolkit was designed for the English language; however, its adaptations for other languages soon followed (Boyd, 2018; Belkebir and Habash, 2021; Uz and Eryiğit, 2023; Yoon et al., 2023).

For Russian, two adaptations of ERRANT have been designed in conjunction with the two datasets described above, RULEC-GEC (Rozovskaya, 2022) and ReLCo (Katinskaia et al., 2022). The latter is available on GitHub under the name RuERRANT.[1] For both tools, evaluation results prove to be very good; however, both datasets have strong specificity. RULEC-GEC comprises sentences from students of relatively high proficiency level and of the same dominant language. In ReLCo, the texts are not produced entirely by L2 speakers of Russian; students only insert separate words or constructions in an offered context.

Thus, although much has been done for Russian as a foreign language in the area of grammatical error correction, there is still a lack of a large corpus that would contain reliably annotated free production data coming from students dominant in different languages.

In this paper, we present two large datasets for L2 Russian that include data coming from different types of speakers (heritage speakers and L2 learners), with different levels of proficiency and dominant in different languages. One of the datasets, RLC-GEC, is fully annotated for errors, while the other, RLC-Crowd, contains corrections obtained via a fairly large crowdsourcing experiment. In addition, we introduce a new rule-based ERRANT-like tool, RLC-ERRANT, enabling automated error tagging with the error-type system utilised in RLC-GEC.

3. Russian Learner Corpus

We base our work on the Russian Learner Corpus (RLC) (Rakhilina et al., 2016). This corpus consists of texts mainly produced by college and university students of the Russian language from different countries. Some of these texts were collected by tutors of Russian as a second language outside Russia; others came from the Russian Language Centre or the International Prep Year Centre at HSE University. The corpus currently comprises above 2,000,000 tokens (193,189 sentences), which include the production of both L2 learners of Russian and heritage speakers of Russian (HL), i.e., bilinguals who have a limited command of Russian as their mother tongue and are dominant in a different language. The number of learners' dominant languages registered in RLC is currently 48. The RLC team

[1] https://github.com/Relco/relco.github.io

has primarily relied on personal and institutional contacts to collect texts from data contributors at universities in various countries where Russian is taught as a foreign language, which shaped the dominant language sample. RLC includes RULEC, a large longitudinal subcorpus of L2 academic writing (Alsufieva Yatsenko et al., 2012), which also provides L2 material for the RULEC-GEC dataset mentioned above.[2] For the full list of partners who provided texts to the corpus see the RLC website.[3]

All respondents signed a special consent form, and their names are anonymized in the corpus. RLC contains author- and text-specific metatextual markup. Almost half of the texts in the corpus are manually annotated for error types and provided with corrections. Up until recently, RLC only enabled the user to query data. We now release its part as a standalone dataset suitable for automatic processing and training machine learning models.

3.1. RLC Error Types

Designing a tagset for a learner corpus is a complicated endeavor: there are often multiple hypotheses about the target structures behind errors, and the choice of the appropriate annotation is not always obvious. This requires the annotation scheme to be easily operational, unambiguous, and interpretable. One way to approach these requirements is to base the annotation scheme on misused grammatical markers of the language. This approach is effective, and the strong side of it is that the grammatical system provides a robust systematic ground for error classification. However, it is more descriptive than explanatory and offers little information about the causes of errors. Thus, errors in inflections can be accounted for by different reasons, such as:

- violations in syntactic government, where the case form depends on the verb used: *читать книга/chitat' kniga* "read a book (Nom)" → *читать книгу/chitat' knigu* "read a book (Acc)";
- violations in syntactic agreement, where the case form depends on the noun used: *большое зеркалу/bol'shoe zerkalu* "big (Nom) mirror (Dat)" → *большому зеркалу/bol'shomu zerkalu* "big (Dat) mirror (Dat)";
- the failure to attach proper inflections to given stems: *свинцем/svincem* "lead (Instr, invalid inflection)" → *свинцом/svincom* "lead (Instr)".

To our knowledge, among many papers that discuss annotation for learner corpora, a rare work aimed at developing error-cause annotation is (Kotani and Yoshimi, 2017), which addresses violations in using English articles by L2 learners of English. The classification developed for RLC is aimed at showing the relations and processes that may present difficulties for speakers of Russian as a non-dominant language. Thus, it takes into account errors in derivational morphology, syntactic relations (government, case/gender/person agreement), and constructional violations, which indicate directly the problems that learners experience when they produce texts in Russian (using derivational patterns, producing grammatical forms in accordance to syntactic relations in a sentence, etc.).

The classification takes into account existing work on developmental errors in L1 Russian, speech errors made by monolingual adult speakers of Russian, adult speakers of Russian as a heritage language (Polinsky, 2010), annotation schemes designed for other languages as L2 (Reznicek et al., 2013; Rosen et al., 2014), and feedback received from teachers of Russian as a foreign language (Rakhilina et al., 2016).

The resulting tagset comprises primary 35 tags for spelling, morphology, syntax, and lexis constructions. Edits may receive multiple tags if the error can be attributed to different classes. In addition to the primary tags, there are three secondary tags that cannot stand on their own and should be combined with a primary tag. These include tags for transfer, e.g, **`Lex+Transfer`** for lexical transfer, and extra or missing elements, e.g., **`Ref+Miss`** for omitting a referential marker (pronoun). Punctuation errors are currently not annotated.

An obvious drawback of using this tagset is low inter-annotator agreement (Rakhilina et al., 2016). This issue mainly arises because errors are often hard to classify or may be interpreted differently. At the same time, low Cohen's Kappa score is not uncommon in evaluating manual tagging results (Bryant et al., 2017).

For the full list of tags, their description and corresponding examples, see Appendix A. Below, we highlight certain aspects of the RLC error classification by contrasting it with tagsets utilized in RULEC-GEC and ReLCo.

In these datasets, error annotations refer to the token's part of speech and the grammatical category affected by the error (for example, noun case). In RLC, part-of-speech tagging is automatically applied using the Mystem tool[4] immediately after the text is uploaded to the

[2] Sentences from RULEC have been annotated for RLC and RULEC-GEC independently.

[3] https://www.web-corpora.net/RLC/

[4] https://yandex.ru/dev/mystem/

corpus; however, RLC error tags do not explicitly refer to parts of speech. For example, the following error will be classified as **`AgrGender`** (violation of gender agreement) because the verb does not agree with the subject noun:

фараон забыла/faraon zabyla
"pharaoh (Masc) forgot (Fem)"

→ *фараон забыл/faraon zabyl*
"pharaoh (Masc) forgot (Masc)".

Adjectives should also agree in gender with their controller nouns. Therefore, the same tag **`AgrGender`** will be assigned to the following correction:

гордой критиком/gordoj kritikom
"a proud (Fem) critic (Masc)"

→ *гордым критиком/gordym kritikom*
"a proud (Masc) critic (Masc)".

In RULEC-GEC and ReLCo, these errors would be assigned two different labels corresponding to verb gender and adjective gender respectively. In RLC, they receive the same error tag, since they are instances of the same phenomena. Still, by combining RLC error tags with POS tags, differentiation between subject-verb and noun-adjective agreement is possible.

As mentioned in Section 2, ReLCo comes equipped with an annotation tool, RuERRANT. We used it to compare the tag systems of RLC and ReLCo on a small manually tagged subset of RLC introduced in Section 3.2.1 as RLC-Test. The results are presented as a confusion matrix in Figure 1 with ReLCo tags plotted against the horizontal axis and RLC tags plotted against the vertical axis. Here, we show only tag pairs occurring more than four times. For the complete confusion matrix, see Appendix B.

One aspect of the RLC classification seen from the matrix is worth noting. While RuERRANT usually classifies out-of-vocabulary words as spelling errors, Russian language learners may end up with non-existent words for various reasons, for example, when they fail to follow a proper inflectional pattern (which is labeled as **`Infl`** in RLC) or a proper derivational pattern (labeled as **`Morph`**):

долгожительность/dolgozhitel'nost'
"longevity"
(formed with an invalid abstract noun suffix)

→ *долгожительство/dolgozhitel'stvo*
"longevity".

In comparison to ReLCo, RLC offers a tagset that is better suitable for distinguishing between orthographic and morphological errors.

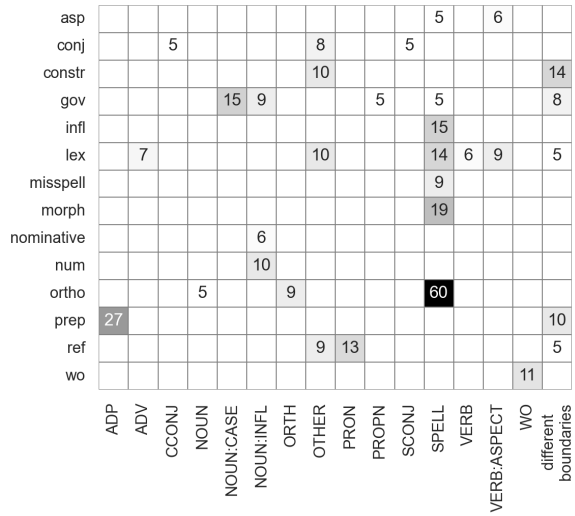

Figure 1: Partial confusion matrix for RuERRANT applied to the RLC-Test dataset from Section 3.2.1. Rows correspond to RLC tags, and columns correspond to tags assigned by RuERRANT; cases where RuERRANT delimits errors differently than it is done in the dataset are labeled as "different boundaries".

3.2. Datasets Derived from RLC

3.2.1. RLC-GEC: Annotated Subset of RLC

We present a new dataset, RLC-GEC, which is a partial dump of RLC consisting of corrected and annotated texts written by Russian learners. It is split into three files. The first file contains meta-information about the texts, such as the dominant language of the authors, their language background (L2 or heritage speaker) and language level, as well as the size of the text in words and sentences. The second file includes original sentences and their corrections. The third file consists of individual error annotations, where, for each error, a correction and type tags according to the error type system presented in Section 3.1 are indicated. There are currently 2004 texts comprising 31519 sentences with 41410 error annotations. The dominant languages are listed in Table 1, while Table 2 shows the most frequently-occurring error tags.

Since we plan to regularly update the dataset, the numbers above are subject to change.

In addition to the full-text dataset, we release a smaller dataset, RLC-Test, that contains 204 individual sentences and annotated corrections for 519 errors therein. These sentences are different from those in the bigger dataset. We use this second dataset to evaluate performance of our error-annotation tool presented in Section 4.1.

Language	Texts	Language	Texts
English	760	Dutch	9
Chinese	304	Norvegian	9
French	214	Bulgarian	3
Kazakh	157	Farsi	3
Spanish	123	Portuguese	3
Turkmen	98	Estonian	2
Italian	72	Mongolian	2
Serbian	65	Tajik	2
German	42	Urdu	2
Slovenian	34	Abkhazian	1
Arabic	21	Bengal	1
Macedonian	19	Greek	1
Turkish	19	Kurdish	1
Korean	15	Vietnamese	1

Table 1: Dominant languages in the RLC dataset. For 21 texts, the dominant language is not known.

Tag	%	Tag	%
Lex	19.7	Conj	3.4
Ortho	15.8	Extra	3.3
Syntax	13.8	AgrCase	3.1
Gov	8.3	Morph	2.8
Constr	6.9	AgrNum	2.8
Miss	5.7	WO	2.7
Prep	5.3	AgrGender	2.6
Ref	4.6	Num	2.5
Asp	3.6	Infl	2.4

Table 2: The most frequently occurring error tags in the RLC dataset.

Both datasets are available on GitHub[5].

3.2.2. RLC-Crowd: Crowdsourced Corrections

A large part of RLC consists of sentences without corrections. As an experiment, we used the crowdsourcing platform Toloka[6] to obtain corrections for 34,150 sentences. Each sentence has been corrected by at least five users, and 4,866 users have been involved in total. The resulting dataset, RLC-Crowd[7], contains 213,683 corrected sentences, together with the corresponding original sentence and the ID of the user who corrected the sentence (all IDs are local to the dataset and do not coincide with IDs assigned to users by Toloka).

The quality of corrections varies greatly. To obtain high-quality corrections, it is necessary to develop techniques for aggregating corrections from several users. It also seems that sentences should be offered for correction to a larger number of users. At the same time, it may be possible that machine learning models can benefit from crowdsourced data as is: our preliminary experiments (not covered in this paper) with a Transformer-based error-correction model suggest that augmenting professionally annotated data with crowdsourced data during training may result in better quality of error correction. Of course, further research is needed to be able to confirm this hypothesis.

The dataset includes 33 sentences slightly modified to reduce the number of ways errors therein can be corrected. These sentences were injected into tasks offered to all users, and users who repeatedly failed to produce expected corrections for such sentences were disqualified from performing further tasks. As a result, most of these sentences have been corrected by over a thousand users. We release a separate file with only these sentences and their corrections, indicating, for each correction, the number of users who proposed it. Although limited in size, this data still provides very interesting material that can help put forward initial hypotheses regarding suitability of crowdsourcing technologies for correcting errors of various types.

4. Automatic Error Annotation

4.1. RLC-ERRANT

We developed a tool, RLC-ERRANT, for automatic error annotation following the RLC error classification outlined in Section 3.1.[8] Similar to RuERRANT from (Katinskaia et al., 2022), RLC-ERRANT is an adaptation of ERRANT, error-annotation software for English (Bryant et al., 2017). Since, as discussed above, the RLC error type system substantially differs from the one in (Katinskaia et al., 2022), the two corresponding tools are necessarily quite different.

ERRANT-based tools receive as input a sentence and its correction. The annotation process consists of two steps: error extraction and error classification. Error extraction requires solving an alignment problem to identify the boundaries of every corrected fragment. This is, however, not sufficient, since a single error may cover two or more edits. Consider the example in Table 3, which shows an optimal alignment between a sentence and its correction. The original sentence contains two errors, one of which is an extra space. The alignment algorithm based

[5] https://github.com/Russian-Learner-Corpus/rlc-annotated

[6] https://toloka.ai

[7] https://github.com/Russian-Learner-Corpus/rlc-crowd

[8] https://github.com/Russian-Learner-Corpus/annotator

on Damerau–Levenshtein distance minimization handles this case by suggesting a deletion and a substitution, whereas we would rather treat it as a single substitution. For this reason, some of aligned edits should be merged.

ERRANT provides a set of merging rules, which we augmented with rules specific to Russian. For instance, if adjacent words in the original sentence have the same number and case different from those of the corresponding words in the correct sentence, we consider this to be a single error. In the example from Table 4, three adjacent words are in a wrong case. However, there is only one error in the sentence, since two of the words, a determinant and an adjective, simply agree in case with the third, a noun.

For the current version of our tool, we somewhat simplified the error tagging system used in RLC. Similarly to how this is implemented in ERRANT for English, the classification in RLC-ERRANT is rule-based. The classification algorithm receives as input original and corrected token sequences of a single edit together with the information on their parts of speech and POS-specific morphological properties. This information is obtained by parsing the entire original and correct sentences using libraries from the Natasha toolset.[9] The result is not always accurate, especially, when the original sentence contains non-existing words or its grammatical structure is totally skewed, which if often the case for RLC sentences. For this reason, in some rules, we resorted to morphological information provided by pymorphy2[10], a popular morphological analyzer for Russian and Ukrainian. Although it processes individual words and cannot take the context into account when determining POS or morphological properties of a word, we empirically found that it may still be safer to rely on pymorphy2 than on Natasha in certain situations.

The algorithm features a rule for every error type. When classifying an edit, the algorithm tries to match it against the rules; the class is assigned by the first rule that is successfully matched. The rules are considered in the following order:

> **WO** (word order), **CS** (code switching), **Brev** (short/long forms of adjectives), **Tense** (verbal tense), **Passive** (passive constructions), **Num** (noun number), **Gender** (noun gender), **Nominative/Gov/AgrCase** (syntactic subject in oblique case/syntactic government/case agreement), **AgrNum** (number agreement), **AgrPers** (person agreement), **AgrGender** (gender agreement), **Refl** (reflexive forms), **Asp** (verbal aspect), **Impers** (impersonal constructions), **Com** (comparative constructions), **Mode** (conditional constructions), **Hyphen+Ins** (extra hyphen), **Hyphen+Del** (missing hyphen), **Space+Ins** (invalid separate spelling), **Space+Del** (invalid merged spelling), **Conj** (conjunctions), **Ref** (referential markers), **Prep** (prepositions), **Graph** (alphabet mixing), **Infl** (invalid inflection), **Lex** (wrong lexical choice), **Constr** (constructional error), **Ortho** (orthographic error), **Morph** (derivational error), **Ortho**

If none of the rules is matched, the edit is classified as **Misspell** (complex orthographic error).

The sequence above contains two occurrences of **Ortho**, which correspond to two rules. The first rule checks for common spelling errors, namely, for confusions between ‘е’ and ‘э’ or ‘и’ and ‘ы’. The second rule, applied only after the error has been deemed non-morphological, decides between **Ortho** and **Misspell** by checking if the normalized indel (insertions/deletions) similarity between the original and correct words is below a certain threshold, currently set at 0.8.

The rules are available as part of the software. Below, we give only a few (somewhat simplified) examples. Every rule is applied to an extracted pair of original and correct token sequences.

Nominative/Gov/AgrCase: The sequences must be of the same length. For each sequence, determine all case/number pairs simultaneously applicable to all its tokens. The sets of pairs for the two sequences must be non-empty and disjoint, while the sets of numbers should intersect (otherwise, the error may be in the number, rather than in the case. The corresponding tokens in the two sequences must have matching lemmas. If none of the tokens in the sequences is a noun or a pronoun, classify as **AgrCase**. If not, classify as **Nominative** or **Gov** depending on whether the correct sequence contains a token in the nominative case.

Conj: At least one of the sequences contains a conjunction.

Ref: All tokens in at least one of the two sequences are pronouns or determinants.

Prep: Both sequences consist of prepositions.

Lex: Both sequences consist of a single token each; the original word exists, and its normal form is different from that of the correct word. The words are not negation particles *не/ne*

[9] https://github.com/natasha

[10] https://github.com/pymorphy2/pymorphy2

Можно	увлечься	чем-то	более	**полезней**	и	**при**	**том**	отдохнуть
mozhno	uvlech'sya	chem-to	bolee	poleznei	i	pri	tom	otdokhnut'
Можно	увлечься	чем-то	более	**полезным**	и		**притом**	отдохнуть
mozhno	uvlech'sya	chem-to	bolee	poleznym	i		pritom	otdokhnut'
You can	get involved	in something	more	useful	and		still	get some rest

Table 3: An alignment between a sentence and its correction: extra space.

Ремонт	делает	**этим**	**великолепным**	**зданием**	идеальным	для	жилья
remont	delaet	etim	velikolepnym	zdaniem	ideal'nym	dlya	zhil'ya
Ремонт	делает	**это**	**великолепное**	**здание**	идеальным	для	жилья
remont	delaet	eto	velikolepnoe	zdanie	ideal'nym	dlya	zhil'ya
Renovations	make	this	gorgeous	building	perfect	for	living

Table 4: An alignment between a sentence and its correction: error in case government.

and *ни/ni*, and they consist of different letters (errors related to double letters or the order of letters are usually classified as **Ortho**).

As can be seen from these examples, some of the rules are fairly involved, while others are quite simple. The latter often deviate from the RLC classification principles. For example, we would rather classify *за слова/za slova* "for words" → *в словах/v slovakh* "in words" as **Prep**, since the different case of the noun (the second word) is determined by the choice of the preposition (the first word) and thus does not constitute an independent error. The rule for **Prep** has to be updated to allow for such cases.

4.2. Experimental Evaluation

We tested our current implementation on the RLC-Test dataset described in Section 3.2.1. Figure 2 shows a confusion matrix for this experiment. In some cases, RLC-ERRANT may fail to identify error boundaries as expected; these are labeled as "different boundaries" in the matrix. In our experiment, this happens for fewer than 12% cases, many of which are due to the inability of Natasha and/or pymorphy2 to correctly recognize certain morphological features. For example, RLC-ERRANT can merge spelling errors in two adjacent words into one edit: *маленкий былий/malenkii bylii* → *маленький белый/malen'kii belyi* "small white". This happens because of the merging rule described in Section 4.1, which is needed to correctly handle the example in Table 4. In this case, the rule fires only because the morphological analyzer reports the nominative case for the correct words and (wrongly) the accusative case for the original words.

The overall classification accuracy is 0.58. Table 5 shows precision and recall for the most frequently occurring RLC tags supported by the tool. As can be seen, it is fairly good at dealing with most such tags. The precision for **Asp** is lower than it could be due to the difficulties the morphological analyzer has in determining the aspect of words with spelling or morphological errors. The recall can be improved by using a verb dictionary of aspectual pairs; the current version of RLC-ERRANT is able to recognize only pairs with similar stems and thus misses morphologically unrelated pairs such as *взял/vzyal* "has taken" → *брал/bral* "used to take", which are classified as **Lex**, reducing the precision for the latter.

Tag	Precision	Recall
Lex	0.70	0.77
Ortho	0.73	0.10
Gov	0.91	0.75
Constr	0.62	0.38
Prep	0.97	0.78
Ref	0.76	0.81
Asp	0.71	0.71
Conj	0.77	0.87

Table 5: Precision and recall for the most common tags supported by RLC-ERRANT.

Constr errors are diverse in their structure and thus may be hard to catch. Often, the tool splits what annotators would prefer to see as a single **Constr** error into several errors. For example, *как это/kak eto* "as this" → *такое/takoe* "such" is parsed as a combination of a **Conj** and a **Ref** errors, which may also be considered a valid interpretation.

The biggest problem seems to be the low recall for spelling errors, **Ortho** and **Misspell**. The main reason for this is that the rule for spelling errors is applied only if all the other rules fail. Therefore, for example, spelling errors in endings, e.g., *интервию/interviyu* → *интер-*

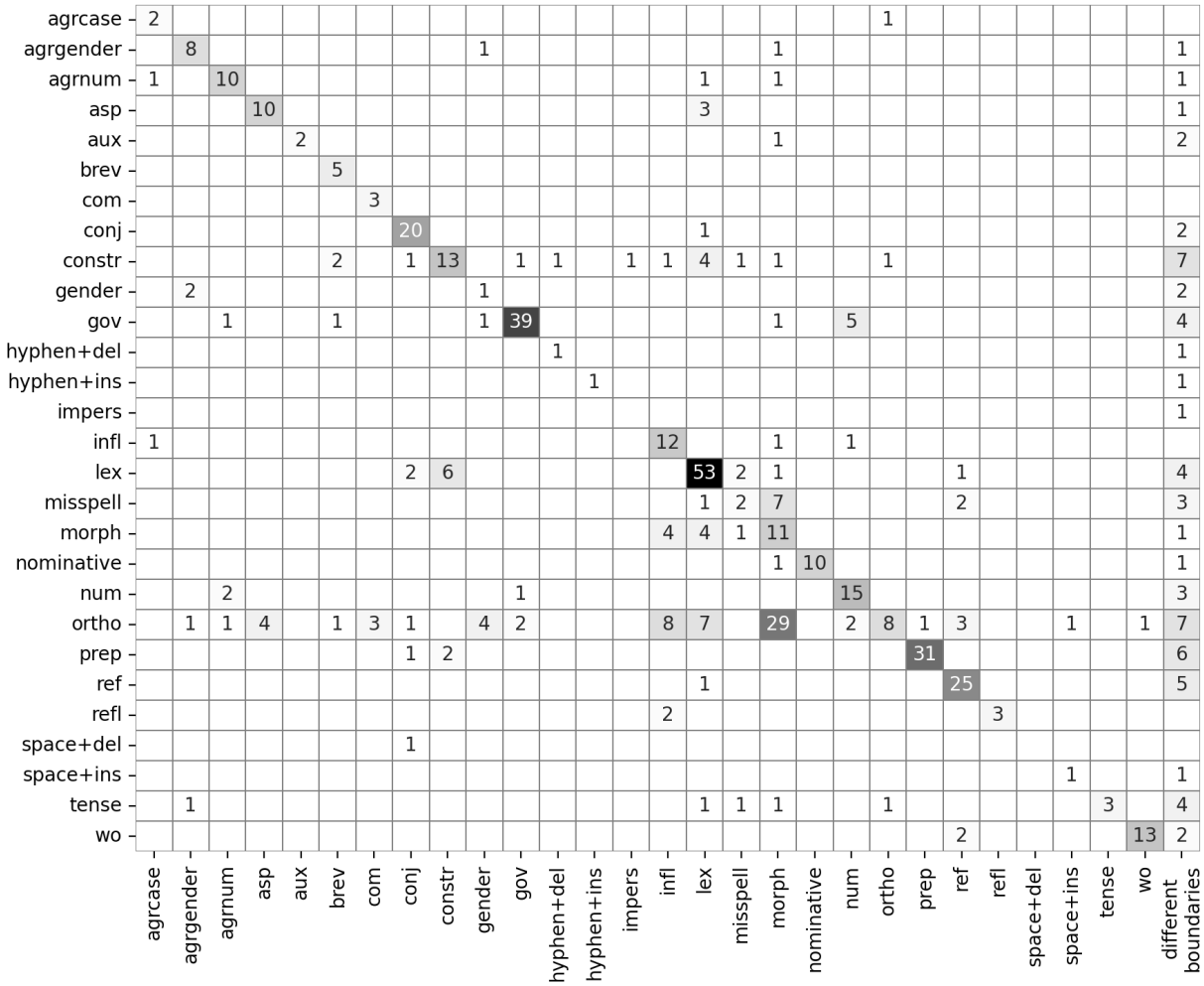

Figure 2: Confusion matrix for RLC-ERRANT applied to the test dataset from Section 3.2.1. Rows correspond to true tags, and columns correspond to predicted tags.

вью/interv'yu "interview", are often classified as **Infl**. The major issue here is that many spelling errors, e.g., *популярным/popul'yarnym* → *популярным/popul'yarnym* "popular", are classified as morphological. One possible solution is to resort to machine learning, build a classifier for distinguishing between spelling and morphological errors, and use it in the tool instead of the respective rules.

To get some first evaluation of the idea, we fine-tuned DeepPavlov (Kuratov and Arkhipov, 2019), a pretrained language model for Russian, on over 16000 spelling and morphological errors from RLC. For training we used extended categories, which included related tags in addition to **Ortho**, **Misspel**, and **Morph**. Using this model, we were able to slightly increase the overall accuracy of RLC-ERRANT to 0.59, however, at the expense of decreasing the recall for **Morph**. We plan to further explore this approach in the future.

5. Conclusion

We release two datasets derived from the Russian Learner Corpus containing sentences written by Russian learners and heritage speakers. One dataset, RLC-GEC, consists of full texts and their corrections annotated by linguists. It includes a smaller dataset, RLC-Test, with individual sentences rather than full texts. Among other things, these datasets can be used for, respectively, training and testing machine-learning models for GEC in non-native Russian writing.

The second dataset, RLC-Crowd, contains corrections obtained via a crowdsourcing platform. Each sentence has been corrected by several users, and most of the users involved have corrected multiple sentences. We believe that this dataset can help study the potential of crowdsourcing technologies in obtaining corrections for large corpora of non-native writing, as well as help identify strategies employed by users of crowdsoursing platforms when

correcting errors. Although individual corrections obtained via crowdsourcing are often far from perfect, we hypothesize, based on our preliminary experiments, that machine-learning models can still benefit from such data when used together with better-quality data in training. We plan to verify this hypothesis in our future studies.

RLC uses an error-type system that is quite different from the other systems developed for Russian (Rozovskaya, 2022; Katinskaia et al., 2022). It is not based on POS properties of tokens but rather indicates linguistic relations and patterns that cause problems for L2 learners when they produce coherent texts in Russian (syntactic relations, lexical choice, derivational patterns, etc.).

We developed a first version of an error-annotation tool, RLC-ERRANT, for the RLC error-type system. As a similar tool for English (Bryant et al., 2017), our tool is rule-based and performs annotation in two separate steps: error extraction and error classification. We present an experimental evaluation of its performance for various error types. While its output must still be verified by an expert, our preliminary experiments indicate that using the tool to obtain suggestions for error boundaries and tags significantly reduces annotation time and improves accuracy. RLC-ERRANT is already used by annotators of the Russian Learner Corpus and is currently being integrated into the RLC annotation platform. We hope to improve its performance by melding the extraction and classification steps and using machine learning for at least some error types.

6. Acknowledgements

This work is partly supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) in project 389792660 (TRR 248, Center for Perspicuous Systems), by the Bundesministerium für Bildung und Forschung (BMBF, Federal Ministry of Education and Research) in the Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI), and by BMBF and DAAD (German Academic Exchange Service) in project 57616814 (SECAI, School of Embedded Composite AI). The publication is supported by the grant for research centers in the field of AI provided by the Russian Federation (ACRF) in accordance with the agreement on the provision of subsidies (identifier of the agreement 000000D730321P5Q0002) and the agreement with HSE University No. 70-2021-001139.

We would like to express our deep gratitude to Nikita Remnev, Daniil Fedorov, Maria Zambrzhitskaia, Polina Egorova, and the undergraduate students of Fundamental and Computational Linguistics at HSE University who participated in the data review for this paper.

7. Bibliographical References

Anna A. Alsufieva Yatsenko, Olesya V. Kisselev, and Sandra G. Freels. 2012. Results 2012: Using flagship data to develop a russian learner corpus of academic writing. *Russian Language Journal*, 62:79–105.

Riadh Belkebir and Nizar Habash. 2021. Automatic error type annotation for Arabic. In *Proceedings of the 25th Conference on Computational Natural Language Learning*, pages 596–606, Online. Association for Computational Linguistics.

Adriane Boyd. 2018. Using Wikipedia edits in low resource grammatical error correction. In *Proceedings of the 2018 EMNLP Workshop W-NUT: The 4th Workshop on Noisy User-generated Text*, pages 79–84, Brussels, Belgium. Association for Computational Linguistics.

Adriane Boyd, Jirka Hana, Lionel Nicolas, Detmar Meurers, Katrin Wisniewski, Andrea Abel, Karin Schöne, Barbora Štindlová, and Chiara Vettori. 2014. The MERLIN corpus: Learner language and the CEFR. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 1281–1288, Reykjavik, Iceland. European Language Resources Association (ELRA).

Christopher Bryant, Mariano Felice, and Ted Briscoe. 2017. Automatic annotation and evaluation of error types for grammatical error correction. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 793–805, Vancouver, Canada. Association for Computational Linguistics.

Daniel Dahlmeier, Hwee Tou Ng, and Siew Mei Wu. 2013. Building a large annotated corpus of learner English: The NUS corpus of learner English. In *Proceedings of the Eighth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 22–31, Atlanta, Georgia. Association for Computational Linguistics.

Mariano Felice, Christopher Bryant, and Ted Briscoe. 2016. Automatic extraction of learner errors in ESL sentences using

linguistically enhanced alignments. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 825–835, Osaka, Japan. The COLING 2016 Organizing Committee.

Sylviane Granger. 1998. The computer learner corpus: A versatile new source of data for SLA research. In Sylviane Granger, editor, *Learner English on Computer*, pages 3–18. Addison Wesley Longman, London and New York.

Anisia Katinskaia, Maria Lebedeva, Jue Hou, and Roman Yangarber. 2022. Semi-automatically annotated learner corpus for Russian. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 832–839, Marseille, France. European Language Resources Association.

Anisia Katinskaia and Roman Yangarber. 2021. Assessing grammatical correctness in language learning. In *Proceedings of the 16th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 135–146, Online. Association for Computational Linguistics.

Olesya Kisselev. 2021. *Corpus-Based Methodologies in the Study of Heritage Languages*, Cambridge Handbooks in Language and Linguistics, page 520–544. Cambridge University Press.

Katsunori Kotani and Takehiko Yoshimi. 2017. Annotation of a learner corpus toward development of an error-cause presenting technique. In *Proceedings of The 2017 International Conference on Advanced Technologies Enhancing Education (ICAT2E 2017)*, pages 78–81. Atlantis Press.

Yuri Kuratov and Mikhail Y. Arkhipov. 2019. Adaptation of deep bidirectional multilingual transformers for russian language. *CoRR*, abs/1905.07213.

Tomoya Mizumoto, Mamoru Komachi, Masaaki Nagata, and Yuji Matsumoto. 2011. Mining revision log of language learning SNS for automated Japanese error correction of second language learners. In *Proceedings of 5th International Joint Conference on Natural Language Processing*, pages 147–155, Chiang Mai, Thailand. Asian Federation of Natural Language Processing.

Hwee Tou Ng, Siew Mei Wu, Ted Briscoe, Christian Hadiwinoto, Raymond Hendy Susanto, and Christopher Bryant. 2014. The CoNLL-2014 shared task on grammatical error correction. In *Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task*, pages 1–14, Baltimore, Maryland. Association for Computational Linguistics.

Maria Polinsky. 2010. Russkij jazyk pervogo i vtorogo pokolenija emigrantov, zhivuschix v SShA. *Slavica Helsingiensia*, 40:336–352.

Ekaterina Rakhilina, Anastasia Vyrenkova, Elmira Mustakimova, Alina Ladygina, and Ivan Smirnov. 2016. Building a learner corpus for Russian. In *Proceedings of the joint workshop on NLP for Computer Assisted Language Learning and NLP for Language Acquisition*, pages 66–75, Umeå, Sweden. LiU Electronic Press.

Marc Reznicek, Anke Lüdeling, and Hagen Hirschmann. 2013. Competing target hypotheses in the falko corpus. In Ana Díaz-Negrillo, Nicolas Ballier, and Paul Thompson, editors, *Automatic Treatment and Analysis of Learner Corpus Data*, pages 101–124.

Alexandr Rosen. 2016. Building and using corpora of non-native Czech. In *Proceedings of the 16th ITAT Conference Information Technologies - Applications and Theory, Tatranské Matliare, Slovakia, September 15-19, 2016*, volume 1649 of *CEUR Workshop Proceedings*, pages 80–87. CEUR-WS.org.

Alexandr Rosen, Jirka Hana, Barbora Štindlová, and Anna Feldman. 2014. Evaluating and automating the annotation of a learner corpus. *Language Resources and Evaluation*, 48(1):65–92.

Alla Rozovskaya. 2021. Spelling correction for Russian: A comparative study of datasets and methods. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 1206–1216, Held Online. INCOMA Ltd.

Alla Rozovskaya. 2022. Automatic classification of Russian learner errors. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 5637–5647, Marseille, France. European Language Resources Association.

Alla Rozovskaya and Dan Roth. 2019. Grammar error correction in morphologically rich languages: The case of Russian. *Transactions of the Association for Computational Linguistics*, 7:1–17.

Toshikazu Tajiri, Mamoru Komachi, and Yuji Matsumoto. 2012. Tense and aspect error correction for ESL learners using global context.

In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 198–202, Jeju Island, Korea. Association for Computational Linguistics.

Viet Anh Trinh and Alla Rozovskaya. 2021. New dataset and strong baselines for the grammatical error correction of Russian. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 4103–4111, Online. Association for Computational Linguistics.

Harun Uz and Gülşen Eryiğit. 2023. Towards automatic grammatical error type classification for Turkish. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop*, pages 134–142, Dubrovnik, Croatia. Association for Computational Linguistics.

Soyoung Yoon, Sungjoon Park, Gyuwan Kim, Junhee Cho, Kihyo Park, Gyu Tae Kim, Minjoon Seo, and Alice Oh. 2023. Towards standardizing Korean grammatical error correction: Datasets and annotation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6713–6742, Toronto, Canada. Association for Computational Linguistics.

Appendix A. RLC Error Classification

The RLC tagset includes 38 tags for spelling, morphology, syntax, lexis and constructions. A token can receive more than one tag if it can be attributed to more than one error type. Similarly, one tag can cover several tokens. The total of 38 tags include 35 primary tags, as well as three secondary tags that need to be paired up with the primary one. The secondary tags include tags for transfer, e.g., **`Lex+Transfer`** for lexical transfer, and extra or missing elements, e.g., **`Ref+Miss`** for omitting a referential marker (pronoun). Punctuation errors are currently not annotated.

Error Tag	Description and Examples
Graph	Mixing alphabets
	В тот момент, когда *м**m**ы *переходи**m** эту границу… V tot moment, kogda my perekhodim etu granitsu... В тот момент, когда **м**ы переходи**м** эту границу... V tot moment, kogda my perekhodim etu granitsu... *The moment we cross this border...*
Hyphen	Errors in hyphenated spelling
	Она уговаривает мужчину ***кудато*** собирается. Ona ugovarivayet muzhchinu kudato sobiratsya. Она уговаривает мужчину **куда-то** собираться. Ona ugovarivayet muzhchinu kuda-to sobirat'sya. *She persuades the man to get ready to go somewhere.*
Space	Extra or missing spaces
	***На конец**, может быть человек узнает как заработовать деньги за дела. Na konets, mozhet byt' chelovek uznayet kak zarabotovat' den'gi za dela. **Наконец**, может быть, человек узнает, как зарабатывать деньги за дела. Nakonets, mozhet byt' chelovek uznayet kak zarabatyvat' den'gi za dela. *Finally, maybe a person will learn how to make money for doing things.*
Ortho	Violation of standard Russian orthography, (except for hyphenation and spaces). This includes, in particular, errors in unstressed vowels and misuse of letters in the root of the word.
	И до сих пор *общ**я**емся. I do sikh por obshchyaemsya. *И* до сих пор общ**а**емся I do sikh por obshchayemsya. *And (we) still communicate.*

Error tag	Tag description
Misspell	Complex spelling errors stemming from the author's intuitive notion of how the word is spelled or pronounced. The author may have heard the word and may have a rough idea of what it means but has only distant memories about how it is spelt or even pronounced. This tag is applied to non-morphological errors. Он дал нам ***деяк***. On dal nam deyak. *Он дал нам **денег*** On dal nam deneg. *He gave us money.*
Morph	Derivational errors: a wrong suffix or prefix is added to a word stem resulting in a non-existent word. После ***осмотрения*** современного Санкт-Петербурга... Posle osmotreniya sovremennogo Sankt-Peterburga... После **осмотра** современного Санкт-Петербурга... Posle osmotra sovremennogo Sankt-Peterburga... *After visiting modern Saint-Petersburg...* Suffix -ени- (-enij-) is used to derive an abstract noun from the verb осмотреть (osmotret') 'inspect', although a null suffix should be used
Altern	Errors in stem alternation Я ***любю** моих друзей. Ja lyubyu moikh druzey. *Я **люблю** моих друзей.* Ja lyublyu moikh druzey. *I love my friends.* The 1.Sg of the verb *lyubit'* 'love' is formed with an alternated stem *lyubl-*.
Infl	Inflectional error: using an existing inflection results in a non-existent form. Такие условия легко способствуют отравлению ***свинцем*** жителей. Takiye usloviya legko sposobstvuyut otravleniyu svintsem zhiteley. Такие условия легко способствуют отравлению **свинцом** жителей. Takiye usloviya legko sposobstvuyut otravleniyu svintsom zhiteley. *Such conditions easily contribute to lead poisoning among residents.* The instrumental case ending *-ем* is unstressed; under stress, the ending *-ом* should be used.

Error tag	Tag description
Num	Errors in number as a nominal category (not number agreement) that result in non-existent words Существует множество других способов *__познаний__ мира. x Sushchestvuet mnozhestvo drugikh sposobov poznanijx-Pl mira. Существует множество других способов __познания__ мира. Sushchestvuet mnozhestvo drugikh sposobov poznaniya-Sg mira. *There are many other ways to learn about the world.* The word *poznanie* 'learning about' is not used in plural.
Gender	Errors in gender as a nominal category (not gender agreement) that result in non-existent words Автор даёт серьезн__ыю__ *__комментарию__. Avtor dayot seriyoznyyu (Fem, Acc) kommentariyu. Автор даёт серьезный __комментарий__ Avtor dajot serijoznyj kommentarij (Masc, Acc). *The author gives a serious commentary.* The inflection *-ю* for the accusative case is used with soft stems of feminine nouns.
Tense	Errors in tense forms. This tag does not apply to aspectual errors such as errors in using analytical forms of future tense. Также, она позвонила всем, кому *__может__ позвонить. Takzhe, ona pozvonila vsem, komu mozhet (Pres) pozvonit' Также, она позвонила всем, кому __могла__ позвонить. Takzhe, ona pozvonila vsem, komu mogla (Past) pozvonit'. *Also, she called everyone whom she could call.*
Asp	Errors in aspectual forms: misuse of aspectual forms including those xderived through suffixation or prefixation, analytical future tense forms, unidirectional vs. multidirectional motion verbs Помню, как мне сразу она *__нравилась__. Pomnyu, kak mne srazu ona nravilas' (Imp). Помню, как мне сразу она __понравилась__. Pomnyu, kak mne srazu ona ponravilas' (Perf). *(I) remember how I liked her straight away.*
Refl	Errors in reflexive forms: failure to use or misuse of verbs ending in reflexive *-ся* ...*__встречала__ с мою подрушку... ...vstrechala (non-reflexive) s moyu podruzhku... ...__встречалась__ с моей подружкой... ...vstrechalas' (reflexive) s moej podruzhkoj... *...met my friend...*

Error tag	Tag description
Brev	Errors in the use of full vs. short forms of adjectives
	Чай был такой *__вкусен__. Chai byl takoj vkusen (Brev). Чай был такой __вкусный__. Chai byl takoj vkusnuj (Full). *Tea was so delicious.*
Gov	Errors in syntactic government: misuse of noun case forms
	Люди думают, что легко найти хороших *__друзья__. Lyudi dumayut, chto legko najti khoroshikh druz'ya (Nom) Люди думают, что легко найти хороших __друзей__ Lyudi dumayut, chto legko najti khoroshikhx druzej (Acc) *People think it's easy to find good friends*
Ref	Misuse of referential markers including missing, extra, or badly chosen pronouns
	Жена Фараона считала *__ее__ жизнь чудесной Zhena Faraona schitala ee (invalid use of possessive pronoun) zhizn' chudesnoj. Жена Фараона считала __свою__ жизнь чудесной Zhena Faraona schitala svoyu zhizn' chudesnoj. *The pharaoh's wife considered her life wonderful.*
Aux	Errors in auxiliary or copula verb usage
	Когда мне грустно, друг всегда *__есть__ со мной. Kogda mne grustno, drug vsegda est' (no copula needed) so mnoj. Когда мне грустно, друг всегда со мной Kogda mne grustno, drug vsegda so mnoj. *When I'm sad, a friend is always with me.*
AgrNum	Errors in number agreement
	...взял все *__своё__ принадлежности. ...vzyal vse (Pl) svoyo (Sg) prinadlezhnosti (Pl). ...взял все __свои__ принадлежности. ...vzjal vse (Pl) svoi (Pl) prinadlezhnosti (Pl). *...(he) took all his possessions.* Possessive pronoun *свой* 'his' must agree with the controller noun *принадлежности* 'possessions' in number.

Error tag	Tag description
AgrCase	Errors in case agreement с ***этом** значением s etom (Loc) znacheniem (Instr) с **этим** значением s etim (Instr) znacheniem (Instr) *with this meaning* Demonstrative pronoun *этот* 'this' must agree with the controller noun *значение* 'meaning' in case.
AgrGender	Errors in gender agreeement в ***скошенном** траве v skoshennom (Adj, Masc) trave (Noun, Fem) в **скошенной** траве v skoshennoj (Adj, Fem) trave (Noun, Fem) *in the cut grass* Adjective *скошенный* 'cut' must agree with the controller noun *трава* 'grass' in gender.
AgrPers	Errors in person agreement Он эти люди очень мало ***знаю**. On (3.Sg) eti lyudi ochen' malo znayu (1.Sg). Он этих людей очень мало **знает**. On (3.Sg) etikh lyudej ochen' malo znaet (3.Sg). *He knows very little of these people.* Verb *знать* 'know' must agree with the controller pronoun *он* 'he' in person.
Passive	Errors in passive constructions 30-го апреля 2012, ***провели** дебаты. 30-go aprelya 2012, proveli (Active) debaty. 30 апреля 2012 г. **были проведены** дебаты 30 aprelya 2012 g. byli provedeny (Passive) debaty. *On April 30, 2012, a debate was held.*
Com	Errors in comparative constructions С дипломом, вы можете найти работу ***более легко**. S diplomom, vy mozhete najti rabotu bolee legko. С дипломом вы можете **легче** найти работу. S diplomom vy mozhete legche najti rabotu. *With a degree, it may be easier for you to find a job.*

Error tag	Tag description
Impers	Errors in impersonal constructions: using a personal construction instead for an impersonal one or vice versa *__Туристы будут__ интересно увидеть озеро Turisty budut (3.Pl, Active) interesno uvidet' ozero. __Туристам будет__ интересно увидеть озеро Turistam budet (Imp) interesno uvidet' ozero. *It will be interesting for tourists to see the lake.*
Mode	Errors in conditional constructions Это возможно, *__если__ участники согласились. Eto vozmozhno, esli uchastniki soglasilis'. Это было бы возможно, __если бы__ участники согласились. Eto bylo by vozmozhno, esli by uchastniki soglasilis'. *It would have been possible if the participants had agreed.* Particle *бы* is omitted in the conditional construction.
Gerund	Errors in gerundive constructions *__Поступив__ в колледж, начинается взрослая жизнь. Postupiv v kolledzh, nachinayetsya vzroslaya zhizn'. *Entering a college, adult life begins.* __Когда поступаешь__ в колледж, начинается взрослая жизнь. Kogda postupayesh' v kolledzh, nachinayetsya vzroslaya zhizn'. *When you go to college, adult life begins.*
WO	Errors in word order Хотя ситуация *__ясна очень__, есть противники... Khotya situatsiya yasna ochen', est' protivniki... Хотя ситуация __очень ясна__, есть противники... Khotya situaciya ochen' yasna, est' protivniki... *Although the situation is very clear, there are opponents...*
Syntax	Errors in the basic syntactic structure of the sentence (e.g., POS misuse) and other syntactic errors После *__мыть__ одежд она будет чистить окон. Posle (Prep) myt' (Verb) odezhd ona budet chistit' okon. После __стирки__ одежды она будет мыть окна. Posle (Prep) stirki (Noun) odezhd ona budet myt' okna. *After washing the clothes she will clean the windows.x*
Constr	Multiple word errors *__Я имею__ две собаки Ya imeyu dve sobaki __У меня есть__ две собаки. U menya est' dve sobaki. *I have two dogs.*

Error tag	Tag description
Lex	Lexical errors: wrong lexical choice, including collocations Ясно было, что надо *__посещать__ родственников. Yasno bylo, chto nado poseshchat' rodstvennikov. Ясно было, что надо __навещать__ родственников. Yasno bylo, chto nado naveshchat' rodstvennikov. *It was clear one had to visit (their) relatives.* Verb *посещать* 'visit' normally collocates with names of buildings and sites rather than people.
Prep	Errors in the use of prepositions *__На__ комнате есть хороший диван. Na (invalid preposition) komnate est' horoshij divan. __В__ комнате есть хороший диван. V komnate est' horoshij divan. *There's a nice sofa in the room.*
CS	Code-switching У берега бегала *__dog__. U berega begala dog. У берега бегала __собака__. U berega begala sobaka. *A dog was running along the shore.*
Idiom	Errors in idioms От дубов *__простыл и свет__. Ot dubov prostyl i svet. От дубов __простыл и след__. Ot dubov prostyl i sled. *Oaks are all gone* Wrong word is used in idiomatic expression *простыл и след* 'trace is gone'.

Secondary tags	
Miss	Missing item
Extra	Extra item
Transfer	Errors accounted for by dominant language transfer

Appendix B. RuERRANT applied to RLC-Test

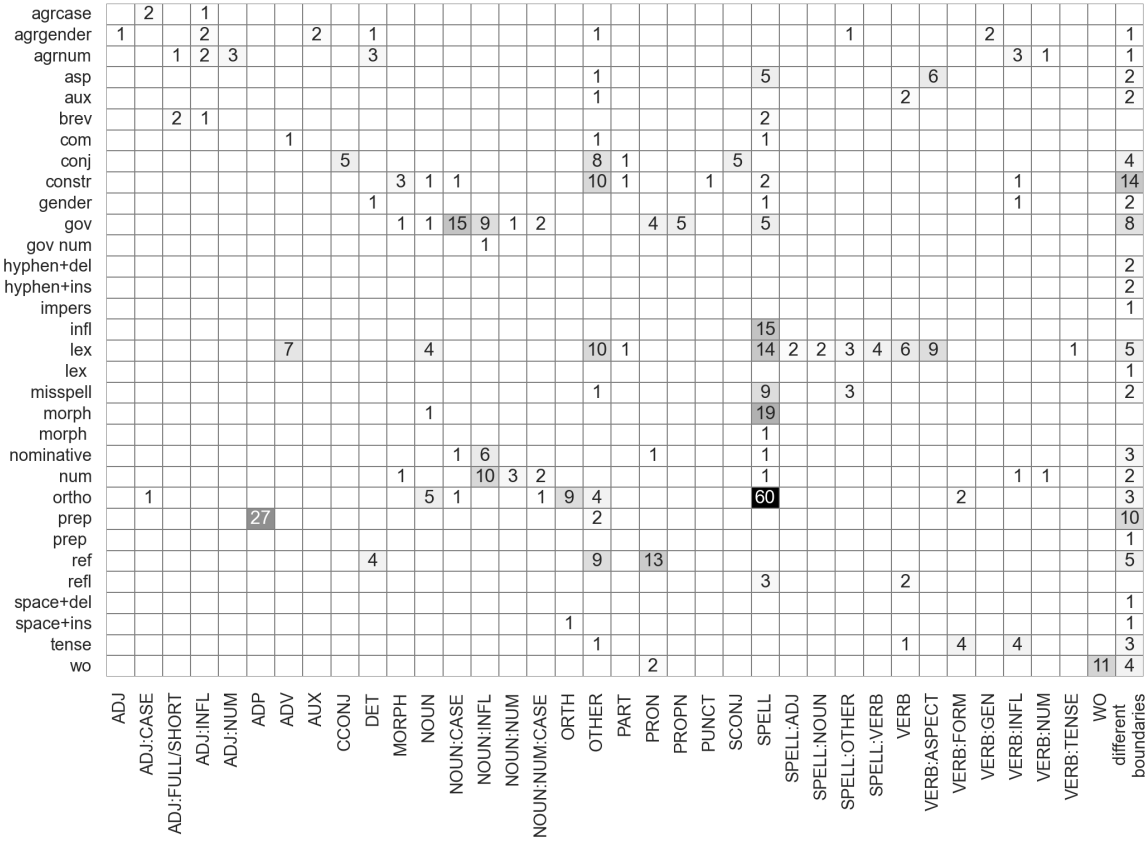

Figure 3: Confusion matrix for RuERRANT (Katinskaia et al., 2022) applied to the RLC-Test dataset. Rows correspond to RLC tags, and columns correspond to tags assigned by RuERRANT; cases where RuERRANT delimits errors differently than it is done in the dataset are labeled as “different boundaries”.