Computational Linguistics and Intellectual Technologies:
Proceedings of the International Conference "Dialogue 2017"

Moscow, May 31—June 3, 2017

MULTI-LEVEL STUDENT ESSAY FEEDBACK IN A LEARNER CORPUS[1]

Vinogradova O. I. (olgavinogr@gmail.com)[1],
Lyashevskaya O. N. (olesar@yandex.ru)[1,2],
Panteleeva I. M. (irapanteleeva@rambler.ru)[1]

[1]National Research University Higher School of Economics,
[2]Vinogradov Institute of the Russian Language RAS, Moscow, Russia

The paper presents the results of using computer tools and of designing an inspection program for the purposes of the automated and semi-automated syntactic, lexical, and grammar error analysis of student essays in a learner corpus. The texts in the corpus were written in English by Russian learners of English. In our experiment we compare the parameters of the essays graded by professional examiners as the best and those graded the lowest in the pool of about 2000 essays. At the first stage in the experiment we applied a syntactic tool for parsing the sentences and collected data regarding mean sentence depth and the average number of relative, other adnominal, and adverbial clauses, then analyzed the results of lexical observations in those texts (such as average word length, number of academic words, number of linking words and some others), and finally collected the statistics related to the errors pointed out in manual expert annotation. The parameters that had very different values for the "good" and for the "bad" essays are regarded by the authors as worthy parts of the feedback a student can get for the text uploaded into the learner corpus.

Key words: learner corpora, corpus research, essay evaluation, automated feedback, lexical complexity, syntactic complexity

[1] The article was prepared within the framework of the Academic Fund Program at the National Research University Higher School of Economics (HSE) in 2016–17 (grant № 16-05-0057 "Learner corpus REALEC: Lexicological observations") and by the Russian Academic Excellence Project "5-100".

АВТОМАТИЗИРОВАННЫЙ КОММЕНТАРИЙ К ЭССЕ В УЧЕБНОМ КОРПУСЕ

Виноградова О. И. (olgavinogr@gmail.com)[1],
Ляшевская О. Н. (olesar@yandex.ru)[1,2],
Пантелеева И. М. (irapanteleeva@rambler.ru)[1]

[1]Национальный Исследовательский университет «Высшая школа экономики», [2]Институт русского языка им. В. В. Виноградова РАН, Москва, Россия

В статье представлены результаты применения компьютерных инструментов оценки лексического и грамматического уровня текстов для автоматического или полуавтоматического анализа студенческих эссе в обучающем корпусе. Тексты в корпусе написаны на английском языке русскими студентами, изучающими английский язык. В нашем исследовании мы сравниваем разные параметры примерно двух тысяч эссе, которые оценены экзаменаторами как лучшие и как худшие работы. На первом этапе мы применили синтаксический инструмент для разбора предложений и собирали данные о средней глубине предложения и среднем количестве разных типов придаточных предложений, затем проанализировали результаты работы лексического инспектора (например, данные о средней длине слова, количестве слов из научного лексикона, количестве связующих слов), и, наконец, собрали статистику, связанную с ошибками, указанными в ручной экспертной аннотации. Параметры, существенно разнящиеся в «хороших» и «плохих» эссе, предполагается включить в форму, которую студент будет получать в режиме обратной связи после загрузки своей работы в корпус.

Ключевые слова: учебные корпуса, корпусные исследования, оценка эссе, автоматическая оценка текста, лексическая сложность, синтаксическая сложность

1. Introduction

It has many times been demonstrated over more than 20 years of learner corpora research that access to a learner corpus contributes greatly to the efficiency of L2 acquisition for both learners and instructors alike (Granger 2012; Granger et al. 2013). REALEC is a learner resource which has been in active use by English instructors teaching at the university level. It is the first collection in the open access of English texts written by Russian university students learning English. It is available with all errors in the student essays outlined by expert annotators (Vinogradova 2016). This paper looks at the syntactic complexity and at the lexical diversity range in the best essays of the past examination in comparison with those that were considered the worst among the examination essays written by the university students in a 2015 administration of the 2nd-year examination in English. This paper aims at evaluating which features constitute the indications of successful / unsuccessful text and can thus be included in automatic essay feedback that a student can get after uploading his / her essay in the corpus.

2. Related work

Published in 2015 Cambridge Handbook of Learner Corpus Research includes a few papers describing approaches to providing learners of a second/foreign language with automatic commentary on the quality of their written production. The papers with the focus on or related to the lexical features of the student text are Adel, 2015 and Granger, 2015. Tom Cobb and Marlise Horst in their chapter of Cambridge Handbook of Learner Corpus Research spoke about the generalizing role of a learner corpus in shedding light on second language acquisition by allowing the use of many computing tools inapplicable to separate texts (Cobb & Horst, 2015, pp. 185–206)

The choice of lexical parameters to be included in evaluation is discussed, for example, in Lavallée & McDonough, 2015. The adjacent filed—comparisons of student texts with authentic academic texts—were reported by researchers from University of Grenoble-II in their work which presents Apex, a system for automatic assessment of a student essay based on the use of Latent Semantic Analysis (Dessus & Lemaire, 2001). McCarthy & Jarvis, 2010 report the comparative assessment of different lexical features in the process of automated evaluation.

Application of syntactic parsing in corpus studies is the topic of many a work of the recent years. Many publications of authors working in cooperation with Daniella McNamara (like McNamara et al. 2011) relate diagnostics of advanced measures of linguistic complexity of a text to the application of an automated tool called Coh-Metrix designed to assess the characteristics of texts for different purposes, and syntactic sophistication is just one of them. The syntactic complexity analyzer by (Lu, Ai 2015), which provides a set of simple yet detailed measurements such as the mean length of clause, the number of dependent clauses and coordinate phrases, has currently become a state-of-the-art benchmark, although the need for more sophisticated measures is discussed in the professional community. The report of the TREACLE project with its reference to the use of the Stanford Parser in a learner corpus of works written by Spanish learners of English (Murcia-Bielsa & MacDonald 2013, page 337) became one of the starting points for our experiment.

The vast literature on the parameters of student writing used in pedagogical expert evaluation has been discussed for examinations of different types administered by different institutions. In view of situation with the English examination at the Higher School of Economics, we have chosen for reference one of the recent and most detailed reports that consider writing potential indications in IELTS examination (Cotton & Wilson 2011). According to this report, the four parts of the grade assigned by examiners are measured by looking at:

- the number of words, relevance to the topic in the question, and coverage of all parts of the question (Task Achievement/Task Response);
- organisation in paragraphs, connection of sentences and paragraphs with logical links and referencial tools, no or little repetition (Coherence and Cohesion);

- use of appropriate academic words and collocations, use of paraphrase to avoid repetition, correct spelling (Lexical Resource);
- use of a variety of grammatical forms, combination of short and complex sentences, and not too many grammatical mistakes (Grammatical Range and Accuracy).

The parameters outlined in this work have defined our selection of the features to be included in the experiment.

3. Experiment setup

The objective of our corpus experiment was to establish the correlation between the grades that examination essays were given by experts, on the one hand, and the indices of the automated analyses of student texts from a learner corpus, on the other, with the more distant goal of outlining the best features for automated essay feedback. The experiment was carried out over essays in IELTS format[2] written by 2nd-year Bachelor students in their final English examination in 2015. The writing part of this examination includes two tasks requiring that each testee writes one essay not less than 150 words long (essay1), the other about 250 words long (essay2), both within the period of one hour. The essays are assessed just as was stated in Section 2—by the following criteria: task response, coherence and cohesion, lexical resource, grammatical range and accuracy. The tasks were given to almost a thousand students. After the examination, the essays were evaluated by independent EFL raters, who assigned each task a holistic grade in the percentage points up to 100. When the essays were uploaded to REALEC, expert annotators spotted errors in the essays and added manual annotations clasifying those errors. For the purposes of the experiment, two groups of essays were chosen out of almost two thousand essays—those that the experts graded at 75% and over (33 essays), and those that got the grade of 30% and lower (43 essays). Essays in either group were subjected to the three stages of analytical procedures: 1) POS and dependency parsing; 2) automatic evaluation of each text using the built-in lexical tool REALEC-Inspector designed at the School of Linguistics, Higher School of Economics; and 3) statistical analysis of the expert annotation. The results of all three stages in two groups were compared with each other, and the conclusions are reflected upon in the final section of the paper.

4. Data analysis

4.1. POS and syntactic parsing

The sentences were processed with the open-source tagger and dependency parser UDpipe (Straka 2015). Each word was tokenized and tagged for POS and dependency types, so that the depth of the tree was easy to calculate for the sentence.

---

[2] IELTS (International English Language Testing System) is a test of English language profi-ciency for non-native speakers of English. IELTS certificates are recognized in more than 120 countries round the world and cover all four language skills—listening, reading, writing, and speaking.

For each essay, the average syntactic depth was counted with the maximum and minimum depths stated. The efficiency of UDpipe on REALEC essays was comparable to that achieved with the Stanford parser in (Murcia-Bielsa, MacDonald, 2013): the cases of wrongly defined arcs (unlabeled attachment) were minimal and could mainly be accounted for by a learner-driven gaps in syntactic structures and by distant dependencies. As for dependency relations, we only took into account the following labels: relative clauses (acl:relcl), groups with participles as the head (acl), and adverbial clauses (advcl). These relations were checked manually for false positives.

The mean syntactic depth of the sentence ranges from 1 (no more than one dependency down from the sentence head, as in *It is wrong!*) to 10. The analysis has revealed insignificant difference between the best and worst essays in their average depth (best: mean = 4.61, sd = 1.66; worst: mean = 4.12, sd = 1.57). The amount of particular subordinate clause types per essay, on the contrary, differs significantly between the best and worst essays, see Table 1 (mean and 95% CI values are shown).

Table 1 Subordinate clause types per essay

Grade Cat	mean.acl	mean.acl:relcl	mean:advcl
best	3 ± 0.82	3.25 ± 0.85	5.41 ± 1.07
worst	1.21 ± 0.42	1.43 ± 0.38	1.86 ± 0.5

Table 2 shows Pearson's pairwise correlation between the following factors: expert's grade (absolute values), mean sentence depth, number of adnominal clauses including participle groups (N_acl), number of relative clauses (N_acl:relcl), number of adverbial clauses (N_advcl), total number of subordinate clauses. The correlation between the number of relative and adverbial clauses is moderate, while the acl score behaves differently. Furthermore, it is predictable that neither of these features correlate with the mean syntactic depth. The amount of adverbial clauses correlates best with the grade since they are used more frequently, and it shows the students' ability to express a variety of causal, temporal, and other relations between propositions.

Table 2 Correlation of the syntactic features

	Mean Depth	N_acl	N_acl: relcl	N_advcl	N_All SubordCl
Grade	0.203	0.397	0.462	0.599***	0.630
MeanDepth		0.375	0.311	0.179	0.346
N_acl (adnominal clauses)			0.355	0.383	0.698
N_acl:relcl (relative clauses)				0.548	0.785
N_advcl (adverbial clauses)					0.867

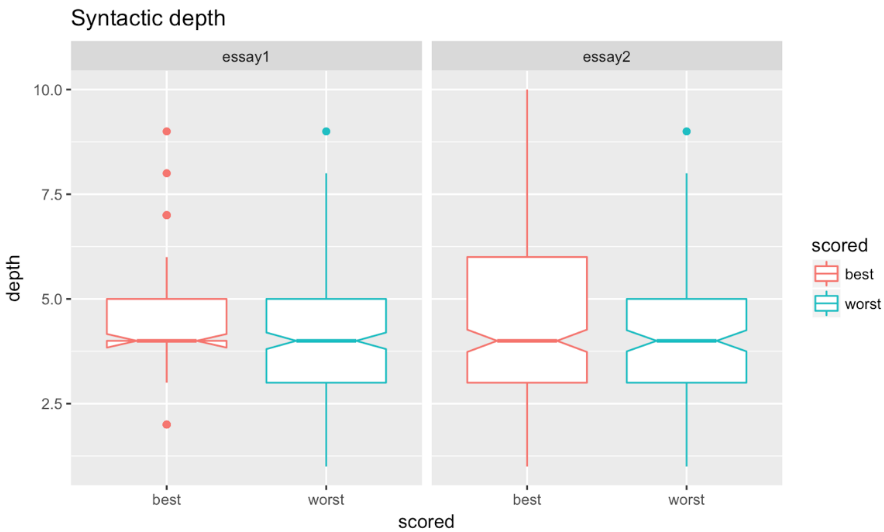

Fig. 1. Mean syntactic depth of sentences by essay type and grade category

Thus, these features should constitute the basis of the student automatic feedback, although more detailed analysis is needed.

4.2. Lexical evaluation with REALEC-Inspector

While considering which parameters to take up for the lexical inspection, we decided to start with those that have been described by the authors working in corpus linguistics as automatically indicative of the level of lexical variety. McCarthy & Jarvis, 2010 pointed out the importance of the length of words and length of sentences as the criteria for the automated lexical evaluation. Frequency of a word in the Corpus of Contemporary American English [3] was justified as a parameter in lexical evaluation in Crossley, Cobb, & McNamara, 2013 and Vongpumivitch, Huang, & Chang, 2009, while for checking the use of academic vocabulary the three lists have been argued in corpus linguistics works: the Coxhead Academic Word List (cf. Coxhead, 2000 and Coxhead, 2011), the one in the Corpus of Contemporary American English[4] and the Pearson academic collocation list[5]. That was why all those parameters were included in our experiment:

1) Number of words in the essay
2) Average length of a sentence in the essay
3) Length of the longest sentence in the essay
4) Average length of word in the essay

[3] COCA (The Corpus of Contemporary American): http://corpus.byu.edu/coca/, http://www.wordfrequency.info/, http://www.academicvocabulary.info/

[4] http://www.academicvocabulary.info/

[5] http://pearsonpte.com/research/academic-collocation-list/

5) Length of the longest word in the essay
6) Number of words of each level of CEFR in the essay
7) Number of words from the COCA frequency lists
8) Number of academic words in the essay with repetitions and without them
9) Number of repetitions of words used in the essay. The word most frequently repeated.
10) Number of linking words and expressions in the essay

For the purposes of the experiment and for the broader perspectives of providing automated lexical analysis for any learner text, we developed the application REALEC-Inspector. Its homepage was placed in the Moodle environment with the options either to upload a text in an input window, or browse for the text in REALEC. For the indices listed above the inspection of the text with REALEC-Inspector opens right after the text with the short statistical summary, of which a sample is given in Figure 2.

Statistical summary

Number of words: 290
Average sentence length: 18.875 words.
Max sentence length: 32 words.
Average word length: 5.10104529617 letters.
Max word length: 18 letters.
CEFR
A1: 49
A2: 16
B1: 11
B2: 7
C1: 1
C2: 0
Unclassified: 38
Stopwords: 36
Frequency:
1-500: 39
501-3000: 36
>3000: 47
Academic words: 71 (51 unique)
Word repetitions: 44 (('children', 6) is the most repeated)
Linking phrases: 12
Pearsons collocations: 7 (5 unique)

Fig. 2. List of statistics for the essay under inspection

The statistical summary is then followed by detailed comments for each item on the list, and for some of them the Inspector provides diagrams. Here we show some of them giving the sources of the reference materials applied.

For the histogram of CEFR words distribution (Fig. 3), Word Family Framework was used (the possibility to use English Vocabulary Profile instead has been reserved), and each word—with the exception of stop words (there are 153 of them in the application)—is lemmatized with the help of NLTK. Words that the system was unable to relate to a particular CEFR level (among them are misspelled words) are categorized as “Unclassified” and given on the histogram in column 0.

CEFR

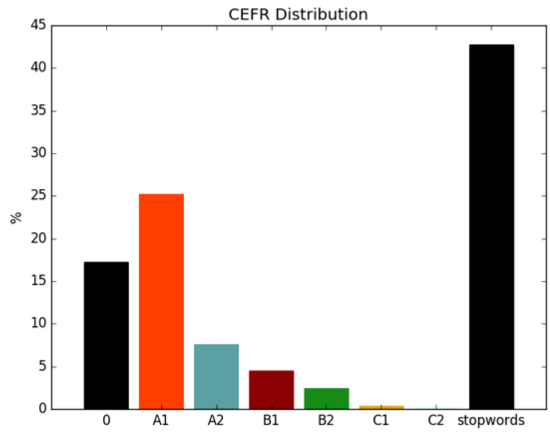

Fig. 3. Sample picture of distribution of CERF-level words

After that the author of the text gets the list of words from the essay that are among the 500 most frequent words in COCA, and then those that are among the 3,000 most frequent words in COCA. Stop-words are again excluded.

The next comment is on the occurrence of academic words from the list which is a combination of two—the Academic Word List Coxhead and the Corpus of Contemporary American English List of Academic Words. As a result, if a word from an essay belongs to either of these lists, it will be counted.

In this section the author will see a diagram showing the distribution of the number of academic vocabulary items across all essays in the corpus, with the red line marking the average index in the essay under inspection for the author to compare with other essays. Useful as it may be, we don't bring in an example of the diagram here, as it goes beyond the scope of this experiment to research whether the comparison with all essays in the corpus gives students the way to understand where their writing stands as far as sophistication is concerned.

Next, five most frequently repeated words (those that are not stop-words) are shown (see Figure 4). The need for demonstrating the ability to paraphrase can be emphasized here.

Word Repetitions

Overall there are 44 word repetitions in this text. The most common of them are:
children: 6 times
parents: 6 times
family: 5 times
society: 4 times
work: 4 times

Fig. 4. Sample list of repetitions in the essay

The number and the list of linking words and introductory expressions used in the essay are accompanied at the next stage next by the indication of their categories (Comparison, Time and sequence, Addition, Cause and Effect, Conclusion and summary, Examples, Concession, Repetition, Giving reasons, Explanations, Contrast (Figure 5).

Linking Phrases

There are 12 introductory phrases.
Comparison: 0
Time and sequence: 5
then: 2
now: 2
nowadays: 1
Addition: 4
also: 3
moreover: 1
Cause and Effect: 0
Conclusion and summary: 1
in conclusion: 1
Examples: 1
for example: 1
Concession: 0
Repetition: 0
Giving reasons, explanations: 0
Contrast: 1
however: 1

Fig. 5. Sample list of linking words in the essay

The comparison of the use of linking phrases in the essays under inspection with all other essays in the corpus can also be presented to the author on the histogram as an additional option, see Figure 6.

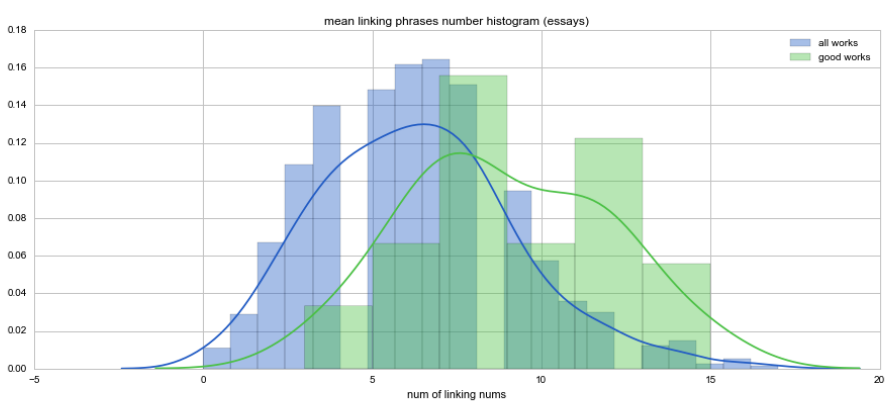

Fig. 6. Distribution of linking phrases number: good essays vs all essays

The inspector then gives the number and the list of collocations from the essay if they are on the Pearson Academic Collocation List (see Figure 7).

Pearsons Collocations

There are 7 collocations, 5 of which are unique.
nuclear family; dominant position; closer look; wide range; modern society;

Fig. 7. Sample list of collocation in the essay

There is also an option to ask for the visualization of the text with one of the three features:

1) with words of different CEFR levels presented in different colours;
2) with words of different COCA frequencies presented in different colours;
3) with academic words highlighted.

Table 4 below gives the summary of the comparative analysis of lexical features under investigation for the two sets of essays (the best and the worst).

Table 4. Synopsis of the comparison between the experimental sets

Parameters for automated lexical inspection	Essays scored 75% and higher		Essays scored lower than 30%	
	Task 1	Task 2	Task 1	Task 2
1) Number of words in the essay	203	292	174+	161***
2) Average length of a sentence in the essay	21	20	17^{ns}	16^{ns}
3) Length of the longest sentence in the essay	37	39	33^{ns}	30^{ns}
4) Number of academic words in the essay (with repetitions/without repetitions)	41/28	69/51	33^{ns}/18+	42***/ 29***
5) Number of linking words and expressions in the essay	5	7	3^{ns}	4^{ns}
6) Number of collocations from the Pearson academic collocation list in the essay (with repetitions/ without repetitions)	0,8/0,8	0,73/0,73	0,38***/ 0,35***	0,38***/ 0,38***

The parameters that have not been included in the table are those whose values were approximately the same for essays scored highly and for those scored very low: the average word length, maximum word length, number of word repetitions, words of different CEFR levels. In general, “good” essays have more CEFR scale words at each level, as well as more words of high frequency in COCA, but not many more of them. This is rather due to the fact that texts showing better writing proficiency have higher overall number of words. So, the figures have not been included in the comparison.

It is clear from the table that the best characteristics distinguishing texts that are more likely to get a good score from those that are less so are the following:

- average sentence length;
- the number of words from academic vocabulary lists;
- the number of academic collocations.

As these parameters can be evaluated by a software application, they will be included in the automated feedback provided to authors of learner texts.

4.3. Error analysis

Error annotation in REALEC is based on the classification scheme of about 150 specific error tags organized into a tree-like structure with 7 classes of errors—Spelling, Capitalisation, Grammar (Morphology), Grammar (Syntax), Vocabulary and Discourse. The overall number of errors spotted by the annotators varies both in the "best" and "worst" essays, and it can be explained by the following consideration: authors with stronger writing potential make more effort to apply sophisticated morphological and syntactic features than those with less proficiency, and as a result the former run a greater risk of making mistakes than the latter. The approach in the examination of IELTS type is to encourage attempt at higher sophistication rather than penalize incorrectness in complicated constructions, either grammar or vocabulary, so the first group of authors get higher grades more often than the second. On the other hand, weaker writers are more prone to making mistakes in simple cases than those with better writing skills. And these two opposing arguments lead to the situation, in which the average numbers of errors in an essay is not a good indicator of the writing proficiency, nor does the average number of syntactic and/or discourse errors demonstrate the level of syntactic complexity of the text. The tagging statistics across the "best" and the "worst" essays within the scope of our research shows exactly the same distribution in Table 5.

Table 5. Error annotation indices in the experimental folders

	Essays scored 75% and higher	**Essays scored lower than 30%**
Average number of all error tags in one essay	19	19.5
Minimum and the maximum number of all error tags	3 to 60	10 to 66
Average number of syntactic tags	2	3
Average number of discourse tags	3	3

The two possible ways of demonstrating annotation results in the feedback are either to give the overall number of tags in comparison with the average number across the folder with similar essays, as well as the number of the tags for repeated categories of errors in the essay, also against the average in the folder, or just summarise the numbers like this:

> The expert pointed out 14 errors (19 average) (5 syntactic, 4 discourse errors, and 3 morphological). You may need to review the use of different syntactic constructions.

automated tagging as well.

5. Conclusions

The observations over the features of many student essays in the learner corpus have confirmed the following points important for working out approaches to automated evaluation of student writing and to automated feedback for student writing:

- word length and number of repetitions are insignificant as indicators of the writing proficiency.;
- the numbers of words at each CEFR level and of those with high COCA frequency are to some extent larger in essays highly evaluated by experts, but their relevance as a part of automated feedback has to be confirmed further. Dependence of the lexical variety and complexity on the length of a piece of writing has many times been emphasized in corpus linguistic research, but the texts of essays at our disposal were of two types—not less than 150 and not less than 250 words, so for the purposes of our experiment all statistical analysis in the lexical inspection was carried out separately for the two types of essays—descriptions of the illustration(s) given in the task and argumentative essays.

The results of the comparison allow us to state that automatic application of both syntactic parsing and lexical inspection will provide good suggestions for improving students' writing potential and can be considered to be good predictions of the success/failure in the examination.

To cater for those users who may look for independent training, we are thinking of giving the results of the reported research one more use in a way of introducing a few computational modules that will show a user the basic characteristics of the text he/she is composing right in the process of typing in an essay, namely, instant demonstration of such features as superfluous repetitions, misspelled words, low variability of syntactic constructions, and some others.

References

1. *Annelie Ä.* (2015), Variability in learner corpora, In Granger S., Gilquin G., Meunier F. (eds.), The Cambridge Handbook of Learner Corpus Research, Cambridge, UK: Cambridge University Press, pp. 401–422.
2. *CEFR* (2001), The Common European Framework of Reference for Languages: Learning, Teaching, available at: https://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf
3. *Cobb T., Horst M.* (2015), Learner corpora and lexis, In Granger S., Gilquin G., Meunier F. (eds.), The Cambridge Handbook of Learner Corpus Research, Cambridge, UK: Cambridge University Press, pp. 185–206.
4. *Cotton F., Wilson K.* (2011), An investigation of examiner rating of coherence and cohesion in the IELTS Academic Writing Task 2, IELTS research reports. Vol. 12, IELTS Australia and British Council, pp. 1–76.
5. *Coxhead A.* (2000), A new academic word list, TESOL Quarterly, Vol. 34 (2), pp. 213–238
6. *Coxhead A.* (2011), The academic word list 10 years on: Research and teaching implications, TESOL Quarterly, Vol 45 (2), pp. 355–362.

7. *Crossley S. A., Cobb T., McNamara D. S.* (2013), Comparing count-based and band-based indices of word frequency: Implications for active vocabulary research and pedagogical applications, System, Vol. 41 (4), pp. 965–981.
8. *González-López S., López-López A.* (2015), Lexical analysis of student research drafts in computing, Computer Applications in Engineering Education, Vol. 23 (4), pp. 638–644.
9. *Granger S.* (2012) How to use Foreign and Second Language Learner Corpora, Research Methods in Second Language Acquisition: A Practical Guide, Blackwell, Oxford. Ch.2, pp. 5–29.
10. *Granger S.* (2015), The contribution of learner corpora to reference and instructional materials design, The Cambridge Handbook of Learner Corpus Research, Cambridge, UK: Cambridge University Press, pp. 485–510
11. *Granger S., Gilquin G., Meunier F.* (eds.) (2013), Twenty Years of Learner Corpus Research. Looking Back, Moving Ahead, Proceedings of the First Learner Corpus Research Conference, Vol. 1, Presses universitaires de Louvain.
12. *Lavallée M., McDonough K.* (2015), Comparing the Lexical Features of EAP Students' Essays by Prompt and Rating, TESL Canada Journal, Vol 32 (2), pp. 30–44.
13. *Lemaire B., Dessus P.* (2001), A System to Assess the Semantic Content of Student Essays, Journal of Educational Computing Research, Vol. 24(3), pp. 305–320.
14. *Lu X., Ai H.* (2015), Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds, Journal of Second Language Writing, Vol. 29, pp. 16–27.
15. *McCarthy P. M., Jarvis S.* (2010), MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment, Behavior Research Methods, 42, no 2, pp. 381–392.
16. *McNamara D. S., Graesser A. C., Cai Zh., Kulikowich J. M.* (2011), Coh-Metrix Easability Components: Aligning Text Difficulty with Theories of Text Comprehension, Proceedings of the Annual meeting of the American Educational Research Association (AERA 2011), New Orleans, LA, available at: https://www.researchgate.net/publication/228455723_Coh-Metrix_Easability_Components_Aligning_Text_Difficulty_with_Theories_of_Text_Comprehension
17. *Murcia-Bielsa S., MacDonald P.* (2013), The TREACLE project: Profiling learner proficiency using error and syntactic analysis, Twenty Years of Learner Corpus Research. Looking Back, Moving Ahead, Proceedings of the First Learner Corpus Research Conference, Presses universitaires de Louvain.
18. *Straka M., Hajič J., Straková J.* (2015), UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing, Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC'16), 23–28 May 2016, Portorož, Slovenia, pp. 4290–4297, UDpipe page available at: http://lindat.mff.cuni.cz/services/udpipe/run.php
19. *Verhelst N., Van Avermaet Piet, Takala S., Figueras N., North B.* (2009), Common European Framework of Reference for Languages: learning, teaching, assessment, Cambridge University Press.

20. *Vinogradova O.* (2016), The Role and Applications of Expert Error Annotation in a Corpus of English Learner Texts, Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog 2016" [Komp'yuternaya Lingvistika i Intellektual'nye Tekhnologii: Trudy Mezhdunarodnoy Konferentsii "Dialog 2016"], Moscow, Russia, pp. 740–751
21. *Vongpumivitch V., Huang J. Y., Chang Y. C.* (2009), Frequency analysis of the words in the Academic Word List (AWL) and non-AWL content words in applied linguistics research papers, English for Specific Purposes, Vol. 28 (1), pp. 33–41.