AUTOMATED VOCABULARY EVALUATION IN A LEARNER CORPUS

O.I. Vinogradova

Research University Higher School if Economics
20, Myasnitskaya str., Moscow, 101000, Russian Federation

When learners of English have to write texts in academic English, and in preparation for written parts of English examinations in particular, students need feedback when they submit their written works. Besides tips concerning content, coherence, and cohesion, grammatical range and accuracy, all of which are standardly included in instructor's comments, students need advice on directions towards lexical improvements. However, such feedback requires a huge amount of time and effort on the part of instructor, whose workload is heavy enough to make any extra effort undesirable. A learner corpus may help in this, as its wealth of student texts allows researchers to make use of the many samples of student writing by applying certain computer tools. This paper reports the development of a system of automated lexical inspection of student works. Initially, we used essays in the corpus to work out which formal parameters in the essays demonstrate in what ways essays that have been evaluated highly by the examination experts can be distinguished, and then we applied those parameters in the process of automated inspection, after which we proceeded to checking the correlation between the inspection results and the traditional grading. Finally, after a system of lexical inspection of student essays was established, which paves way to the development of automated lexical feedback in order to orient students in how to improve the quality of their writing.

Key words: learner corpus, lexical proficiency, corpus research, feedback on student essay

1. INTRODUCTION

The role of access to a learner corpus has proved to increase efficiency of L2 acquisition for learners as well as teaching efficiency for EFL instructors [1; 2]). This paper presents a computer tool for a learner corpus designed at the computer linguistics department of the Higher School of Economics for both categoties of users.

REALEC, Russian annotated Learner Corpus, set up at the School of Linguistics, is the first collection of English texts written by Russian students learning English easily available in the open access at (http://www.realec.org/). All errors made by Russian students in their academic writing in English are pointed out to them with special tags by expert annotators (EFL instructors, as a rule). The annotation process is controlled by the research team responsible for consistency in tagging, as well as development of the learner corpus. One of the directions of the development is to look at the lexical features used in student essays. Our approach in this research was to find such lexical features in the essays scored highly by experts which will be different from those features in the essays scored with the lowest grades.

Methods and related research

The essays for the research were those from the past examination in IELTS format administered to 2nd-year Bachelor students at the Higher School of Economics. Each student had to write two essays — the description of an illustration showing results of a particular research, about 150 words long, and an argumentative essay, about 250 words long, both within the period of one hour. After the examination, the essays were evaluated by EFL experts and assigned a grade for either task in the percentage points. For the purposes of the experiment, two groups of essays were selected — those that the experts graded at 75% or higher, and the those with grades of 30% or lower. We designed a special procedure of lexical inspection, i.e. computing the performance of a set of lexical features commonly applied in automated evaluation of a written text. The adjacent field — comparisons of student texts with authentic academic texts — were reported by Canadian researchers from University of Grenoble-II Benoi t Lemaire and Philippe Dessus in their work which presents Apex, a system for automatic assessment of a student essay based on the use of Latent Semantic [3]. The procedure was called REALECInspector, and the essays in either group were subjected to it. The results of this inspection in two groups were compared with each other. The code for REALEC Inspector was developed from the automated readability system suggested by Konstantin Druzhkin [4; 5].

The choice of the parameters to be included in evaluation is discussed in [6—9].

For the purposes of our experiment the comparisons were drawn across the following features:

— Length of words;

— Length of sentences;

— Distribution of words across the Common European Framework scale levels (A1-C2);

— Frequency of each word in the Corpus of Contemporary American English;

— Use of academic vocabulary from the two lists — the Coxhead Academic Word List in [10; 11]) and in list included in the Corpus of Contemporary American English COCA);

— Repetitions;

— Use of linking words;

— Use of collocations (as attested by the presence on the Pearson academic collocation list).

The objective of our experiment was to establish the correlation between the grades that were given by experts and the automated evaluation of lexical content on the basis of certain criteria.

Our hypothesis was that the criteria applied in the developed application would be sufficient for a valid preliminary evaluation of lexical variability of written papers.

2. DISCUSSION

The works in the experiment consisted, on the one hand, of 45 sets (2 essays in each set — an argumentative essay and a description of a diagram or diagrams), — those that were marked by experts at 75% (out of 100%) and higher, and in the second group 900 sets (the same types of essays) marked below 75%. The results of lexical inspection application in the two groups were analysed in comparison. It was revealed that certain

characteristics (for example, the average sentence length, the number of words from academic vocabulary lists) are significantly different in the two groups, which proves that grades assigned by experts do have certain correlations with parameters that can be evaluated by a software application. For one, there are more words on average in "good" texts.

The following are the main numerical results of the comparative analysis:

There are 188 words on average in a diagram description in a "good" essay versus exactly the same number of words in a diagram description from the big collection of essays.

There are 294 words on average in a "good" argumentative essay versus 268 words on average in an argumentative essay from the big collection.

On average, sentences are longer in "good" essays: there are 21.15 words in a "good" diagram description versus 18.4 words in a diagram description from the entire collection, and 20.34 words in a "good" argumentative essay versus 18.1 words in an argumentative essay from the entire collection. The maximum sentence length is longer in "good" texts: it is 37.5 words in a "good" diagram description versus 33.14 words in a diagram description from the entire collection, and 37.4 words in a "good" argumentative essay versus 35.75 words in an argumentative essay from the big collection. All in all, the mean sentence length positively correlates with the quality of learner writing, while the extremes in the length of a sentence rather relate to poor quality in an essay.

At the same time, the average word length was approximately the same for both "good" and "average" essays in the first collection of essays, both argumentative and diagram descriptions. However, with the increase in the number of essays scored highly, the positive correlation of this feature brought us to understanding that the word length is a factor to consider when giving feedback to students. The same stands true for the longest words in the papers and the number of word repetitions.

The number of linking words in "good" diagram descriptions was recognized as a positive factor, though not overwhelmingly so — 3.6 versus 3.23. However, "good" argumentative essays demonstrate a significant difference: 8.97 versus 6.33.

The number of collocations from Pearson's list (with repetitions) for diagram descriptions was 1.35 in "good" texts versus 0.4 in "ordinary" descriptions, and 1.62 in "good" argumentative essays versus 0.71 in "average" argumentative essays.

The number of collocations without repetitions was for diagram descriptions — 0.88 in "good" ones versus 0.71 in "ordinary" descriptions, and for argumentative essays — 1.46 in "good" versus 0.67 in "average".

In general, "good" essays have more CEFR scale words at each level, but not many more. This is rather due to the fact that the good papers have more words altogether. So, these figures will not be considered as a separate feature for a positive feedback.

The same stands true for COCA frequencies. The "good" essays on the whole have more words at each level.

At the same time, there are notably more words from academic vocabulary lists in "good" essays: for diagram descriptions it is 43 pieces of academic vocabulary on average versus 36 (with repetitions), and 28 versus 22 (without repetitions), and for argumentative essays, the average number of academic vocabulary items is 70 versus 56 (with repetitions), and 51 versus 40 (without repetitions).

Table below gives the synopsis of the significant differences between essays scored highly and the rest of the essays.

Table

Features of "good" essays in comparison with the mean values of the rest of the REALEC essays

Parameters for automated lexical inspection	Essays scored as high as 75% and higher		Essays scored lower than 75% by experts	
	Task 1	Task 2	Task 1	Task 2
1. Number of words in the essay	188	294	188	268
2. Average length of a sentence in the essay	21.15	20.34	18.4	18.1
3. Length of the longest sentence in the essay	37.5	37.4	33.14	35.75
4. Number of academic words in the essay (with repetitions/without repetitions)	43/36	70/56	28/22	51/40
5. Number of linking words and expressions in the essay	3.6	8.97	3.23	6.33
6. Number of collocations from the Pearson academic collocation list in the essay (with repetitions/without repetitions)	1.35/0.88	1.62/1.46	0.4/0.71	0.71/0.67

The results of the comparisons between "good" and average essays have allowed us to set up an automated application called REALEC Inspector, which presents automated feedback to authors of learner texts uploaded to the corpus (argumentative essays or description of diagrams) and provides some statistical information based on the comparison of the essay's formal features with the average figures for an essay of this type collected in REALEC, as well as offers some recommendations for improvements.

The stages of work with this application are the following.

There is an input window on its homepage with the "inspect" button to open the page for the lexical analysis.

The first thing on the page that appears after pressing Inspect button is the essay itself. Then comes the short statistics on:

— Number of words in the essay;
— Average length of a sentence in the essay;
— Length of the longest sentence in the essay;
— Average length of word in the essay;
— Length of the longest word in the essay;
— Number of words of each level of CEFR in the essay;
— Number of words from the COCA frequency lists;
— Number of academic words in the essay with repetitions and without them;
— Number of repetitions of words used in the essay. The word most frequently repeated;
— Number of linking words and expressions in the essay;
— Number of collocations from the Pearson academic collocation list in the essay.

In this list of short statistics, each line may be open to reveal detailed lists, comments, or the necessary diagrams. For the histogram of CEFR words distribution (Fig. 1), Word Family Framework was used (the possibility to use English Vocabulary Profile instead has been reserved as well), and each word is lemmatized with the help of NLTK. Stopwords

(153 on the list) are excluded. Words that the system was unable to relate with a particular CEFR level are categorized as "Unclassified" (some misspelled words are among them).
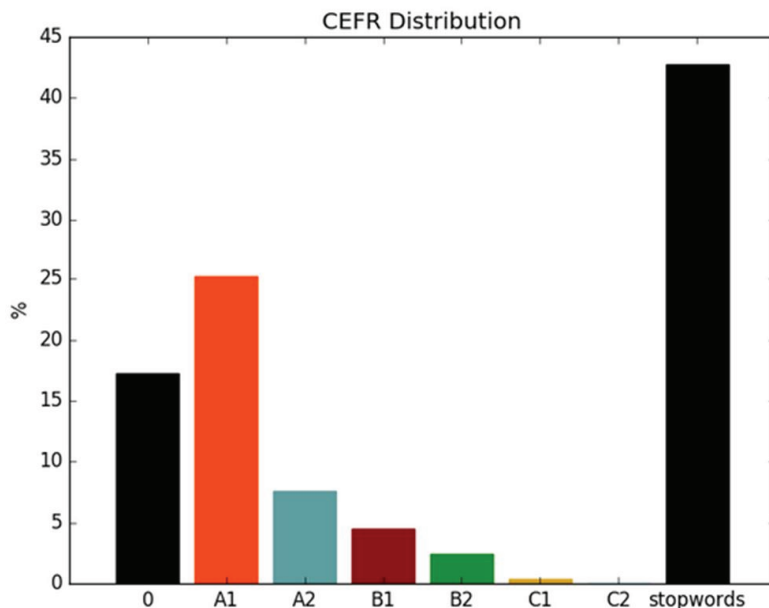

Fig. 1. Distribution of words in a short learner text by CEFR levels

For "Number of words from the COCA frequency lists" 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 3000 most frequent words in COCA. Stop-words are again excluded. The statistics is presented in the following way:

> **Frequency**:
> 1-500: 39
> 501-3000: 36
> >3000: 47

The occurrence of academic words is the statistics drawn from the list which is a combination of two — the Academic Word List Coxhead and the Corpus of Contemporary American English. As a result, if a word belongs to either of these lists, it will be counted. The result record is returned in this way:

> **Academic words:** 71 (51 unique)

After the number of words that have been used more than once in the essay the author will see the word that was repeated the highest number of times added to the statistics, namely:

> **Word repetitions:** 44 ('children' 6)

The next lines presents an impirtant index of the number of linking tools in the text, which looks like the following:

> **Linking phrases**: 12

The number and the list of linking words and introductory expressions used in the essay are accompanied 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.

The inspector then gives the number and the list of collocations from the essay if they are on the Pearson Academic Collocation List (see Fig. 2 below).
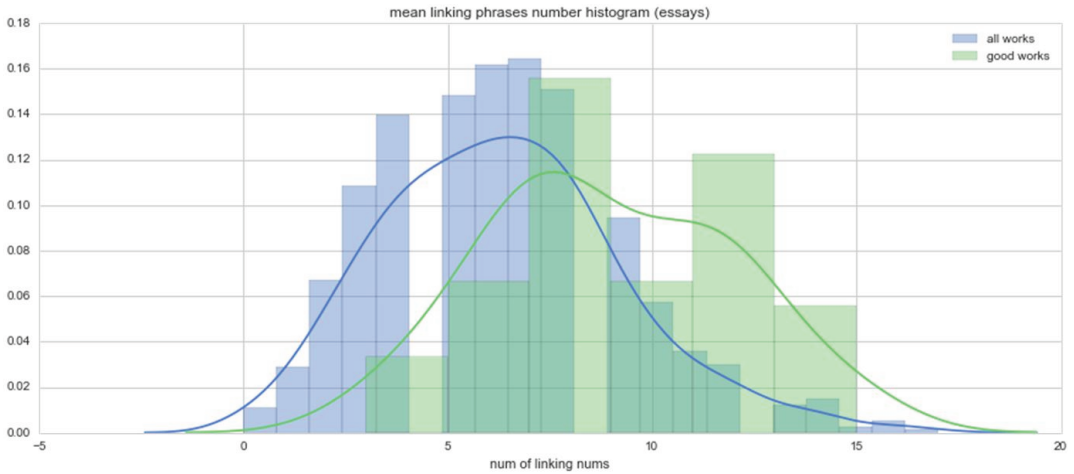

Fig. 2. Diagram showing "good" and "average" distribution of the number of collocations from the Pearson Academic Collocation List

48.0% of texts in the realec corpus have mean sentence length less than in the source text.
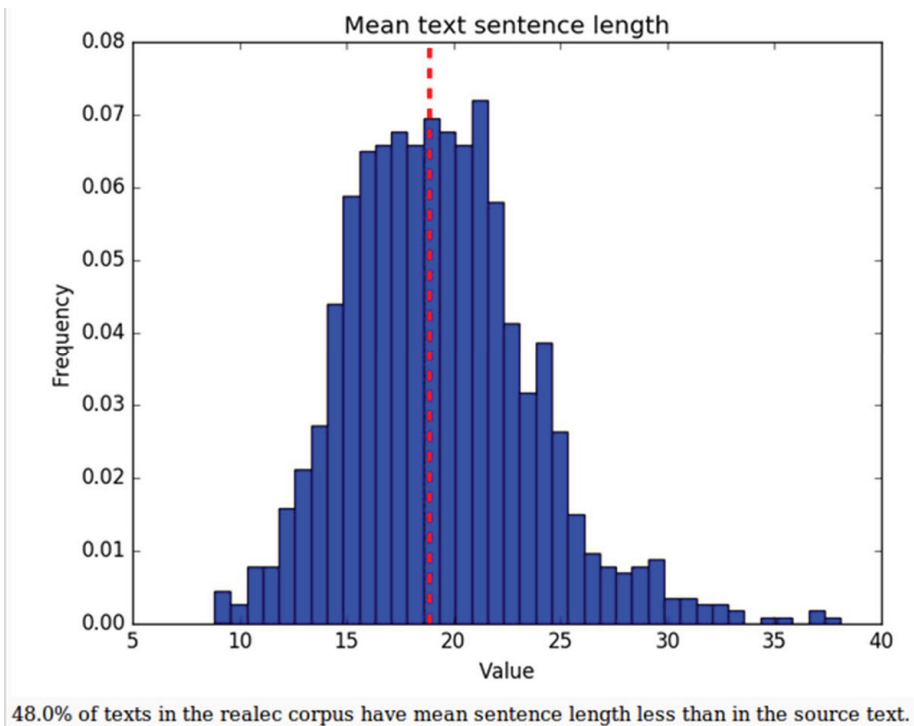
Fig. 3. Histogram for the mean sentence length in the text

If one clicks on any lines with statistics, a diagram or detailed list are presented. Fig. 3 gives the example of the distribution of average sentence length in the corpus. and one can get a similar diagram for the distribution of average word length. The red line on the diagram marks the average index in the essay under inspection for the author to compare with other essays. The comparison can also be numerical, as percentage is given here as well. For example, if under the diagram with the average sentence length there is a figure of 90%, it means the average sentence length is longer than in 905 of all essays in the corpus. More often than not, it is a feature of a good essay, as it implies that sentences are more sophisticated than in the majority of essays. On the contrary, low percentage number is the result of oversimplified sentence structure and has to bepointed out to the author as deficiency.

The last stage of inspection implies the use of syntactic parser UD pipe, which will be described in detail in a separate publication, but we need to mention it here, because the combination of the statistical analysis described above and the results of parsing account for final recommendations to the author, which — in case of mostly positive — may looks like this:

You have sufficiently complex sentences in your essay. Keep it up!

3. CONCLUSION

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

1. Overall, sentence length and numbers of linking words, collocations, and academic words are larger in essays highly evaluated by experts.

2. Word length and number of repetitions are insignificant as indicators of the writing proficiency.

3. The numbers of words at each CEFR level and of those with high COCA frequency are greater in essays highly evaluated by experts.

4. All parameters in automated inspection except word length are valid in distinguishing essays scored highly by experts, so the application can work as the preliminary stage in evaluating writing proficiency, but not instead of expert evaluation. Nevertheless, REALECInspector makes up a good suggestion for students' independent work on how to expand their writing potential.

With independent training in mind, we are thinking of introducing a few more computer tools that will be of use in the process of writing an essay in the corpus, which will increase the convenience of writing in the corpus in future — time management system, instant demonstration of low-profile features like superfluous repetitions, misspelled wirds, etc.

REFERENCES

1. Granger, S. 2012. "How to use Foreign and Second Language Learner Corpora". In A. Mackey and S.M. Gass (eds), Research Methods in Second Language Acquisition: A Practical Guide Blackwell, Oxford. Ch. 2. Pp. 5—29. Print.

2. Granger, S., G. Gilquin, and 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. Print.

3. Lemaire B., and P. Dessus. 2003. "Modèles cognitifs issus de l'analyse de la sémantique latente". In Cognito — Cahiers Romans de Sciences Cognitives, In Cognito, INPG, 46 Avenue Felix Viallet, 38031 Grenoble Cedex. 1(1): 55—74. Web: http://www.in-cognito.net/new/index.php

4. Druzhkin, K.Yu. "Readability: onlajn-servis" [Readability]. Web: http://web-corpora.net/wsgi3/readability/index

5. Druzhkin, K.Yu. 2016. "Metriki udobochitaemosti dlya russkogo yazyka" [Readability Metrics for the Russian Language]. Master's Thesis, NRU HSE, Moscow. Web: https://www.hse.ru/edu/vkr/184791276

6. McCarthy, P.M., and S. Jarvis. 2010. "MTLD, vocd-D, and HD-D: A Validation Study of Sophisticated Approaches to Lexical Diversity Assessment". In Behavior Research Methods. 42 (2): 381—392. Print.

7. Lavallée, M., and K. McDonough. 2015. "Comparing the Lexical Features of EAP Students' Essays by Prompt and Rating". TESL Canada Journal. 32 (2): 30—44. Print.

8. Crossley, S.A., T. Cobb, and McNamara D.S. 2013. "Comparing Count-based and Band-based Indices of Word Frequency: Implications for Active Vocabulary Research and Pedagogical Applications". System: 41(4): 965—981. Print.

9. Vongpumivitch, V., J.-Y. Huang, and Chang Y.-C. "Frequency Analysis of the Words in the Academic Word List (AWL) and non-AWL Content Words in Applied Linguistics Research Papers". In English for Specific Purposes. 28 (1): 33—41. Print.

10. Coxhead, A. "A New Academic Word List". TESOL Quarterly. 34 (2): 213—238. Print.

11. Coxhead, A. "The Academic Word List 10 Years on: Research and Teaching Implications". TESOL Quarterly. 45 (2): 355—362.

Bio Note:
Olga Ilinichna Vinogradova is a PhD in Philology, Associate Professor of the School of Linguistics, Research University Higher School if Economics. E-mail: olgavinogr@gmail.com

АВТОМАТИЗИРОВАННАЯ ОЦЕНКА ЛЕКСИКОНА ОБУЧАЮЩИХСЯ ПРИ ПОМОЩИ УЧЕБНОГО КОРПУСА

О.И. Виноградова

Национальный исследовательский университет Высшая школа экономики
Российская Федерация, 101000, Москва, ул. Мясницкая, 20

При обучении академическому английскому языку студентам необходима обратная связь в процессе предоставления письменных работ. Это касается оценки содержания работы, релевантности теме, грамматической согласованности и точности. Все перечисленные критерии, как правило, включены в комментарий инструктора. Обучающимся также необходимы подробные комментарии по улучшению лексических навыков. Такая обратная связь требует существенных усилий и временных затрат со стороны преподавателя. Эффективным инструментов в таком случае может выступить учебный корпус, насчитывающий многочисленные образцы студенческих работ в самых разных жанрах. В настоящей статье представлена информация о разработке системы автоматизированной проверки студенческих работ. На материале жанра эссе мы выделили параметры, которые учитываются экспертами при оценке работ такой направленности. Эти параметры были введены в систему автоматизированной проверки, после чего была произведена корреляция полученных результатов с традиционной системой оценок.

Ключевые слова: учебный корпус, лексическое мастерство, исследование корпусов, отзывы о студенческом эссе

Сведения об авторе:
Виноградова Ольга Ильинична — кандидат филологических наук, доцент Школы лингвистики факультета гуманитарных наук НИУ ВШЭ. E-mail: olgavinogr@gmail.com