AUTOMATIC GENERATION OF LEXICAL EXERCISES

Fenogenova A. (alenush93@gmail.com)
Kuzmenko E. (lizaku77@gmail.com)
National Research University Higher School of Economics, Moscow, Russia

Abstract

The paper presents our approach towards automatic generation of lexical exercises. We describe the corpora and tools used for generation of several types of exercises for English learners, provide examples and evaluate the quality of generated exercises by conducting an experiment in which students complete automatically generated exercises and exercises from coursebooks.

Keywords: language exercises generation, corpus, English as foreign language, lexical exercises

АВТОМАТИЧЕСКОЕ СОЗДАНИЕ ЛЕКСИЧЕСКИХ ТЕСТОВ

Феногенова А. С. (alenush93@gmail.com)
Кузьменко Е. А. (lizaku77@gmail.com)
НИУ ВШЭ, Москва, Россия

Аннотация

Эта статья посвящена автоматическому созданию лексических тестов, то есть упражнений, используемых при изучении лексики иностранного языка. Мы описываем корпуса и инструменты, использованные при создании упражнений и представляем нашу методологию их генерации. Также мы проводим оценку качества полученных упражнений при помощи эксперимента, в котором изучающие язык выполняют сгенерированные нами упражнения и упражнения, взятые из учебников иностранного языка.

Ключевые слова: автоматическое создание упражнений, лексические упражнения, корпусные технологии

AUTOMATIC GENERATION OF LEXICAL EXERCISES[1]

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 (grant № 16-05-0057 Learner corpus REALEC: Lexicological observations) and supported within the framework of a subsidy granted to the HSE by the Government of the Russian Federation for the implementation of the Global Competitiveness Program.

1. Introduction

Language tests play a huge role in the process of learning a foreign language. They help to master new grammatical constructions and to put the recently learned words and collocations into real-world contexts.

At the same time language tests and exercises are expensive to create manually. If instructors make up exercises, there is alway a risk that the exercises will not sound authentic. Therefore, it has become a trend in the second language teaching to construct exercises on the basis of a suitable corpus (chosen by the mode of language, genres, topics and specific characteristics of texts). This adds validity to the exercises, but doesn't facilitate the process of their creation.

The present paper considers the possibility and results of unsupervised generation of exercises. When automatically generated, they are flexible, adaptive, less costly and from the pedagogic perspective have proved to be popular among learners [Sumita, 2005]. More specifically, we elaborate on exercises for collocations. Their importance in language learning is often underestimated, and while students practise grammatical constructions and enrich their vocabulary with single words, collocations are often simply learnt by heart or even go unnoticed [Meunier et Granger, 2008]. This is particularly true for the learners at lower levels of language acquisition, or for specific domains of English. We are dealing with one such domains – English for academic purposes. We aim to tune up the process of exercises generation for academic collocations and to test them among students.

Our exercises generated on the data of the British National Corpus [Leech, 1992] and the British Academic Written English Corpus [Alsop et Nesi, 2009]. We compare the exercises generated on the basis of the two corpora with the ones taken from traditional coursebooks for English learners.

2. Related work

The history of automatically generated exercises can be traced back to 1997, when Roger Levy introduced the notion of computer-assisted language learning. As stated in his work, "Computer Assisted Language learning (CALL) may be defined as the search for and study of application of the computer in language teaching and learning." [Levy, 1997]. Since then the automatic generation of learning content for exercises, presentations, teaching materials and courses has become a widely spread practice. Another established educational trend is blended learning [Graham, 2006], which satisfies the requirement for using online materials as well.

Various systems for automatic language exercise-generation have been developed in recent years. They vary in numerous aspects: 1) language supported, 2) types of exercises, 3) sources of creation, 4) the target aspect of learning (lexical or grammatical).

Most state-of-the-art approaches use various corpora as a source of data. There are a number of papers describing development of exercises from different corpora. In [Wilson, 1997] the advantages and drawbacks of using corpora as a source of generation are outlined and possible improvements of such an approach are proposed. All in all, corpora are used as a general solution in automatic generation of various exercises for different languages [Aldable, 2006; Bick, 2004]. Apart from corpora, the research on exercises generation

involves other sources such as ontologies and thesauruses [Knoop, 2013; Brown, 2005], especially for question generation tasks and vocabulary assessment of the students [Heilman, 2007]. Another method uses the web as the resource of data – there are a lot of systems that are designed on the basis of different word lists, dictionaries and datasets parsed from online web resources. Such resources are compiled together to produce exercises manually according to written rules [Malafeev, 2015]. A relatively novel approach is to apply statistical methods and machine learning in this task. [Goto, 2010] proposes a machine learning method for multiple-choice cloze questions – the system is able to select sentences from input student text based on Preference Learning [Fürnkranz et Hüllermeier, 2011], to estimate blanks by Conditional Random Fields [Lafferty et al. 2011] and to generate distractors based on statistical patterns of existing questions.

The task of automatically generated exercises provides a wide range of possible uses of various NLP methods and techniques. The next section will discuss the methods we use in our system.

3. Materials and methods

3.1. Exercises description

Current study occurs within the research team project of the National Research University Higher School of Economics, and our team's task is to maintain the Russian Error-Annotated Learner English Corpus (REALEC)[2], investigate errors, assess the lexical level of student works, and produce recommendations for them. One of the most efficient ways for students to improve and transform their English writing skills and their proficiency in academic English is by learning collocations. For the purposes of the project special lexical exercises are needed to set up a system of developing lexical skills. Our exercises focus on the ACL (academic collocation lists) [Ackerman, 2013] which can help to increase students' lexical competence.

We have 5 types of exercises:

- **Match collocations.** Two columns with collocation parts in random order are offered to the learner. A student has to match the first part of the collocation with the second one from the given list.
- **Multiple choice.** A student has to fill in the gaps in sentence. There are 4 choices of one of the part of the collocation are given. Only one answer is correct.
- **Open cloze**. A learner has to fill in gaps with suitable whole collocation. No candidates for answers are given.
- **Word bank.** A learner has to fill in the gaps with a suitable collocation. The full list of answer choices is given. No distractors are presented.
- **Word formation.** A student has to fill in the gaps with derivatives of a part of a given collocation.

In our choices we have been inspired by comparable tests from different English exams such as IELTS, FCA, etc. However, most of the exercises there are not lexical but grammatical. At the same time, lexical exercises found in the tests do not focus on the

2 http://realec.org/

collocations training. Therefore, we designed our exercises guided more by our project aims than the conventionally accepted exercises.

3.2. Data

For generating exercises we have used two different English corpora: BNC and BAWE.

1) The British Academic Written English Corpus (BAWE) [Alsop et Nesi, 2009] is an english corpus of academic written texts. The BAWE corpus contains about 6,700,000 tokens in 2761 assessed student writings, ranging in length from 500 words to 5000 words. Texts are evenly distributed across four broad disciplinary areas (Arts and Humanities, Social Sciences, Life Sciences and Physical Sciences) and across four levels of difficulties, altogether thirty five disciplines are represented. The corpus is available online free of charge for academic purposes to researchers who agree to the conditions of use.

2) British National Corpus (BNC) [Leech, 1992] is a 100 million word collection of texts of written and spoken British English from the late twentieth century. Collection includes the extracts from regional and national newspapers, specialist journals for all ages and interests, academic books and popular fiction, school and university essays and a lot of other texts. Corpus is available online and can be downloaded free of charge for academic purposes as well.

3.3. Methodology

The generation of all exercises is based on the corpora mentioned above and the Pearson's Academic collocation list[3]. The whole system of exercises generation is written by means of scripts in the Python programming language.

Firstly, we take a list of collocations and generate for every item from the list its paradigm. That helps us to detect all possible variants of the collocations from the texts in the corpora. Secondly, our program is looking for sentences in the corpus which contain the required collocation. It is possible to configure the length of the sentence and the context (plus/minus one or two sentences around). For specific type of exercises other techniques are implemented.

For the *match a collocation* exercise nothing but the collocation is needed. Word bank and open cloze exercises are generated only with corpora. An interesting type of exercises is *word formation* exercises. It presents a head form of the part of collocation to a student, for him to produce the real form. For this type of exercises we used the wiktionary list[4]. The most sophisticated exercise type is *multiple choice*. In this case we need for generation not only texts with collocations, but also 3 distracting answers apart from the correct one. For this purpose we use a novel word2vec[5] technology. Word2vec is a language model based on neural networks, which computes vector representations of words from a big dataset, in our case from the BNC corpus. A word2vec model is trained to reconstruct linguistic contexts:

3 http://pearsonpte.com/research/academic-collocation-list/

4 https://simple.wiktionary.org/wiki/Wiktionary:BNC_spoken_freq_01

5 https://code.google.com/archive/p/word2vec

the network gets a word and guesses the closest words that occurred in adjacent positions in the input text. Taking as input the word from collocation, the word2vec model returns us several candidates that are likely to occur in the same context as our word. There is a risk of getting completely interchangeable words, but finding the border between too close and too distant words, we find the ideal variants for multiple choice.

Here are the examples of generated exercises (answers are given within the # # signs):

Word formation

- *The children are producing their own spelling dictionary which is #freely# available to the whole class. (free)*
- *The provision of private medicine both within and without the NHS has remained a #controversial# issue. (controversy)*

Multiple choice

Their married life in its essentials will be largely independent of their respective families even when accommodation is shared there will seldom be parental domination such as exists in many societies. The relative #status# and esteem accorded to husband and wife will be roughly equal. Usually each will have some paid employment outside the home and each may have his or her career.

Choices: status, autonomy, competence, identity

Match

universally	*income*
seminal	*accepted*
preceding	*work*
total	*new*
entirely	*work*

Open cloze

- *Environmental and #climatic conditions# have combined with agricultural techniques to produce in Japan exceptionally high yields.*
- *This approach she adopted in all her #subsequent work# thereby introducing a revolutionary style of attack on problems of algebra.*

The problems we face when generating the exercises are as follows:

- inability to induce the right answer from the provided context;
- general complexity of exercises due to the authenticity of texts;
- the possibility of having more than one right answer.

We conducted an experiment to evaluate how well students cope with our exercises.

Word Bank

Choices: renewed interest, dynamic system, stress level, unrelated topic

1. *They could see that nature was not static and unchanging, but that it was a _____ that ever changing.*
2. *For example, excessive noise can raise _____ and also gives the impression of a lack of privacy.*
3. *Science is a very broad field comprising of many varied and seemingly ____ from, zoology to astronomy and geology to medicine.*
4. *The black feminist movement again threw the crisis of African-American masculinity and gender relations into relief, and so inspired _____ in men's studies by the late 1980s.*

We will describe the experiment and its outcomes in the next section.

4. Evaluation of the exercises

In our experiment we aimed to compare the exercises generated by our program with the exercises manually compiled by language instructors. To do this, we have made three sets of exercises:

- exercises generated on the BNC data;
- exercises generated on the BAWE data;
- exercises taken from English coursebooks.

In order to make the tasks as similar as possible, we have tested only one type of exercises, namely, multiple choice. The reason for this is that in coursebooks there were few exercises of other types that resembled ours.

These three sets of tests were given to 21 students studying in the 2nd year of the program “Fundamental and Applied Linguistics” at National Research University Higher School of Economics. Each set consisted of 5 variants. The students could indicate which questions seemed ambiguous to them (the task is unclear, or there is no right answer, or the context does not provide enough information). This parameter of evaluating exercises is subjective and not very reliable, but still it helps to identify the quality of exercises. Table 1 shows the percentage of correct answers per person in each set and the number of ambiguous questions found in each variant. The figures were normalized by variants because the number of variants passed in each set of exercises was uneven in our experiment.

Table 1. Results of passing the exercises of type “multiple choice” in each set.

	Coursebooks	**BAWE**	**BNC**
Correct answers, %	73.6	62.5	46.01
Ambiguous questions, %	0	21	45

As we can see, the original exercises taken from coursebooks seem to be more appropriate for learners. They raise no questions at all, and the percentage of correct answers is high. At the same time, the exercises generated with the BAWE corpus are very promising. The percentage of correct answers is still high, and ambiguous questions are not very frequent in the generated tests. The poorest performance was found in the tests generated with the BNC corpus. The participants answered correctly less than half of the exercises, and another half was considered ambiguous.

Below in the Table 2 the percentage of errors made by students in different types of exercises is shown. This table shows results from not all students, but from those students who have answered all questions in each section.

Table 2. Percentage of errors made by students.

Student №	Original exercises	Exercises build on BAWE	Exercises built on BNC
1	0.11	0.14	0.27
2	0.45	0.57	0.63
3	0.31	0.35	0.5
4	0.15	0.28	0.63
5	0.20	0.35	0.45
6	0.09	0.21	0.09
7	0.045	0.14	0.16
8	0.13	0.42	0.18
9	0.5	0.5	0.9
10	0.4	0.5	0.7
11	0.72	0.5	0.45
12	0.31	0.35	0.54
13	0.13	0.21	0.27
14	0.22	0.21	0.54

We can see that most students perform similarly on exercises from different sets - if the result of a student is bad for classical exercises, he will also perform badly on the generated exercises, and vice versa. This pattern can also be proved by finding the degree of correlation between the percentage of errors in different exercises types. Correlation between tests generated from BAWE and original tests was found to be equal to 0.8, and correlation between tests generated from BNC and original tests is equal to 0.65. Thus, there is no significant deterioration of the students' performance on different types of exercises.

Our experiment shows that the quality of generated exercises is heavily dependent on the corpora used for their creation. At the same time, the computational approach towards the creation of exercises still can be adopted in any learning environment. In our experiment exercises generated with the BAWE corpus were not significantly worse than the manually compiled exercises while they are more flexible and easier to generate.

5. Conclusion

In our work we presented our approach towards automatic generation of lexical exercises. The data for exercises was taken from the British National Corpus and the British

Academic Writing Corpus. Exercises of several types were developed: multiple choice, match, word formation, open cloze.

We evaluated the quality of the generated exercises by comparing them to the exercises from coursebooks for learners of English compiled manually by language instructors. Only one type of exercises was evaluated, namely, multiple choice. The comparison included how well language learners perform while completing original and generated exercises. This experiment showed that the quality of exercises can be different depending on the corpus used for their generation. However, the automatically generated exercises are found to be comparable in quality to the ones published in textbooks.

The usage of generated exercises in the classroom raises many issues, which can be resolved in subsequent work. In particular, other types of exercises should be evaluated, and we are going to experiment with written English corpora to find out what data is more suitable for learners. Also we intend to improve the quality of exercises by identifying and eliminating ambiguous questions. These are our goals for future work.

Acknowledgments

We express our gratitude to Olga Vinogradova for commenting on our generated exercises and helping us with the organization of the experiment.

References

1. Aldabe I. et al. Arikiturri: an automatic question generator based on corpora and nlp techniques //Intelligent Tutoring Systems. – Springer Berlin Heidelberg, 2006. – С. 584-594.
2. Alsop, S., & Nesi, H. (2009). Issues in the development of the British Academic Written English (BAWE) corpus. *Corpora*, *4*(1), 71-83.
3. Ackermann K., Chen Y. H. Developing the Academic Collocation List (ACL)–A corpus-driven and expert-judged approach //Journal of English for Academic Purposes. – 2013. – Т. 12. – №. 4. – С. 235-247.
4. Bick E. Live use of Corpus data and Corpus annotation tools in CALL: Some new developments in VISL //H. Holmboe (red.), Nordic Language Technology, Årbog for Nordisk Sprogteknologisk Forskningsprogram. – 2000. – Т. 2004. – С. 171-186.
5. Brown J. C., Frishkoff G. A., Eskenazi M. Automatic question generation for vocabulary assessment //Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing. – Association for Computational Linguistics, 2005. – С. 819-826.
6. Fürnkranz, J., & Hüllermeier, E. (2011). *Preference learning* (pp. 789-795). Springer US.
7. Goto T. et al. Automatic generation system of multiple-choice cloze questions and its evaluation //Knowledge Management & E-Learning: An International Journal (KM&EL). – 2010. – Т. 2. – №. 3. – С. 210-224.
8. Graham C. R. Blended learning systems //CJ Bonk & CR Graham, The handbook of blended learning: Global perspectives, local designs. Pfeiffer. – 2006.
9. Heilman M., Eskenazi M. Application of automatic thesaurus extraction for computer generation of vocabulary questions //SLaTE. – 2007. – С. 65-68.
10. Knoop S., Wilske S. WordGap - Automatic generation of gap-filling vocabulary exercises for mobile learning //Proceedings of Second Workshop NLP Computer-Assisted Language Learning at NODALIDA. – 2013. – С. 39-47.

11. Meunier, F., & Granger, S. (Eds.). (2008). *Phraseology in foreign language learning and teaching*. John Benjamins Publishing.

12. Lafferty, J., McCallum, A., & Pereira, F. C. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data.

13. Leech, G. (1992). 100 million words of English: the British National Corpus (BNC). *Language Research*, *28*(1), 1-13.

14. Levy M. Computer-assisted language learning: Context and conceptualization. – Oxford University Press, 1997.

15. Malafeev A. Y. Exercise Maker: Automatic Language Exercise Generation, in: *Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue" (2015)* Issue 14(21). M.: Russian State University for the Humanitie, 2015. P. 441-452.

16. Sumita, E., Sugaya,F., and Yamamoto, S. (2005). Measuring Non-native Speakers' Proficiency of English by Using a Test with Automatically-generated Fill-in-the-blank Questions. In Proceedings of the Second Workshop on Building Educational Applications Using NLP, (Stroudsburg, PA, USA: Association for Computational Linguistics), pp. 61–68.

17. Wilson E. The automatic generation of CALL exercises from general corpora //Teaching and language corpora. – 1997. – C. 1-23.