

# Automatic Generation of Text-Based Open Cloze Exercises

Alexey Malafeev<sup>(✉)</sup>

National Research University Higher School of Economics,  
Nizhny Novgorod, Russia  
aumalafeev@hse.ru

**Abstract.** This article presents an approach to the automatic generation of open cloze exercises that are based on real-life English texts. The exercise format is similar to the open cloze test used in Cambridge certificate exams (FCE, CAE, CPE). Two experiments were conducted to evaluate the usefulness on the machine-generated exercises and compare them with authentic Cambridge tests. The experiments showed that the generation method used was quite effective. With some customization, the presented method can be applied to generating similar exercises based on texts written in other languages.

**Keywords:** Exercise generation · Open cloze · Language exercises · Computer-assisted language learning (CALL) · English as a foreign language (EFL)

## 1 Introduction

Vocabulary and grammar exercises are widely used in teaching English as a foreign language (TEFL), but creating them manually is time-consuming and expensive. In response to this, many methods for the automated generation of language exercises have been proposed in the past two decades. These solutions rely on various NLP tools and techniques and can produce different types of exercises.

This paper presents an effective approach to the generation of text-based open cloze exercises similar to those used in Cambridge certificate exams (FCE, CAE and CPE). The method used is simple and does not rely on any sophisticated NLP tools, yet it is powerful enough to generate realistic and useful exercises. In fact, as shown in the evaluation section of this paper, experienced EFL instructors find it somewhat difficult to tell the difference between exercises generated with the help of the method presented here and authentic Cambridge tests.

In the most general sense, the cloze is a test of language ability or reading comprehension, which is created by removing certain words from a text. The gaps are to be filled in with appropriate words. In the open cloze, the test-taker is to guess the suitable words from the context, without seeing any multiple choice options. It may therefore be a challenging task, requiring a deep understanding of language structure [1, 2].

Although the open cloze may be of different varieties [2], the approach presented in this paper is aimed at the generation of exercises<sup>1</sup> emulating the open cloze test used in Cambridge certificate exams (FCE, CAE and CPE). In this test, “[t]he focus of the gapped words is either grammatical, such as articles, auxiliaries, prepositions, pronouns, verb tenses and forms; or lexicogrammatical, such as phrasal verbs, linkers and words within fixed phrases” [1].

The reasons for choosing the Cambridge open cloze as the target exercise type are the following:

1. Cambridge certificate exams are well-established and highly regarded, and they tend to emphasize a close relationship between teaching and testing [3]. Open cloze exercises are a useful stimulus in integrated reading, writing and vocabulary instruction [2].
2. This exercise type largely focuses on using function words in English, an analytic language. These may be difficult for learners to master and often require extensive practice. This is especially true for learners whose mother tongue (L1) differs from English in its structure. For example, Russian learners typically struggle with English prepositions, auxiliary verbs and articles, which either work very differently or are not altogether found in their L1. The open cloze may therefore be a helpful tool for practicing the use of function words.

The method discussed in this paper is part of a larger system developed by the author, called ELEM (English Language Exercise Maker), which is aimed at generating English vocabulary and grammar exercises of various types based on real-life texts. Being able to use arbitrary texts (e.g. news articles, blog entries, film reviews etc.) for generating exercises gives the user a lot of freedom in choosing interesting and relevant material. Research has shown that learners’ motivation can be improved by tailoring texts to their interests [4].

This paper is structured as follows. Section 2 discusses related work. Section 3 describes the method of generating open cloze exercises used in ELEM. Section 4 reports on an evaluation of the exercises generated. Section 5 concludes and outlines future work.

## 2 Related Work

Some recent research has been conducted with a view to facilitating exercise creation. Among the more general solutions are multi-domain exercise or test generation systems, e.g. [5–7], as well as authoring tools, e.g. Hot Potatoes,<sup>2</sup> MaxAuthor<sup>3</sup> and others.

There are also several systems that are designed for generating exercises of one or more specific types to aid learners of the supported language(s). Exercise types include

---

<sup>1</sup> Or tests, depending on whether they are used for practice or assessment; for simplicity, they will be referred to as exercises.

<sup>2</sup> <http://hotpot.uvic.ca/>

<sup>3</sup> <http://cali.arizona.edu/docs/wmaxa/>

reading comprehension questions [8], morphological transformation [9–13] error correction [10], finding related words [14], shuffle questions (putting words in the right order to form a correct sentence, adding appropriate inflections and/or function words) [15], translation [16], grammatical or lexico-grammatical open cloze (fill-in-the-blank without multiple choice) [12, 13, 15], and, probably the most common, multiple choice questions or cloze tests [10, 13, 17–20]. All these are language exercises, unlike, for example, open cloze tests checking students’ factual knowledge [21].

As mentioned before, this article will focus on a specific exercise type, the open cloze that tests the learner’s language proficiency. The majority of the solutions listed above concentrate on other exercise types, with the exception of KillerFiller, an exercise-building tool that is part of the VISL project<sup>4</sup> [12], the gap-filling activity maker in the VIEW project<sup>5</sup> [13], and GramEx [15]. It is important to outline the principal differences between these systems and the method presented in this paper.

KillerFiller supports multiple languages including English. It extracts sentences from annotated corpora, replacing words of a given part of speech with blanks that the user has to fill in. At the time of writing, five word classes are available: verbs, adjectives, prepositions, adverbs and nouns. Open cloze exercises are only generated if the user chooses prepositions; in all other cases the lemma of the target word is given next to the blank, in which case the focus of the task is on morphological transformation. When prepositions are chosen as the target word class, they are simply removed from the sentence and replaced with gaps. Often, there is arguably not enough context for guessing correctly.

VIEW is an ICALL tool for enhancing authentic web pages in English and some other languages. It uses a blend of state-of-the-art NLP techniques: tokenization, lemmatization, morphological analysis, part-of-speech tagging, chunking, and parsing. VIEW can transform authentic web pages into language exercises. One of the supported exercise types is a fill-in-the-blank activity, where the user first selects the language phenomenon to practice (e.g. articles, determiners, gerunds vs. infinitives etc.). After providing a URL, the user proceeds to the enhanced version of the web page, where he/she is to fill the blanks made by VIEW. The system removes some of the words representing the target language phenomenon that are found on the page. The resulting activity is effectively an open cloze if the user has chosen one of the following: articles, determiners, phrasal verbs or prepositions.

GramEx is a framework for generating grammar exercises in French that are sentence-based. Unlike in KillerFiller, however, the sentences are not extracted from a corpus, but generated automatically in a strictly controlled way. Like in the two systems discussed above, the user first specifies the target word category or a more specific language phenomenon such as ‘adjectives that precede the noun’. Open cloze exercises are generated if either prepositions or articles are selected. The resulting sentences are very simple, e.g. “She loves small armadillos.” This may be seen as an advantage for

<sup>4</sup> Visual Interactive Syntax Learning, <http://beta.visl.sdu.dk/>.

<sup>5</sup> Visual Input Enhancement of the Web (<http://sifnos.sfs.uni-tuebingen.de/VIEW/>), an extension of WERTi, or Working with English Real Texts interactively (<http://sifnos.sfs.uni-tuebingen.de/WERTi/>).

elementary or pre-intermediate learners of French; on the other hand, higher-level learners might find the exercises too repetitive and not challenging enough.

The approach presented in this article is different in that:

1. It is developed specifically for English, and is quite language-specific. However, similar principles may be applied to the generation of this type of open cloze exercises in other languages.
2. The proposed method targets a specific exam format, and may thus be helpful in preparation for taking FCE, CAE or CPE. For many learners, the open cloze may be one of the more challenging tasks in these exams.
3. The input for the exercises is plain text files. Thus, the difficulty level of the exercise does not depend solely on the system settings (e.g. how many gaps to make or which words to remove), but also on the complexity/readability of the source text. Tools like Lexile [22] can be used to evaluate texts before generating exercises.
4. The proposed exercises are focused on words of various parts of speech, not only prepositions or determiners. Among other word classes used are conjunctions, pronouns, auxiliary and modal verbs, adverbs and particles. Furthermore, these word classes are not practiced in isolation from one another, but rather in one combined activity. This makes the task more challenging because of the larger number of gap-filling options available to the learner.
5. Although the proposed method is quite effective in generating open cloze exercises, it does not require sophisticated NLP tools such as annotated corpora (VISL), a tree adjoining grammar, syntactic and morpho-syntactic lexicons (GramEx), or even morphological analysis and part-of-speech tagging (VIEW). Therefore, it should be relatively easy to apply the method to other languages. Moreover, some NLP tools may simply not be readily available for some languages, although this is obviously not the case with English.

There is also a web-based cloze generator<sup>6</sup> that is somewhat similar to the proposed solution. It accepts arbitrary English texts as input and produces open cloze tests. However, much like in KillerFiller, VIEW and GramEx, it removes all words of the part of speech specified by the user, and the gaps may be very close to each other, e.g. with prepositions: *Michael Hussey picked it \_\_\_\_\_ and swept that \_\_\_\_\_ the gap* (up from outside off; into). Although the method itself is similar (using a hardcoded list to replace certain words in the text with gaps) the resulting exercises are very different from Cambridge open cloze tests in that:

- There is always only one target word category per task;
- The same word forms can be removed from the text multiple times;
- Gaps can be too close to each other, which sometimes makes it hard, if not impossible, to restore the words.

To sum up, the presented method differs from most existing work in that it efficiently generates open cloze exercises emulating those used in Cambridge certificate

---

<sup>6</sup> <http://l.georges.online.fr/tools/cloze.html>

exams, even though it does not rely on advanced NLP tools. The next section will describe the proposed method in more detail.

### 3 Generation of Open Cloze Exercises

In the target type of exercise, there is a text of about 200–220 words with a certain number of gaps<sup>7</sup> placed at irregular intervals. Each gap is to be filled in with a single word. Consider the following example: “It is not unusual for objects only about a metre or \_\_\_\_\_ away to become unrecognizable”. The blank in this sentence can be filled in with *so*, *less* or *two*. In the exam, it is enough for the candidate to give just one correct answer [1].

An empirical study of 29 FCE, CAE and CPE open cloze tests (408 gaps total) was carried out to determine what words can be removed. According to the answer keys, 198 unique word forms were accepted as correct fillers of the gaps. While a more complete list could be compiled from a larger sample of tests, the one obtained seemed representative enough. Most words on the list were function words. In fact, although forms of content words comprised about 40 % of the list, they were only used in 8 % of the gaps. This means that, roughly, in every nine out of ten cases, the test taker was to make a decision about which function word to use.

Given the dominance of function words in the chosen exercise type, it seemed possible to use a very straightforward approach to generation. It relies on a predefined set of specific word forms, rather than all words belonging to a particular class (such as prepositions) or words used in certain high-frequency combinations (collocation-based approaches). It is presumed that the target word forms can be ‘safely’ removed from almost any sentence in a given text. However, enough context should be provided so that the user would be able to fill in the gap with the missing word or other suitable words. Two research questions were raised:

1. Is it possible to automatically generate useful Cambridge-like open cloze exercises from English texts by relying on a static list of target word forms?
2. How similar would the resulting exercises be to open cloze tests used in Cambridge certificate exams?

Clearly, a robust list of target word forms becomes crucial for this approach. These are some key characteristics of the list:

- It should be large enough to ensure sufficient variety and difficulty of the generated exercises.
- The list should mostly contain function words: articles, prepositions, conjunctions, pronouns, particles and auxiliaries. A handful of modal verb forms can be confidently used too, provided these forms have no high-frequency homonyms in the English language. This is important because no part-of-speech disambiguation is used in the presented approach. For example, ‘can’ is not a very good candidate,

---

<sup>7</sup> 12 gaps in the FCE cloze and 15 in both CAE and CPE.

because there is a rather high-frequency homonymous noun in English. ‘Could’, on the other hand, is perfectly suitable.

- The list should not contain any nouns, adjectives or verbs (except for some forms of ‘be’, ‘do’ and ‘have’ and the already mentioned modal verb forms). A limited number of adverbs can be used, especially those that are known to be frequently misused by many learners of English.

After several months of extensive testing with EFL students at various levels of proficiency in English, 81 word forms (mainly content words) were removed and 29 (mainly function words) added to the list of 198 unique word forms obtained from the empirical study. The main criteria were word frequency, part of speech and ‘restorability’. The resulting 146 forms are listed below:

*a about above after again against ago all although am an and another any anybody anyone anything anywhere apart are around as at away back be because been before behind being below besides between but by could despite did do does doing done down during each either enough every everybody everyone everything few for from had hardly has have having how however if in into is it its itself just least less many more most much never no nobody none nor not nothing of off on one only onto or other others ought out over rather regardless same scarcely should since so some somebody someone something somewhere such than that the then there therefore these this those though through throughout till to too under until up was were what whatever when where whereas whether which while whilst who whose why will with within without would yet*

The words on this list are known to cause learners significant difficulty, though in varying degrees. At the same time, these word forms are restorable if removed from most sentences. Thus, they can be used to make interesting and relevant gaps, i.e. ones that are neither too obvious nor too ambiguous.

The list is heterogeneous enough to provide for sufficient variety, and has mostly high-frequency words, which means that virtually any text can be used as input for exercise generation. For example, in the previous 29-word sentence thirteen of the words can potentially be removed, although, naturally, not at the same time:

*(1) The list (2) is heterogeneous (3) enough (4) to provide (5) for sufficient variety, (6) and (7) has mostly high-frequency words, (8) which means (9) that virtually (10) any text can (11) be used (12) as input (13) for exercise generation.*

Replacing any word of the thirteen with a gap, provided there is enough context, would result in an arguably useful open cloze question.

The exercise generation script was written in Python. It works as follows: the source text is normalized and tokenized. All words in the text that are on the above list are marked. Words of at least two characters and written in all capital letters are skipped to filter out abbreviations. A number of marked words are removed from the text at random. Note that each unique word form on the above list has an equal chance to be chosen and can only be used once per text. This is done to ensure that highest-frequency words such as articles do not dominate the exercise. Also, this is important for emulating Cambridge tests, where all gapped words in a single cloze tend to be different.

The removed words are replaced with blanks. The number of words to remove is specified at the beginning of generation. The minimum distance between the gaps is

three words, irrespective of punctuation. Thus each gap has a minimum context of at least six words, which was thought to be sufficient to restore the word in most cases.

Somewhat counterintuitively, this simple and straightforward approach to the generation of open cloze exercises yielded good results. The evaluation methods used and results achieved are described in the next section.

## 4 Evaluation

Exercises generated with the proposed method were extensively used by language instructors with students at different levels. It was noticed that such exercises can accommodate any proficiency level, from beginner to advanced. This is largely due to the fact that the difficulty of the given exercise depends crucially on the complexity of the input text.

The final version of the solution was evaluated more formally. Two experiments were conducted to test the quality of the generated exercises and answer the research questions. The participants were expert teachers of English, all non-native speakers.

### 4.1 First Experiment

The aim of the first experiment was to evaluate the usefulness of individual gaps. A random sample of ten article extracts was taken from `cnn.com` and `guardian.com`. The size of each extract was about 300 words. These texts had not been previously used for exercise generation. One open cloze exercise with 12 gaps was generated from each of the sample texts using the method described in Sect. 3. The exercises were not post-edited or modified in any way.

Two expert EFL instructors with a background in preparing language students for taking Cambridge certificate exams were asked to assess the gaps in the exercises (the answer key was also provided). The instructors were to answer two questions about each of the 120 gaps:

1. Can the removed word be restored from the context?
2. Is the gap useful for teaching intermediate learners any relevant aspects of English grammar or collocation?

**Table 1.** Expert evaluation of the generated open cloze exercises

	Restorable gaps		Useful gaps	
	n	%	n	%
Expert 1	118	98 %	110	92 %
Expert 2	118	98 %	104	87 %
Agreement	118	98 %	106	88 %

**Discussion.** As can be seen from Table 1, the experts considered almost all of the removed words to be restorable from the context. As for the usefulness of the gaps, the results were predictably worse, but still rather good – about 90 %. The experts commented that most of the gaps considered ‘not useful’ seemed too obvious for the target proficiency level.

It seems that for the type of exercise under consideration, there might be an inverse relationship between the frequency of the removed word and the difficulty of the gap. However, further investigation might be necessary to confirm this. In any case, better results could probably be achieved by using multiple lists of removable word forms tailored to different target proficiency levels. Lists targeting lower level students should contain higher frequency word forms, and vice versa.

## 4.2 Second Experiment

The second experiment was aimed at comparing the generated exercises and actual Cambridge tests. Two open cloze tests from the CAE examinations in 2008 were randomly selected; these had not been previously used for compiling the list of target words. The original texts were restored by filling the gaps back in with the help of the answer keys. When two or more answers for the same gap were given in the key, the first alternative was always chosen. Open cloze exercises with 15 gaps (the same number as in the original tests) were automatically generated from the restored texts. Only one exercise variant was generated for each of the texts. As in the previous experiment, the exercises were not post-edited.

The two machine-generated and two original activities were given to two groups of EFL instructors (17 and 16 people). All instructors received identical forms with two pairs of exercises based on the two source texts. No answer keys were provided; however, most of the gaps could be easily filled in by comparing the exercises, as the majority of the gaps did not overlap.<sup>8</sup> The experts were told that in each pair of activities, one was machine-generated and one was an actual Cambridge test. The text formatting was identical, including the numbering of the gaps.<sup>9</sup> The experts were instructed to answer the same set of questions for both pairs. First, they were asked to identify the machine-generated exercise, if possible. If they felt they could do it, they were to give their answer and mark their confidence level on a scale of 0 (not sure) to 3 (very certain). Optionally, the experts were asked to specify the criteria they used to identify the machine-generated exercise.

There was no time limit for answering the questions. With the first group of 17 experts, the experiment was conducted in a university classroom. The longest it took for the experts in this group to give their answers was about 45 min. With the second group of 16 experts, the experiment was carried out by email. The results of the experiment are presented in Tables 2 and 3.

---

<sup>8</sup> In the first pair, 5 out of 15 gaps coincided (33 %); in the second, 2 out of 15 (13 %).

<sup>9</sup> In the CAE exam, the numbering of the open cloze gaps starts from 13.



**Table 2.** Telling the difference between automatically generated exercises and authentic tests (individual answers)

Expert group 1					Expert group 2				
n	Answ.1 <sup>a</sup>	Confid.1	Answ.2	Confid.2	n	Answ.1	Confid.1	Answ.2	Confid.2
1	R	2	R	n/a	1	R	3	R	1
2	R	1	W	1	2	R	2	R	2
3	n/a	n/a	R	1	3	W	1	R	1
4	R	1	R	1	4	R	1	R	1
5	R	1	W	1	5	R	2	R	2
6	n/a	0	R	1	6	W	1	R	1
7	R	1	R	1	7	R	2	W	1
8	R	n/a	W	n/a	8	W	1	W	2
9	W	2	R	2	9	R	1	R	1
10	R	2	R	2	10	R	2	W	1
11	R	2	R	1	11	R	1	R	1
12	n/a	n/a	n/a	n/a	12	R	1	n/a	n/a
13	W	n/a	W	1	13	R	1	R	1
14	W	2	R	2	14	R	1	n/a	n/a
15	n/a	0	n/a	0	15	R	2	W	2
16	W	1	R	1	16	R	1	R	2
17	W	1	n/a	n/a					

<sup>a</sup> R and W stand for identifying the machine-generated exercise correctly and incorrectly, respectively.

**Table 3.** Telling the difference between automatically generated exercises and authentic tests (aggregated results)

	Correct answers	Incorrect answers	No answer	Correct, confid. > = 2	Both correct	Both correct, confid. > = 2
n	41/66	16/66	9/66	15/66	13/33	3/33
%	62 %	24 %	14 %	23 %	39 %	9 %

**Discussion.** The results of the experiment show that the machine-generated exercises look quite similar to the authentic Cambridge tests. Granted, this does not imply that they are similarly useful, but it is interesting that experienced EFL instructors had considerable difficulty in differentiating between the two. Indeed, only one of the participants chose the highest confidence level for his answer (which was correct). As can be seen from Table 3, although over 60 % of the answers were correct, the instructors were not absolutely certain about their choices. Only 15 of the 66 answers were given with a confidence of 2 or higher. Remarkably, only 13 of the 33 participants were able to correctly identify both of the machine-generated exercises, and only three of them were ‘almost sure’ of both of their answers (confidence level 2); the other ten were less confident.

As for the optional question about the criteria used, it was answered by the experts in 46 out of the 57 cases (81 %) when one of the two activities was marked as machine-generated. However, no single criterion was consistently used for correct identification. Although the criteria used for giving 18 out of 41 correct answers (44 %) were related to the usefulness of the exercises, the same or very similar criteria were sometimes applied when making incorrect choices (6 out of 16, or 38 %). Both this fact and the low confidence levels seem to indicate that the machine-generated open cloze exercises do not look obviously less useful than the authentic ones. Admittedly, it might be necessary to perform further experiments with more exercise pairs to confirm this.

## 5 Conclusion and Future Work

In this paper, a simple method for generating open cloze exercises was presented. The exercises are text-based and are intended to emulate the open cloze tests used in Cambridge certificate exams (FCE, CAE and CPE). The machine-generated open cloze exercises are used in the ELEM system together with other types of language exercises focused on error correction, word formation, using verb forms etc.

The presented method relies on a list of carefully selected word forms that seem to be restorable from most contexts. Although the method does not use any sophisticated NLP tools, it seems sufficiently reliable and efficient at generating exercises of the target type, based on the evaluation described in the previous section. Presumably, the method can be used for other languages, although it might yield better results for analytic languages.

The solution described in this paper could benefit from added functionality such as checking the user's answers automatically – not simply by comparing them to the words removed from the source text, but rather by tapping into word co-occurrence data from corpora. This would make it possible to check if the word given as the answer could actually be appropriate for the context even though a different word was originally used in the source text.

Another way to continue this work could be experimenting with high-frequency content words that form numerous collocations (some examples are: 'get', 'come', 'time', etc.). Occasionally using some content words might possibly make machine-generated exercises more similar to authentic Cambridge open cloze tests. Also, it might be productive to use several different word lists to generate exercises of varying difficulty.

## References

1. Cambridge English: Advanced Handbook for Teachers (2012)
2. Lee, S.H.: Beyond reading and proficiency assessment: The rational cloze procedure as stimulus for integrated reading, writing, and vocabulary instruction and teacher-student interaction in ESL. *System* **36**, 642–660 (2008)
3. Chalhoub-Deville, M., Turner, C.E.: What to look for in ESL admission tests: Cambridge certificate exams, IELTS, and TOEFL. *System* **28**, 523–539 (2000)

4. Heilman, M., Collins-Thompson, K., Callan, J., Eskenazi, M., Juffs, A., Wilson, L.: Personalization of reading passages improves vocabulary acquisition. *Int. J. Artif. Intell. Educ.* **20**, 73–98 (2010)
5. Sonntag, M.: Exercise generation by group models for autonomous web-based learning. In: 35th Euromicro Conference on Software Engineering and Advanced Applications, SEAA '09, pp. 57–63 (2009)
6. Almeida, J.J., Araujo, I., Brito, I., Carvalho, N., Machado, G.J., Pereira, R.M.S., Smirnov, G.: PASSAROLA: High-order exercise generation system. In: 2013 8th Iberian Conference on Information Systems and Technologies (CISTI), pp. 1–5 (2013)
7. Mitkov, R., Ha, A.L., Karamanis, N.: A computer-aided environment for generating multiple-choice test items. *Nat. Lang. Eng.* **12**, 177–194 (2006)
8. Gates, D.M.: Automatically generating reading comprehension look-back strategy: Questions from expository texts (2008)
9. Antonsen, L., Johnson, R., Trondsterud, T., Uibo, H.: Generating modular grammar exercises with finite-state transducers. In: Proceedings of Second Workshop NLP Computer-Assisted Language Learning, NODALIDA 2013, pp. 27–38 (2013)
10. Aldabe, I., de Lacalle, M.L., Maritxalar, M., Martinez, E., Uria, L.: ArikIturri: An automatic question generator based on corpora and NLP techniques. In: Ikeda, M., Ashley, K.D., Chan, T.-W. (eds.) ITS 2006. LNCS, vol. 4053, pp. 584–594. Springer, Heidelberg (2006)
11. Dickinson, M., Herring, J.: Developing online ICALL exercises for Russian. In: Proceedings of the Third Workshop on Innovative Use of NLP for Building Educational Applications, pp. 1–9. Association for Computational Linguistics, Stroudsburg, PA, USA (2008)
12. Bick, E.: Live use of corpus data and corpus annotation tools in CALL: Some new developments in VISL. *Nord. Lang. Technol. Aarb. Nord. Sprogteknologisk Forskningsprogram*, pp. 171–185 (2005)
13. Meurers, D., Ziai, R., Amaral, L., Boyd, A., Dimitrov, A., Metcalf, V., Ott, N.: Enhancing authentic web pages for language learners. In: Proceedings of the NAACL HLT 2010 Fifth Workshop on Innovative Use of NLP for Building Educational Applications, pp. 10–18. Association for Computational Linguistics, Stroudsburg, PA, USA (2010)
14. Heilman, M., Eskenazi, M.: Application of automatic thesaurus extraction for computer generation of vocabulary questions. In: Proceedings of the SLaTE Workshop on Speech and Language Technology in Education, pp. 65–68 (2007)
15. Perez-Beltrachini, L., Gardent, C., Kruszewski, G.: Generating grammar exercises. In: Proceedings of the Seventh Workshop on Building Educational Applications Using NLP, pp. 147–156. Association for Computational Linguistics, Stroudsburg, PA, USA (2012)
16. Burstein, J., Marcu, D.: Translation exercise assistant: Automated generation of translation exercises for native-arabic speakers learning english. In: Proceedings of HLT/EMNLP on Interactive Demonstrations, pp. 16–17. Association for Computational Linguistics, Stroudsburg, PA, USA (2005)
17. Sumita, E., Sugaya, F., Yamamoto, S.: 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, pp. 61–68. Association for Computational Linguistics, Stroudsburg, PA, USA (2005)
18. Hoshino, A., Nakagawa, H.: WebExperimenter for multiple-choice question generation. In: Proceedings of HLT/EMNLP on Interactive Demonstrations, pp. 18–19. Association for Computational Linguistics, Stroudsburg, PA, USA (2005)
19. Goto, T., Kojiri, T., Watanabe, T., Iwata, T., Yamada, T.: Automatic generation system of multiple-choice cloze questions and its evaluation. *Knowl. Manag. E-Learn. Int. J. KMEL.* **2**, 210–224 (2010)

20. Knoop, S., Wilske, S.: WordGap - Automatic generation of gap-filling vocabulary exercises for mobile learning. In: Proceedings of Second Workshop NLP Computer-Assisted Language Learning at NODALIDA 2013, pp. 39–47 (2013)
21. Kurtasov, A.: A System for generating cloze test items from Russian-language text. In: Proceedings of the Student Research Workshop Associated with RANLP, pp. 107–112 (2013)
22. Stenner, A.J.: Measuring reading comprehension with the lexile framework. In: Fourth North American Conference on Adolescent/Adult Literacy. Washington D.C. (1996)