

Personalized Exercises for Preposition Learning

John Lee, Mengqi Luo

The Halliday Centre for Intelligent Applications of Language Studies

Department of Linguistics and Translation

City University of Hong Kong

{jsylee, mengqluo}@cityu.edu.hk

Abstract

We present a computer-assisted language learning (CALL) system that generates fill-in-the-blank items for preposition usage. The system takes a set of carrier sentences as input, chooses a preposition in each sentence as the key, and then automatically generates distractors. It personalizes item selection for the user in two ways. First, it logs items to which the user previously gave incorrect answers, and offers similar items in a future session as review. Second, it progresses from easier to harder sentences, to minimize any hindrance on preposition learning that might be posed by difficult vocabulary.

1 Introduction

Many learners of English find it challenging to master the use of prepositions. Preposition usage is a frequent error category in various learner corpora (Izumi et al., 2003; Dahlmeier et al., 2013; Lee et al., 2015); indeed, entire exercise books have been devoted to training learners on preposition usage (Wacyn-Jones and Allsop, 2000; Yates, 2010). To address this area of difficulty, we present a system that automatically generates fill-in-the-blank (FIB) preposition items with multiple choices.

Also known as gap-fill or cloze items, FIB items are a common form of exercise in computer-assisted language learning (CALL) applications. Table 1 shows an example item designed for teaching English preposition usage. It contains a sentence, “The objective is to kick the ball into the opponent’s goal”, with the preposition “into” blanked out; this sentence serves as the *stem* (or *carrier sentence*). It is followed by four choices for the blank, one of which is the *key* (i.e., the

correct answer), and the other three are distractors. These choices enable the CALL application to provide immediate and objective feedback to the learner.

Traditional exercise books no longer meet all the needs of today’s learners. The pedagogical benefits of using authentic textual material have been well documented (Larimer and Schleicher, 1999; Erbaggio et al., 2012). One recent approach turns text on web pages into slot-fill items (Meurers et al., 2010). By offering the learner the freedom to choose his or her own preferred text, this approach motivates the learner to complete the exercises.

Our system automatically constructs FIB preposition items from sentences in Wikipedia, a corpus that contains authentic language. As more users own mobile devices, mobile applications are now among the most efficient ways to provide on-demand language learning services. Although user attention on mobile devices can be brief and sporadic, each FIB item can be completed within a short time, and therefore our system offers an educational option for users to spend their idle moments. Focusing on prepositions, the system generates distractors based on error statistics compiled from learner corpora. Further, it maintains an estimate of the user’s vocabulary level, and tai-

The objective is to kick the ball _____ the opponent’s goal. (A) in (B) into (C) to (D) with
--

Table 1: An automatically generated fill-in-the-blank item, where “into” is the key, and the other three choices are distractors.

lors item selection to address his or her areas of weakness. To the best of our knowledge, this is the first system that offers these personalization features for preposition items.

The rest of the paper is organized as follows. Section 2 reviews previous work. Section 3 outlines the algorithms for generating the fill-in-the-blank items. Section 4 gives details about the personalization features in the item selection process. Section 5 reports implementation details and evaluation results.

2 Previous work

The Internet presents the language learner with an embarrassment of riches. A plethora of CALL websites—*Duolingo*, *LearnEnglish Grammar* by the British Council, or *Rosetta Stone*, to name just a few—provide a variety of speaking, listening, translation, matching and multiple choice exercises. In these exercises, the carrier sentences and other language materials are typically hand-crafted. As a result, the number of items are limited, the language use can sometimes lack authenticity, and the content may not match the users’ individual interests.

Promoting use of authentic material, the WERTi system provides input enhancement to web pages for the purpose of language learning (Meurers et al., 2010). It highlights grammatical constructions on which the user needs practice, and turns them into slot-fill exercises. It handles a wide range of constructions, including prepositions, determiners, gerunds, to-infinitives, wh-questions, tenses and phrasal verbs. On the one hand, the system offers much flexibility since it is up to the user to select the page. On the other, the selected text does not necessarily suit the user in terms of its language quality, level of difficulty and the desired grammatical constructions.

A number of other systems use text corpora to create grammar exercises. The KillerFiller tool in the VISL project, for example, generates slot-fill items from texts drawn from corpora (Bick, 2005). Similar to the WERTi system, an item takes the original word as its only key, and does not account for the possibility of multiple correct answers.

Other systems attempt to generate distractors for the key. Chen et al. (2006) manually designed patterns for this purpose. Smith et al. (2010) utilized a thesaurus, while Zesch and Melamud (2014) developed context-sensitive rules.

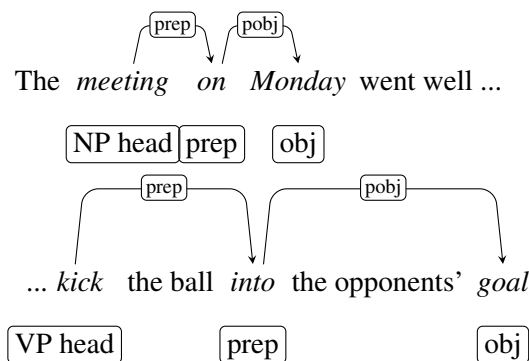


Figure 1: Parse tree for example carrier sentences. Distractors are generated on the basis of the prepositional object (“obj”), and the NP head or VP head to which the prepositional phrase is attached (Section 3). See Table 1 for the item produced from the bottom sentence.

Unlike our approach, they did not adapt to the learner’s behavior. While some of these systems serve to provide draft FIB items for teachers to post-edit (Skory and Eskenazi, 2010), most remain research prototypes.

A closely related research topic for this paper is automatic correction of grammatical errors (Ng et al., 2014). While the goal of distractor generation is to identify words that yield incorrect sentences, it is not merely the inverse of the error correction task. An important element of the distractor generation task is to ensure that distractor appears plausible to the user. In contrast to the considerable effort in developing tools for detecting and correcting preposition errors (Tetreault and Chodorow, 2008; Felice and Pulman, 2009), there is only one previous study on preposition distractor generation (Lee and Seneff, 2007). Our system builds on this study by incorporating novel algorithms for distractor generation and personalization features.

3 Item creation

The system considers all English sentences in the Wikicorpus (Reese et al., 2010) that have fewer than 20 words as carrier sentence candidates. In each candidate sentence, the system scans for prepositions, and extracts two features from the linguistic context of each preposition:

- The **prepositional object**. In Figure 1, for example, the words “Monday” and “goal” are respectively the prepositional objects of the keys, “on” and “into”.

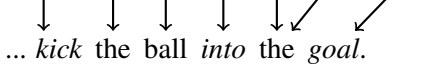
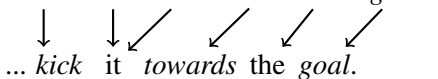
<p>Co-occurrence method</p> <p>... <i>kicked</i> the chair <i>with</i> ...</p> <p>... <i>kicked</i> the can <i>with</i> ...</p> <p>... <i>with</i> the goal of ...</p>
<p>Learner Error method</p> <p>... <i>kicked</i> it <small><error>in</error></small> the goal.</p> <p>... <i>kick</i> the ball <small><error>in</error></small> the other team's goal.</p>
<p>Learner Revision method</p> <p>... <i>kick</i> the ball <i>to</i> his own goal.</p> <p style="text-align: center;">  </p> <p>... <i>kick</i> the ball <i>into</i> the goal.</p> <p>... <i>kick</i> the ball <i>to</i> the goal.</p> <p style="text-align: center;">  </p>

Table 2: The Co-occurrence method (Section 3.1) generates “with” as the distractor for the carrier sentence in Figure 1; the Learner Error method (Section 3.2) generates “in”; the Learner Revision method (Section 3.3) generates “to”.

- The head of the noun phrase or verb phrase (**NP/VP head**) to which the prepositional phrase (PP) is attached. In Figure 1, the PP “into the opponents’ goal” is attached to the VP head “kick”; the PP “on Monday” is attached to the NP head “meeting”.

In order to retrieve the preposition, the prepositional object, and the NP/VP head (cf. Section 3), we parsed the Wikicorpus, as well as the corpora mentioned below, with the Stanford parser (Manning et al., 2014). The system passes the two features above to the following methods to attempt to generate distractors. If more than one key is possible, it prefers the one for which all three methods can generate a distractor.

3.1 Co-occurrence method

This method requires co-occurrence statistics from a large corpus of well-formed English sentences. It selects as distractor the preposition that co-occurs most frequently with *either* the prepositional object *or* the NP/VP head, but not both. As shown in Table 2, this method generates the distractor “with” for the carrier sentence in Figure 1, since many instances of “kick ... with” and “with ... goal” are attested. The reader is referred to Lee and Seneff (2007) for details.

Our system used the English portion of Wikicorpus (Reese et al., 2010) to derive statistics for this method.

3.2 Learner error method

This method requires examples of English sentences from an error-annotated learner corpus. The corpus must indicate the preposition errors, but does not need to provide corrections for these errors. The method retrieves all sentences that have a PP with the given prepositional object and attached to the given NP/VP head, and selects the preposition that is most frequently marked as wrong.

To derive statistics for this method, our system used the NUS Corpus of Learner English (Dahlmeier et al., 2013), the EF-Cambridge Open Language Database (Geertzen et al., 2013) and a corpus of essay drafts written by Chinese learners of English (Lee et al., 2015).

3.3 Learner revision method

Finally, our system exploits the revision behavior of learners in their English writing. This method requires draft versions of the same text written by a learner. It retrieves all learner sentences in a draft that contains a PP with the given prepositional object, and attached to the given NP/VP head. It then selects as distractor the preposition that is most often edited in a later draft. As shown in Table 2, this method generates the distractor “to” for the carrier sentence in Figure 1, since it is most often edited in the given linguistic context. The reader is referred to Lee et al. (2016) for details.

To derive statistics for this method, our system also used the aforementioned corpus of essay drafts.

4 Item selection

Learners benefit most from items that are neither too easy nor too difficult. Following principles from adaptive testing (Bejar et al., 2003), the system tracks the user’s performance in order to select the most suitable items. It does so by considering the vocabulary level of the carrier sentence (Section 4.1) and the user’s previous mistakes (Section 4.2).

4.1 Sentence difficulty

A potential pitfall with the use of authentic sentences, such as those from Wikipedia, is that dif-

difficult vocabulary can hinder the learning of preposition usage. To minimize this barrier, the system starts with simpler carrier sentences for each new user, and then progresses to harder ones.

For simplicity, we chose to estimate the difficulty of a sentence with respect to its vocabulary.¹ Specifically, we categorized each word into one of ten levels, using graded vocabulary lists compiled by the Hong Kong Education Bureau (2012) and the Google Web Trillion Word Corpus.² The lists consist of about 4,000 words categorized into four sets, namely, those suitable for students in junior primary school, senior primary, junior secondary, or senior secondary. Levels 1 to 4 correspond to these four sets. If the word does not belong to these sets, it is classified at a level between 5 and 10, according to decreasing word frequency in the Google corpus. The difficulty level of a sentence is then defined as the level of its most difficult word.

For each new user, the system starts with sentences at Level 4 or lower. It keeps track of his or her performance for the last ten items. If the user gave correct answers for more than 60% of the items from the current level, the system increments the difficulty level by one. Otherwise, it decreases the difficulty level by one.

4.2 Preposition difficulty

In Figure 2, the system presents an item to the user. If the user selects a distractor rather than the key, he or she is informed by a pop-up box (Figure 3), and may then make another attempt. At this point, the user may also request to see a “similar” item to reinforce the learning of the preposition usage (Figure 4). Two items are defined as “similar” if they have the same preposition as key, and the same prepositional object and NP/VP head.

The system records all items to which the user gave incorrect answers; we will refer to this set of items as the “wrong list”. When the user logs in next time, the system begins with a review session. For each item in the “wrong list”, it retrieves a “similar” item from the database (Figure 4), thus facilitating the user in reviewing prepositional usage with which he had difficulty in a previous session. If the user now successfully chooses the key,

¹The difficulty level of a sentence depends also on syntactic and semantic features. Most metrics for measuring readability, however, have focused on the document rather than the sentence level (Miltsakaki and Truitt, 2008; Pitler and Nenkova, 2008).

²<http://norvig.com/ngrams/>

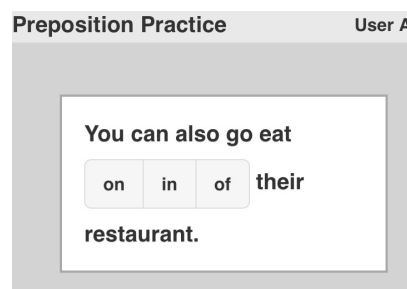


Figure 2: The system displays a carrier sentence with the key “in” and the distractors “on” and “of”.

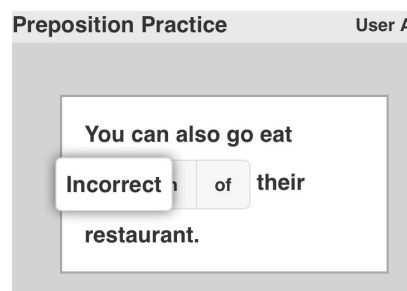


Figure 3: After the user selected the distractor “on” for the item in Figure 2, a pop-up box alerts the user.

the item is taken off the “wrong list”. After the review session, the system resumes random selection of items within the estimated level of sentence difficulty, as described in the last section.

5 Implementation and evaluation

5.1 Architecture

We used the MySQL database, and JSP for the website backend. There are three main tables. The *Question* table stores all carrier sentences selected from the English portion of the Wikicorpus (Reese et al., 2010). To expedite item retrieval and identification of “similar” items, the table stores the key, prepositional object and NP/VP head of each item, as well as the difficulty level of the carrier sentence.

The *Answer* table stores the distractors for each item. Currently, the distractors do not change according to user identity, but we anticipate a future version that personalizes the distractors with respect to the user’s mother tongue.

The *User* table stores the user profile. Information includes the user’s personal “wrong list”, his or her estimated vocabulary level, as well as login time stamps.

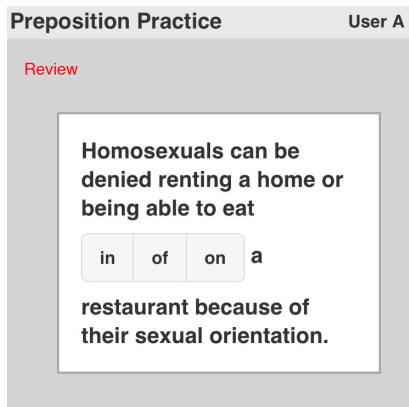


Figure 4: As review for the user, the system offers an item that is similar to the one in Figure 2, which also has “in” as the key, “eat” as the VP head and “restaurant” as the prepositional object.

5.2 Interface

For a better user experience on mobile devices, we used JQuery Mobile for interface development. At the start page, the user can register for a new account, or log in with an existing user name and password. Alternatively, the user can choose to access the system as a guest. In this case, he or she would be treated as a new user, but no user history would be recorded.

The user can attempt an arbitrary number of preposition items before logging out. Each item is presented on its own page, with the distractor and key displayed in random order (Figure 2). The user chooses the best preposition by tapping on its button. If the answer is correct, the system advances to the next item; otherwise, it informs the user via a pop-up box (Figure 3), and then flags the distractor in red. The user may try again until he or she successfully chooses the key.

5.3 Evaluation

To assess system quality, we asked two professional English teachers to annotate a set of 400 items, which included both automatically generated and human-crafted items. For each choice in an item, the teachers judged whether it is correct or incorrect. They did not know whether each choice was the key or a distractor. They may judge one, multiple, or none of the choices as correct.

A distractor is called “reliable” if it yields an incorrect sentence. As reported in Lee et al. (2016), the proportion of distractors judged reliable reached 97.4% for the Learner Revision

method, 96.1% for the Co-occurrence method, and 95.6% for the Learner Error method.

For each incorrect choice, the two annotators further assessed its plausibility as a distractor from their experience in teaching English to native speakers of Chinese. They may label it as either “obviously wrong”, “somewhat plausible”, or “plausible”. The Learner Error method produced the best distractors, with 51.2% rated “plausible”, followed by the Learner Revision method (45.4%) and the Co-occurrence method (34.6%). The number of plausible distractors per item among the automatically generated items compares favourably to the human-crafted ones (Lee et al., 2016).

6 Conclusion

We have presented a CALL system that turns sentences from Wikipedia into fill-in-the-blank items for preposition usage. Using statistics from both standard and learner corpora, it generates plausible distractors to provide multiple choices.

The system tailors item selection for individual learners in two ways. First, it chooses carrier sentences that matches the learner’s estimated vocabulary level. Second, to facilitate learning, it offers review sessions with items that are similar to those with which the learner previously demonstrated difficulty.

In future work, we plan to extend the system coverage beyond preposition to other common learner error types.

Acknowledgments

We thank the reviewers for their very helpful comments. This work was supported in part by an Applied Research Grant (Project no. 9667115) from City University of Hong Kong.

References

- Isaac I. Bejar, René R. Lawless, Mary E. Morley, Michael E. Wagner, Randy E. Bennett, and Javier Revuelta. 2003. A Feasibility Study of On-the-Fly Item Generation in Adaptive Testing. *The Journal of Technology, Learning, and Assessment*, 2(3).
- Eckhard Bick. 2005. Grammar for Fun: IT-based Grammar Learning with VISL. In P. Juel, editor, *CALL for the Nordic Languages*, pages 49–64. Copenhagen: Samfundslitteratur, Copenhagen Studies in Language.
- Hong Kong Education Bureau. 2012. *Enhancing English Vocabulary Learning*.

- ing and Teaching at Secondary Level. http://www.edb.gov.hk/vocab_learning_sec.
- Chia-Yin Chen, Hsien-Chin Liou, and Jason S. Chang. 2006. FAST: An Automatic Generation System for Grammar Tests. In *Proc. COLING/ACL Interactive Presentation Sessions*.
- Daniel Dahlmeier, Hwee Tou Ng, and Siew Mei Wu. 2013. Building a Large Annotated Corpus of Learner English: The NUS Corpus of Learner English. In *Proc. 8th Workshop on Innovative Use of NLP for Building Educational Applications*.
- Pierluigi Erbaggio, Sangeetha Gopalakrishnan, Sandra Hobbs, and Haiyong Liu. 2012. Enhancing Student Engagement through Online Authentic Materials. *The International Association for Language Learning Technology Journal*, 42(2):27–51.
- Rachele De Felice and Stephen Pulman. 2009. Automatic Detection of Preposition Errors in Learner Writing. *CALICO Journal*, 26(3):512–528.
- Jeroen Geertzen, Theodora Alexopoulou, and Anna Korhonen. 2013. Automatic Linguistic Annotation of Large Scale L2 Databases: The EF-Cambridge Open Language Database (EFCAMDAT). In *Proc. 31st Second Language Research Forum (SLRF)*.
- Emi Izumi, Kiyotaka Uchimoto, Toyomi Saiga, Thepchai Supnithi, and Hitoshi Isahara. 2003. Automatic Error Detection in the Japanese Learners' English Spoken Data. In *Proc. ACL*.
- Ruth E. Larimer and Leigh Schleicher. 1999. *New Ways in Using Authentic Materials in the Classroom*. Teachers of English to Speakers of Other Languages, Inc., Alexandria, VA.
- John Lee and Stephanie Seneff. 2007. Automatic Generation of Cloze Items for Prepositions. In *Proc. Interspeech*.
- John Lee, Chak Yan Yeung, Amir Zeldes, Marc Reznicek, Anke Lüdeling, and Jonathan Webster. 2015. CityU Corpus of Essay Drafts of English Language Learners: a Corpus of Textual Revision in Second Language Writing. *Language Resources and Evaluation*, 49(3):659–683.
- John Lee, Donald Sturgeon, and Mengqi Luo. 2016. A CALL System for Learning Preposition Usage. In *Proc. ACL*.
- Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP Natural Language Processing Toolkit. In *Proc. ACL System Demonstrations*, pages 55–60.
- Detmar Meurers, Ramon Ziai, Luiz Amaral, Adriane Boyd, Aleksandar Dimitrov, Vanessa Metcalf, and Niels Ott. 2010. Enhancing Authentic Web Pages for Language Learners. In *Proc. Fifth Workshop on Innovative Use of Nlp for Building Educational Applications*.
- Eleni Miltsakaki and Audrey Troutt. 2008. Real Time Web Text Classification and Analysis of Reading Difficulty. In *Proc. Third Workshop on Innovative Use of NLP for Building Educational Applications*.
- Hwee Tou Ng, Siew Mei Wu, Ted Briscoe, Christian Hadiwinoto, Raymond Hendy Susanto, and Christopher Bryant. 2014. The CoNLL-2014 Shared Task on Grammatical Error Correction. In *Proc. 8th Conference on Computational Natural Language Learning: Shared Task*, pages 1–14.
- Emily Pitler and Ani Nenkova. 2008. Revisiting Readability: a Unified Framework for Predicting Text Quality. In *Proc. EMNLP*.
- Samuel Reese, Gemma Boleda, Montse Cuadros, Lluís Padró, and German Rigau. 2010. Wikicorpus: A Word-Sense Disambiguated Multilingual Wikipedia Corpus. In *Proc. LREC*.
- Adam Skory and Maxine Eskenazi. 2010. Predicting Cloze Task Quality for Vocabulary Training. In *Proc. NAACL HLT 2010 Fifth Workshop on Innovative Use of NLP for Building Educational Applications*.
- Simon Smith, P. V. S. Avinesh, and Adam Kilgarriff. 2010. Gap-fill Tests for Language Learners: Corpus-Driven Item Generation. In *Proc. 8th International Conference on Natural Language Processing (ICON)*.
- Joel Tetreault and Martin Chodorow. 2008. The Ups and Downs of Preposition Error Detection in ESL Writing. In *Proc. COLING*.
- Peter Watcyn-Jones and Jake Allsop. 2000. *Test Your Prepositions*. Penguin Books Ltd.
- Jean Yates. 2010. *The Ins and Outs of Prepositions*. Hauppauge, New York : Barron's.
- Torsten Zesch and Oren Melamud. 2014. Automatic Generation of Challenging Distractors Using Context-Sensitive Inference Rules. In *Proc. Workshop on Innovative Use of NLP for Building Educational Applications (BEA)*.