# How to Generate Cloze Questions from Definitions: A Syntactic Approach

**Donna Gates**
5000 Forbes Ave
Carnegie Mellon Univ.
Pittsburgh, Pa. 15213
dmg@cmu.edu

**Greg Aist**
206 Ross Hall
Iowa State Univ.
Ames, IA 50011-1201
gregoryaist@alumni.cmu.edu

**Jack Mostow**
5000 Forbes Ave
Carnegie Mellon Univ.
Pittsburgh, PA 15213
mostow@cmu.edu

**Margaret McKeown**
743 LRDC
Univ. of Pittsburgh
Pittsburgh PA 15260
mckeown@pitt.edu

**Juliet Bey**
5000 Forbes Ave
Carnegie Mellon Univ.
Pittsburgh, PA 15213
jpbey@cmu.edu

**Abstract**

This paper discusses the implementation and evaluation of automatically generated cloze questions in the style of the definitions found in Collins COBUILD English language learner's dictionary. The definitions and the cloze questions are used in an automated reading tutor to help second and third grade students learn new vocabulary. A parser provides syntactic phrase structure trees for the definitions. With these parse trees as input, a pattern matching program uses a set of syntactic patterns to extract the phrases that make up the cloze question answers and distracters.

## Introduction

Multiple-choice questions are common test and practice exercises for a variety of subject matter including reading and language teaching. These types of questions require human effort to produce. Brown, Frishkoff and Eskenazi (2005) generate cloze questions for vocabulary assessment. They remove a single word from the original sentence and generate single word distracters. Mostow et al (2004) describe a reading tutor which deletes a random word from a sentence in a story to form a cloze question and uses words from elsewhere in the same story as distracters. Mitkov and Ha (2003) developed a program to generate grammar test questions from text using shallow parsing techniques and lexical knowledge. A phrase is removed from a sentence stating a grammar principle and, in turn, it is used as the answer. The sentence is transformed into a WH-question. They generate other grammar instruction phrases of similar grammatical type to become the distracters.

As part of a reading tutor's vocabulary component, we designed a multiple-choice cloze question based on the definition of a target vocabulary word with a phrase removed from the definition. This phrase contains information that explains the meaning of the word. In order to save time and human resources, we chose to automate the generation of these questions and their distracters.

Using Stanford's NLP Parser (Klein and Manning 2003), we parse definitions written for the purpose of learning vocabulary while reading with Project LISTEN's Reading Tutor (Mostow 2007). These parses are then transformed into cloze questions with distracters that originate from other definitions.

We will first describe the definitions and how the cloze questions will be used by the tutor during vocabulary instruction. Then, we will describe how the questions and distracters are generated and filtered. We will then show how the questions compare to hand-written questions. Lastly, we conclude and discuss future work.

## Definitions

The definitions used by the Reading Tutor for teaching vocabulary were written by members of Project LISTEN in the style of Collins COBUILD English Learner's Dictionary (Rammell and Collins 2003). An expert in teaching vocabulary to children edited the definitions. The definitions follow the suggestions described in Beck, McKeown and Kucan (2002) as to how to make definitions appropriate for young children. They are worded simply so that they can be understood by second and third grade readers. The COBUILD-style is preferred over standard definition formats because it gives a context for the word in the definition and it states the definition in a complete sentence. Definitions in this style take several forms: a verb or adjective usually appears as:

*If you VERB something, you DO-SOMETHING to it.*
*If you are ADJECTIVE, you …*

Noun definitions take the typical forms:

*A NOUN is a THING/PERSON that/who …*

*A NOUN is a ... THING/PERSON.*

For example, below is the definition for the verb *abandon*:

*If you abandon something or someone, you leave them and never go back.*

The following is our definition for the noun *steak*:

*A steak is a large flat piece of meat*.

## Definitions in the tutor

Initially, a student sees the definition of a vocabulary word when he first encounters it in a story. The next day the student is asked to perform a vocabulary activity to reinforce learning. Prior to this, the student is reminded of the definition via a cloze question.

The definition cloze questions refresh the student's memory of the definition of the word whether they get the answer correct or not. Every day for four days the student receives a new word practice exercise for a vocabulary word that was introduced while reading a story earlier in the week. When the student receives the cloze question for the definition, he is asked to choose the correct answer from 2 randomly ordered choices: correct answer vs. distracter. For example,

*A steak is _____.*

*an old broken down car*

*a large flat piece of meat*

The task is not meant to test the student's knowledge but to serve as a reminder before starting the word practice activity. Since there are often 3-5 distracters available to choose from for each question, the tutor uses a new distracter each time the cloze question is shown and then recycles the distracters if it runs out.

## Question Generation

The cloze question generator was developed and tested using 308 student definitions. 30 additional word senses and definitions were set aside for evaluating the generator.

The syntactic patterns that we use to find the blanks in the definitions as well as the distracters are based solely on the phrase-structure trees produced when we parsed the definitions with the Stanford NLP Parser. Figure 1 shows an example phrase-structure parse for the definition, *If you abandon something or someone, you leave them and never go back.*

```
(ROOT (S  (SBAR  (IN If)
                 (S (NP (PRP you))
                   (VP (VBP abandon)
                       (NP (NN something)
                           (CC or)
                           (NN someone)))))
          (, ,)  (NP (PRP you))
          (VP  (VP (VBP leave) (NP (PRP them)))
               (CC and)  (ADVP (RB never))
               (VP (VBP go) (ADVP (RB back))))  (. .)))
```

**Figure 1 Parse of Definition for *abandon***

We look for specific syntactic patterns that match the "defining information" (that part that shows the meaning of the word) in the definition. For example, in the above definition, the defining information is *you leave them and never go back*.

Parts of this phrase will be removed to become the blank/answer. The verb phrase (VP) "*leave them*" is a candidate for the answer-blank because it matched a pattern looking for a VP in this context.

One of the syntactic patterns looks for a noun phrase (NP) with the pronoun *you* immediately followed by a VP. This VP becomes the answer and gets added to the list of other phrases stored for distracter generation that match "simple-VP-after-you" (which excludes the coordinate structure). The pattern matches on :

*(SBAR (S (NP (PRP you)) (VP (VP x)...*

where *x* is the content we are looking for. The final cloze question is generated by removing the answer phrase and replacing it with a 5 underscore blank space.

*If you abandon something or someone, you _____ and never go back.* (Answer: *leave them*)

Yet another pattern is used to capture the complete VP with the coordinate structure "complex-VP-after-you" which would match the VP *leave them and never go back.*

There are also patterns that look for and extract noun phrases, prepositional phrases, relative clauses and sentential complements, as shown in Table 1.

| | |
|---|---|
| Noun Phrase after BE | A pace is *a step you take* when you walk. |
| Sentential Complement | If you persuade someone, you tell them *what to think or do*. |
| Prepositional Phrase | Your profile is the outline of your face seen *from the side*. |
| Relative Clause | Your attitude about something is the way *you think and feel about it*. |
| Reduced Relative Clause | Your profile is the outline of your face *seen from the side.* |

**Table 1 Examples of phrases matched by syntactic patterns**

## Distracters

We chose to use the phrases from other vocabulary definitions rather than generate new phrases or extract them from other texts. The advantages are that the phrases are already in language that children can easily read (simple vocabulary and syntax) and there is no chance that another vocabulary word will appear accidently since the definitions were carefully worded to exclude other target vocabulary words. The target vocabulary word only appears at the beginning of a definition (i.e., in the initial subordinate *IF* clause or before a predicate verb such as BE).

Based on advice from our vocabulary expert, it was determined that the following requirements were necessary for a distracter to pass the human review process:

1. Should be of relatively the same syntactic phrase type as the answer,
2. Should have roughly the same length as the answer
3. Should not be a possible answer
4. Should be grammatical when inserted into the blank

The fourth constraint allows for some loosening of person and number agreement constraints, according to the vocabulary expert, since the point is to make sure the child learns the meaning and not memorize the definition.

## Distracter Filters

The length of the distracters were filtered so that no distracter would be more than 20 characters longer or shorter than the answer with preference for ones that are no more than 11 characters longer or shorter. These lengths were determined by hand after repeating the generation with differing length thresholds and checking with the vocabulary expert as to which ones appeared to work best so that the answer and the distracter did not look significantly different from one another.

The distracters were filtered based on phrase type such that questions whose answer is a simple VP, only have distracters that are simple VPs. For example, the answer for the cloze question *If you abandon someone or something, you _____ and never go back.* is the VP *leave them.* A good VP distracter would be *ask them a question*, or *buy them a present* but not *a large flat piece of meat*. Phrases that matched on exactly the same pattern were collected and saved to be used as distracters for definitions with the same pattern.

In an effort to improve the filtering of distracters that are too closely related semantically to the answer, we tried to use WordNet (Fellbaum 1998) similar to Gates (2008) to compare the definitions of the answer word to that of the source word from which a specific distracter originated. While there appeared to be evidence that the distracters could be filtered to with lexical information, an early evaluation showed that, except in one case, there was no change to the final 5 distracters selected for each question when compared to not using any WordNet information. The syntactic pattern and length filters prevented these distracters from ever being considered in the first place. Figure 2 shows the output for the question and distracter generator displaying the filter warning messages which state that the distracter *cause it* is not only too short but also possibly too close in meaning to the answer to be a valid distracter.

The filters applied in a specific order: phrase type, length, and then semantic filter.

*If you construct something, you _____.*
**build it by putting parts together*
*say it in a clear, strong way*
*change it back to the way it was*
*clean it by rubbing it very hard*
*behave badly and are not polite*
*bring it back from the place where it was left*
*aim it at them or say it only to them*
-- *caused it* (Filtered: length too-short, Filtered: WordNet definition for distracter's vocabulary word (attribute) contains target word (construct)

**Figure 2 Example of semantic filter being overshadowed by length restriction filter**

To ensure that we generate enough good distracters so that there is at least one remaining after a human reviews them, the program selects 5 phrases from the pool of possible phrases.

### Human Review

The generated distracters from the cloze questions for the 33 unseen word senses were reviewed by a human to determine whether they were adequate distracters for a given cloze definition. According to our vocabulary expert, each distracter must not be similar in meaning to the answer, should not stand out as being longer or shorter, should not sound completely implausible grammatically, and should not be too vague or too specific so that it could be interpreted as a possible answer. The following example illustrates a cloze question and distracters that are acceptable and unacceptable.

*If you abandon something or someone, you _____ and never go back.*

*leave them* (answer)
*look for it* (ok: clearly conflicts in this definition)
*do it* (too vague: could fit)
*are very mean* (too close/specific: could fit if you think it is mean to abandon something)
*lose it* (too close in meaning)
*look for someone that you have not met before* (too long)

# Evaluation

The program was evaluated on 33 vocabulary word sense definitions that were excluded from being used during the development of the program, the filters and the patterns. A human, expert in writing cloze questions for children, wrote 33 cloze questions (sentences with a blank plus an answer) and another human, expert in writing distracters, wrote a single distracter for each of the cloze questions. It took 31 minutes to write 33 cloze questions and 22 minutes

to write 33 distracters by hand. In contrast, the question generator produced 91 cloze questions for the 33 word senses and over 522 distracters in less than 5 seconds (not including time to parse the definitions). The two human experts judged the generated questions and distracters, taking an average of 49.5 minutes (range 32-67) to review and note whether the questions and the distracters were acceptable. For the purposes of this discussion, an item refers to either a cloze question (definition with blank space) or to a distracter. Based on items that the judges agreed on, 73% of the generated output was acceptable. There were 613 items (91 cloze questions + 522 distracters). This judging process revealed that the generator produced 356 good items (77 cloze questions + 279 distracters) and 135 bad items (1 cloze question + 134 distracters). The judges were not in agreement on the remaining 122 items.

In the case of the single unacceptable cloze question, the answer phrase that was chosen for deletion was simply too short: *If you make a distinction between 2 things, you _____ or say how they are different.* Answer: *show*

Judging the generated output yielded 7.19 acceptable items/minute while hand-writing the examples yielded 1.29 items/minute. Since it takes less time per item to judge the output than write one, a further advantage of generating the cloze questions automatically is that we get a variety of questions (2.8 per definition) in approximately 10 seconds (7 seconds to generate cloze questions and organize phrase and 3 seconds to group and filter distracters for a specific cloze question). The current tutor happens to only take advantage of the extra distracters.

Parsing and pattern matching errors accounted for fewer than 10 unacceptable cloze questions in the development set, and no parsing errors occurred in the evaluation set. Some of the definitions only had 1 or 2 distracters which were not acceptable. Distracter generation requires a large enough pool of syntactically similar phrases from other definitions and, for these cases, there were not enough.

### Discussion

While over-all it took less time to write a cloze question and a single distracter by hand, it was more efficient to have a human review generated items on a per item basis. A further advantage of automatically generating the cloze questions is that it produced a variety of questions in different forms that could be used by the tutor on different days. The current tutor happens to only take advantage of the extra distracters.

## Future Work

The Reading Tutor is currently being used in an experiment with second and third graders learning vocabulary and includes 300 automatically generated definition cloze questions. The experiment is still in the initial stages at the time of this writing. We hope to have some feedback from the experiment soon. We would also like to continue experimenting with a filter to better restrict the distracters and lessen the cost of human reviewing. If the generated results can be improved and automatically thinned out by filters, it would greatly benefit the next step when we scale up the vocabulary tutor.

## Acknowledgements

Presentation of this work was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A080157 to Carnegie Mellon University. The opinions expressed are those of the authors and do not necessarily represent the views of the Institute or the U.S. Department of Education.

The authors wish to thank members of Project Listen for their assistance in preparing the evaluation results and the anonymous reviewers for their helpful suggestions and comments.

## References

Beck, I. L., McKeown, M. G., and Kucan, L. 2002. *Bringing words to life : robust vocabulary instruction*. New York: Guilford Press.

Brown, J., Frishkoff, G., and Eskenazi, M. 2005. *Automatic question generation for vocabulary assessment*. Paper presented at the HLT/EMNLP conference, Vancouver, B.C.

Fellbaum, C. (Ed.) 1998. *WordNet: An Electronic Lexical Database*. Cambridge, Ma: MIT Press

Gates, D. 2008. Generating Look-Back Strategy Questions from Expository Texts. In *Proceedings of the Workshop on the Question Generation Shared Task and Evaluation Challenge*, NSF, Arlington, VA.

Klein, D., and Manning, C. 2003. Fast Exact Inference with a Factored Model for Natural Language Parsing. *Advances in Neural Information Processing Systems 15 (NIPS 2002)*, 3-10.

Mitkov, R., and Ha, L. A. 2003. *Computer-aided generation of multiple-choice tests*. Paper presented at the HLT-NAACL 2003 Workshop on Building Educational Applications Using Natural Language Processing, Edmonton, Canada.

Mostow, J., Beck, J., Bey, J., Cuneo, A., Sison, J., Tobin, B., and Valeri, J. 2004. Using automated questions to assess reading comprehension, vocabulary, and effects of tutorial interventions. *Technology, Instruction, Cognition and Learning, 2*, 97-134.

Mostow, J., and Beck, J. 2007. When the Rubber Meets the Road: Lessons from the In-School Adventures of an Automated Reading Tutor that Listens. *Scale-Up in Education, 2*, 183-200.

Rammell, C. and Collins. (Eds.) 2003. Collins COBUILD Learner's Dictionary. HarperCollins.