1. Abstract

Automatic Construction of Multiple-choice Cloze Questions

How do the BAWE and Wikipedia items compare to manual items?

Aims of this research project:
- Develop a tool to produce vocabulary quizzes from lists.
- Evaluate quiz items produced using the system with feedback from teachers and students.
- Compare the effectiveness of online and offline corpus resources in the production of quiz items.

Multiple-choice cloze is the most studied and is often used in vocabulary instruction, training, and multiple-choice assessment, though Lee et al. 2013 and Liu et al. 2005 are close.

Multiple-choice cloze is a time-honored method of vocabulary instruction, training, and distractor effectiveness that is common in language instruction. The ability to generate similar items is a significant advantage of the automatic construction of test questions (e.g., Goto et al. 2009; Mitkov et al. 2008).

Word Quiz Constructor (WQC) is a Java application designed to construct a graphical user interface, expand capability for other vocabulary lists, and multiple-choice cloze is the most studied and is often used in vocabulary instruction, training, and multiple-choice assessment, though Lee et al. 2013 and Liu et al. 2005 are close.

Multiple-choice cloze is a time-honored method of vocabulary instruction, training, and distractor effectiveness that is common in language instruction. The ability to generate similar items is a significant advantage of the automatic construction of test questions (e.g., Goto et al. 2009; Mitkov et al. 2008).

2. Background

Many types of questions may be used to test and evaluate students’ knowledge.

3. Word Quiz Constructor

Resources

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Sample Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Select key</td>
<td>- Random choice from specified AWL sublist</td>
</tr>
<tr>
<td>Select stem</td>
<td>- Random choice from specified corpus</td>
</tr>
<tr>
<td>Select distractors</td>
<td>- Random choice from specified corpus</td>
</tr>
<tr>
<td>Finalize item</td>
<td>- Output in specified format (text, csv, moodle, XML, quizlet)</td>
</tr>
</tbody>
</table>

4. Experimental Evaluation

Experiment 1: Well-formedness and difficulty

Experienced EFL teachers (N=32, avg teaching exp=21 yrs) judged well-formedness and then difficulty (relative to key) of well-formed items in forced-choice paradigm.

Experiment 2: Facility, discrimination, efficiency

Undergraduate and graduate student teachers (N=32) responded to the stimuli items in a simulated vocabulary test in exchange for 1000 yen (60 €).

5. Summary

How do the BAWE and Wikipedia items compare to manual items?

6. Discussion

One possible reason for the well-formedness drop in low-level Wikipedia items is that Wikipedia’s writing style is normally quite high. Items with low ARI might not be normal writing. However, in abbreviations, foetalise, or academic shorthand (e.g., mathematics), it may be useful to have a lower ARI threshold in addition to the upper threshold to control this. Wikipedia items also take significantly longer to produce, but that is probably caused by networking delays and limits on the number of API requests by the Wikipedia server.

The limitations with Wikipedia likely result from the smaller size of the corpus: Many items may fail to be finalised, thus costing time. Those that are finalised are more difficult, perhaps because of a greater concentration of difficult vocabulary. Nonetheless, on the whole, Word Quiz Creator is capable of producing vocabulary test items on a par with those produced manually. An online corpus (Wikipedia) and an offline corpus (BAWE) perform somewhat similarly in this process, but may potentially complement each other to produce useful items.

7. Future work

Improved improvements to Word Quiz Creator include:
- Use Google n-grams rather than BAWE n-grams. This should increase item acceptability rate, speeding up production time.
- Use a local server installation of Wikipedia rather than the wiki site directly, also speeding up production time.
- Add other question types (e.g., matching, word-ordering).
- Construct a graphical user interface.
- Expand capacity for other vocabulary lists.
- Prepare application for free distribution.

Furthermore, since the current study’s scale is relatively small, future work will include more extensive testing of the Word Quiz Creator’s output in order to validate the usefulness of the items for testing vocabulary knowledge.

Acknowledgements

This research is partially supported by a Waseda University Grant for Special Research Projects (Project number: 2014K-6142). I am also grateful to Laurence Anthony for advice and assistance at the beginning of this stage of work and also to several teachers who participated voluntarily in Experiment 1.

References

