Collocations, or typical combinations of words such as *strong tea* or *crystal clear*, are an essential part of fluent language use. It is well-known that English language learners (ELL), even advanced ones, often struggle with English collocations (Granger 1998, Nesselhauf 2003). The goal of this study is to develop a automated collocation error detection tool to provide a more context-sensitive lexical support for ELL writers than dictionaries or thesauri. To our knowledge, this is the first rigorous evaluation of collocation detection system, since it uses over a thousand strings and 2 raters.

Our basic approach consists of first extracting candidate strings from ELL-written essays that match a target construction. Next, using a reference database (RDB), which is built from about 1 billion words of native-speaker-written texts, the tool checks how likely the target word is in the context of the rest of the string. Our RDB was built using a rank-ratio (RR) metric (Deane 2005) to determine collocational strength using two corpora: the Lexile corpus (informational and literary texts; collected by the Metametrics Corporation); and the SourceFinder corpus (literary and scientific journal articles). The rank-ratio metric measures how likely the occurrence of a word is in a given context using its *actual rank* and *expected rank*. To determine if the target word (e.g. *strong* in *strong tea*) in the original string is the best collocate or not, the tool also replaces the target word with synonyms and checks the collocational strength to compare against that of the original string. In the end, if the original string's collocational strength is below a threshold, or at least one of the synonym is a significantly stronger collocate, the string is judged to contain a collocation error.

To evaluate the performance of the collocation detection tool, a total of 1,446 target strings were extracted from randomly selected 300 essays from Test of English TOEFL essays, from all score points (1 – 6). Two trained native speakers of English were asked to judge whether each string was OK, an error, or a string extraction error. The tool itself returned 3 responses, OK, ERROR, and ND (no data). ND was grouped with ERROR for the purpose of this evaluation. Precision and recall for "OK" judgments were practically the same for inter-rater agreement and machine-rater agreement. The numbers for the former were 0.841 (with rater 1 as the gold standard) and 0.930 (with rater2 as the gold standard). For rater-machine agreement, precision and recall were calculated on the 1,100 strings on which the raters agreed. The precision for this was 0.89 and recall, 0.93. For "Error" judgments, the numbers were lower for both, with inter-rater agreement being somewhat higher (inter-rater: 0.573 and 0.626; machine-learner: precision = 0.28, and recall = 0.43).

Our analysis shows that the two most frequent tool errors were (i) string extraction errors; and (ii) data sparsity in the RDB (strings not found). Of the 1,110 strings on which the two raters agreed, the tool erred on 195 (17.5%). Of these, 75 were due to string extraction errors, which account for 38.4% of all errors and 6.8% of all strings. This was closely followed by 67 data sparsity cases (34.4% of all errors, 6% of all strings). The third most common cause was spelling errors, which caused 37 errors (19% of all errors, 3.3% of all strings). Since there is no similar system which has gone through the same type of evaluation to our knowledge, whehter our system is "good enough" can only be assessed on the basis of whether it performs well enough for our purpose, which is to provide a context-sensitive lexical resource for ELLs learning English writing.

References

Deane, P. (2005). A nonparametric method for extraction of candidate phrasal terms. *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: collocations and formulae. In A.P. Cowle, (Ed.), *Phraseology: Theory, Analysis, and Application*. Oxford: Clarendon Press.

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics 24(2)*.