Aist, G., Slater, T., Oakey, D., & Ramaeke, H. (2011). Recursive search and social media for STEM vocabulary learning. In S. Huffman & V. Hegelheimer (Eds.), *The role of CALL in hybrid and online language courses*. Ames, IA: Iowa State University.

Recursive Search and Social Media for STEM Vocabulary Learning

Gregory Aist
Tammy Slater
David Oakey
Heidi Ramaeke
Iowa State University

Science, technology, engineering, and mathematics (STEM) form a cluster of academic subjects of increasing interest. Lists of academic vocabulary such as the 570-item Academic Word List (AWL) do not generally focus on STEM. Previously (Aist et al., 2010) we delineated four processes for developing a STEM-specific vocabulary from a general resource such as the AWL: selecting STEM words, filtering out non-STEM words, pairing some words with related non-AWL words (e.g., *qualitative*, *quantitative*), and extending some words into STEM-specific phrases (e.g., *genetic factor*, *algebraic factor*). Extension into phrases was aimed in part at recontextualizing vocabulary items to specify their STEM meaning.

This presentation discussed an interesting finding emerging from two extension methods, one tapping written language on the Web, and the other tapping speakers' intuitions about words.

The first method involved searching on Google for documents containing a STEM term (e.g., science) and the word itself, and then recursively searching for any collocation that emerged as obvious -- "popped out" -- of the first page of search results, appearing frequently and towards the top. The method finished successfully when results were stable, when search on a collocation yielded many hits containing that same collocation. For example, of the first page of results for *mathematics parallel* (Figure 1), six of nine contain *parallel lines*; when the collocation *parallel lines* is searched for, seven of nine contain *parallel lines* and thus the search is stable, and *parallel lines* is one of the results (Figure 2). Likewise, when searching for *computer parallel*, the collocation *parallel computing* is found in the first and second of the top nine results; when the collocation *parallel computing* is searched for, ten of the top ten hits contain *parallel computing* and thus the search is stable, and *parallel computing* is one of the results. [This example determined April 7, 2011, based on non-image, non-video, non-book results.]

The second method, based on social media, involved looking for information that might be useful in constructing STEM-specific collocations from the user-edited Wiktionary. For example, the English language url for *parallel* is http://en.wiktionary.org/wiki/parallel. As of March 29, 2010, there were two adjective senses of *parallel*, and three noun senses: "One of a set

of parallel lines", "A line of latitude", and "An arrangement of electrical components such that a current flows along two or more paths; see in parallel." (Figure 3).

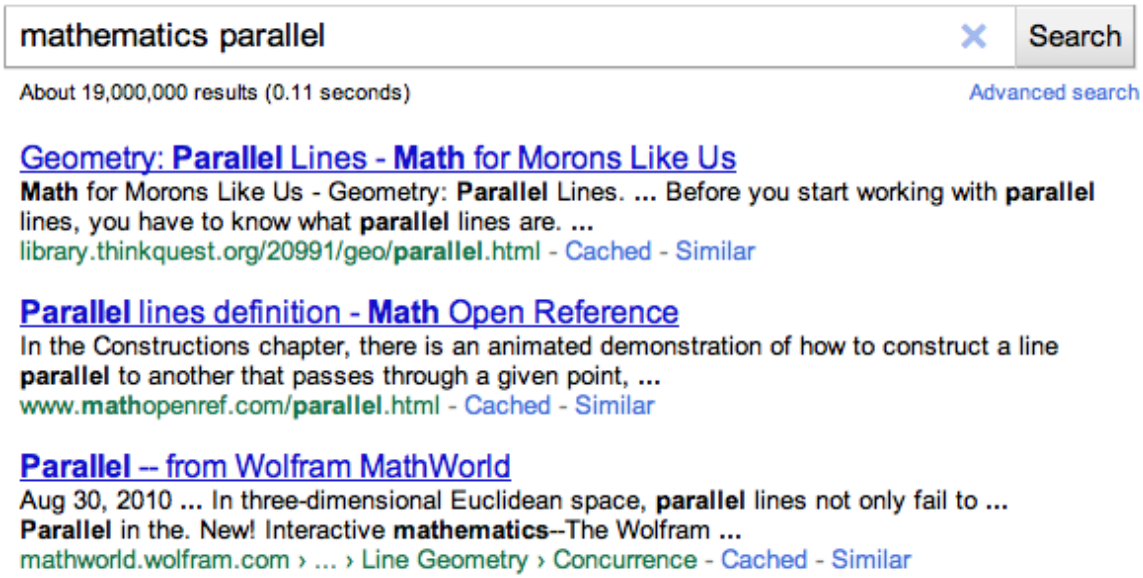

Figure 1. Google search for "mathematics parallel", showing "parallel lines" as a potential collocation.

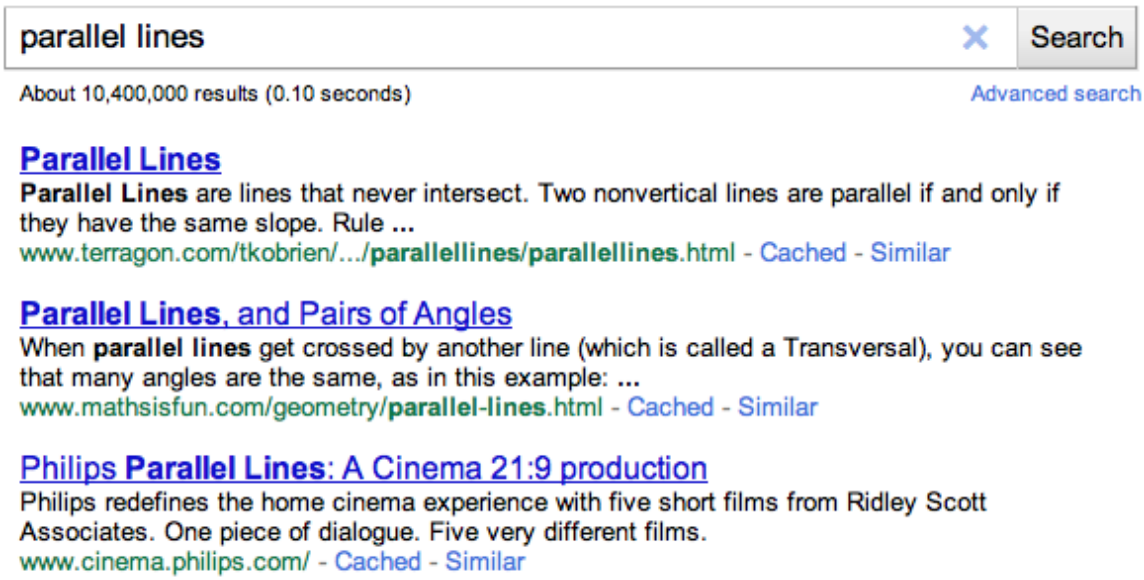

Figure 2. Google search for "parallel lines." The recursive search has stabilized since "parallel lines" is also found in the results.

Adjective

parallel (*not comparable*)

1. Of two or more (straight) lines, (flat) surfaces etc: Equally distant from one another at all points.
   The horizontal lines on my notebook paper are parallel.
2. **parallel to**: Of one thing, relative to another: Equally distant from at all points.
   *The railway line runs **parallel** to the road.*

Noun

parallel (*plural* **parallels**)

1. One of a set of parallel lines.
2. A line of latitude.
   *The 31st **parallel** passes through the center of my town.*
3. An arrangement of electrical components such that a current flows along two or more paths; see in parallel

Figure 3. Wiktionary entries for "parallel" from March 29, 2010.

Each word on the AWL was coded with a three-level coding for each of the two strategies. Zero (0) was assigned if no STEM senses were found. A two (2) was assigned if at least one clear STEM sense was found in Wiktionary, or in the Google search if several/many hits were found, or 4 or more hits across several senses. The middle level (1) was assigned for possible STEM senses in Wiktionary or possible STEM-related hits in Google.

From the analysis and comparison of these methods an interesting pattern emerged. Wiktionary yielded a STEM-specific meaning in an apparently linear pattern: more often for more frequent words. Recursive search, by contrast, yielded a STEM-specific meaning in an apparently normal (bell-shaped) distribution with respect to frequency. Thus for the higher frequency groups on the AWL, Wiktionary is likely to have more useful information than a recursive Google search. For the middle frequency groups, Wiktionary and Google are equally likely to have useful information about phrases, and in some cases Google search may yield more. For the lower frequency groups, Wiktionary is likely to have more useful information than recursive Google search. These differences suggest that learners or teachers would be able to gather more examples of STEM collocations by combining search-based and social media-based methods.

Here is an example to illustrate the contrast between methods. The Google search results for the word *conduct* were dominated by the sense "conduct research" or "conduct"/"misconduct" – which are not limited to STEM, but are generally useful in academic English. Wiktionary, by contrast, listed those senses [as of Sept 11, 2010] as well as STEM-related senses "To serve as a medium for conveying; to transmit, as heat, light, electricity, etc." and "To act as a conductor (as of heat, electricity, etc.); to carry."

Contrary to this pattern is the word *goal*, with *goal node* the end result of a recursive Google search on *computer goal*, but with no useful STEM sense in Wiktionary [as of Sept. 11, 2010].

How might these findings inform the design of language learning activities, particularly in an online / hybrid course? When giving tasks that involve looking up vocabulary, a teacher could direct some students to online search (Google) and some to social media dictionaries (Wiktionary). The resulting differences in information provide a ready-made resource for dynamically constructed information gap activities, where each learner in a pair has some information that the other does not. For example, in computing contexts, one learner could look up *goal*, *network*, and *parallel* on Wiktionary, and the other could look the same words up using recursive Google search. The two learners could then discuss with each other how to arrive at a good phrase that illustrates the use of each word in a computer science context.

One final note: We studied the words on the AWL and how they are presented on the Web (via Google) and in social media. Web pages and social media are like a river – you cannot step into the same Web page twice. Between when this research was conducted in Spring 2010 and the conference presentation that same fall, the Wiktionary entry for *parallel* was edited to include a computing-specific definition: "Describing the processing of multiple tasks at the same time". The findings thus apply to the class of words described (the Academic Word List), and to the parameter we investigated (frequency), so in general words are likely to pattern this way, but a specific word may well vary over time in terms of its appearance in Web searches or in social media dictionaries.

REFERENCE

Aist, G., Oakey, D.J., Slater, T., and Ramaeker, H. (2010). Towards forming a STEM wordlist for ESL. Presented at Computer-Assisted Language Instruction Consortium (CALICO) 2010 conference, Amherst MA, June 10, 2010.