Li-Shih Huang
CANADA

Using Guided, Corpus-Aided Discovery to Generate Active Learning

Over the years, educators have proposed a variety of active learning pedagogical approaches that focus on encouraging students to discover for themselves the principles and solutions that will engage them in learning and enhance their educational outcomes. Among these approaches are problem-based, inquiry-based, experiential, and discovery learning, all of which utilize such techniques as group work, hands-on experience, and social interaction to enable students to discover new concepts on their own.

In recent years, some researchers have pointed out that these unguided or minimally guided instructional approaches lack adequate empirical evidence to support their efficacy, especially for novice learners (e.g., Kirschner, Sweller, and Clark 2006). However, there is no reason for these active learning approaches to be implemented with insufficient guidance, since many teachers realize that effective learning requires intervention to keep students on task and ensure a clear focus on the course's particular educational objectives. Therefore, even researchers who are critical of discovery learning are more supportive when the approach includes a more active role for teachers to intervene (Mayer 2004). In this context, teachers and students form a dynamic partnership and share the responsibility for learning processes and outcomes.

The purpose of this article is to show teachers of English for communicative purposes how guided, corpus-aided discovery teaching can generate their students' active learning. To be successful, this method requires teachers to provide careful guidance while their students research, discover, and reflect on the grammatical and sociolinguistic aspects of English that they are in the process of acquiring.

Why incorporate corpus-aided discovery learning?

A *corpus* is a large collection of naturally occurring texts gathered from, in this case, users of the English language in a wide variety of communicative scenarios. A corpus may include natural spoken, written, computer-mediated, spontaneous, or scripted discourses in diverse contexts. These collected discourses represent a variety of genres, such as everyday conversations, fictional novels, academic texts and lectures, business meetings, radio and television news broadcasts, and radio talk shows, to name a few. In recent years, corpus data have been acknowledged as (1) resources that provide descriptive insights that are relevant to how people use language, and (2) tools that can directly affect learning and teaching processes. Central to the corpus-aided method used in the classroom is data-driven learning (Johns 1994), which encourages learners to take on the role of language researchers who are engaged in discovery learning (Gavioli 2001).

Many corpora are available online and can be accessed by researchers, teachers, and students to analyze how words, phrases, grammatical structures, and idioms are used in a large compilation of meaningful contexts. This can be extremely helpful to students, as they can notice word frequencies, the different forms a word can take, and common and uncommon usages of a word or phrase. English learners can discover how people use language in the real world, in various forms, and at various levels of formality. Students also can see how language fulfills different speech functions across various contexts.

Utilizing an English corpus to analyze authentic written and spoken texts also provides students with a powerful tool to *learn how to learn* as they work independently or collaboratively to observe, analyze, and interpret patterns of language use. In addition, corpus-based learning promotes the transferability of language skills and language-learning strategies (Hunston 2002; Sinclair 2004).

English language corpora online resources

There are several corpora resources on the Internet that English language learners can search and analyze. Although some sites require a subscription for unlimited access to their corpus, many offer a limited search at no cost. The corpora are most useful if students learn the different options that allow them to narrow and focus their searches, which include codes and queries to search for language items by part of speech, speech event, and speaker, among other categories. The three sites described below are examples of online corpora that contain a wide variety of language data from diverse texts and numerous contexts.

1. The *Collins WordbanksOnline* English corpus currently holds about 56 million words from written and spoken texts, including newspapers, books, magazines, websites, and TV and radio shows. Searchers can specify American or British texts, and 40 lines of results are available as a free demonstration. Information on how to access and search the corpus is available at: www.collins.co.uk/corpus/CorpusSearch.aspx.
2. The *British National Corpus* (BNC) contains 100 million words taken mostly from British written English, including newspapers, magazines, academic texts, school essays, and fiction; it also includes spoken texts from business and government conversations, as well as radio shows. Fifty lines of results are available for free as a demonstration. Information on how to access and search the corpus is available at: www.natcorp.ox.ac.uk.
3. The *Michigan Corpus of Academic Spoken English* (MICASE) currently holds over 1.8 million spoken words as they are used in a large university setting, including language spoken by faculty and staff, and by students with various English proficiency levels and native languages. Searches can be narrowed by many variables, including (1) type of speaker (faculty, staff, students, etc.), (2) type of speech event (seminars, lectures, student presentations, etc.), and (3) discipline (Women's Studies, Sociology, English, etc.). Information on how to access and search the corpus at no cost is available at: http://lw.lsa.umich.edu/eli/micase/index.htm.

In addition to the corpora above, there are free online user text concordances such as the one available at www.lextutor.ca/concordancers/text_concord, which allows users to paste in their own texts and create a corpus that can be analyzed.

Using corpora to generate active learning

Before students step into the world of corpus-aided learning, they must know how to use a corpus, with its built-in search tools, to obtain information about lexico-grammatical associations, collocations, and word frequencies in different contexts. Since most corpora contain detailed instructions on how to conduct specific searches, advanced learners can experiment with different searches to get the feel for the tools during the familiarization stage.

Some learners, however, may feel overwhelmed by the complexity of the patterns and usages that emerge from the output of a corpus analysis, a complexity stemming from the myriad ways that linguistic items and structures vary across genres, users, and contexts. Taking this into account, it is easy to see that an unguided approach to corpus-aided discovery learning is unlikely to result in effective learning during the early stages. For this reason, the limited results for non-subscribers to the *Collins WordbanksOnline* corpus and the *British National Corpus*, for example, are not a drawback because beginning language learners will not be overwhelmed by pages and pages of examples. The limited random examples are sufficient to promote self-regulated learning and enable students to gradually take on the role of a researcher in their own language learning endeavors. This will occur provided that teachers are prepared to guide students on how to conduct searches in a precise way that leads to the discovery of relevant, meaningful examples for analysis. In other words, during the first few weeks of working with corpora, students will need to learn how to use the tools to help them identify, analyze, and interpret observations (Bernardini 2001).

Languaging to promote corpus-aided discovery

Swain (2006) defines *languaging* as the process by which learners produce and use language as they attempt to understand, solve problems, create meanings, and make sense of their interpretations to themselves and to others. Since Krashen originated the Input Hypothesis, which emphasizes that providing comprehensible input leads to second language acquisition, many researchers have broadened the scope of the hypothesis (Long 1996; Pica 1994; Swain 1995). For example, Swain (1995) argues that output is essential to learning and stresses that languaging promotes an awareness of how a language works and pushes learners to process language more deeply than does input alone.

In recent years, researchers in second language acquisition (Lapkin, Swain, and Smith 2002; Swain 2005) have been exploring empirically how languaging can be a source of second language learning. The idea that people operate with mediating tools, such as languaging sessions and language logs, originates with Vygotsky's (1986, 1987) sociocultural theory of mind, which has received much attention from those studying second language teaching and learning (Lantolf 2000, 2003; Lantolf and Thorne 2006). Smagorinsky (1998) argues that "the process of rendering thinking into speech is not simply a matter of memory retrieval, but a process through which thinking reaches a new level of articulation" (172–73). This suggests that languaging spurs development by serving as a vehicle through which thinking is articulated and experience is reshaped. As Wells (2000) points out, "one of the characteristics of utterance, whether spoken or written, is that it can be looked at as simultaneously process and product: as 'saying' and as 'what is said'" (73). Wells (2000) suggests that it is often in the effort of "saying" that a speaker "has the feeling of reaching a fuller and clearer understanding for him or herself" (74). Furthermore, verbalized thoughts, whether spoken or written, become available as objects about which questions can be raised and answers can be explored with others or with oneself. In other words, languaging is a process that creates an audible or a visible product about which one can language further, and while speaking and writing, the learner may reach a new or deeper understanding.

The concept of *noticing* also explains why languaging improves learning. Schmidt's (1993) Noticing Hypothesis states that "what must be attended to and noticed is not just the input in a global sense but whatever features of the input are relevant for the target system" (209). It is the learner's attending to and noticing linguistic forms in the input that affords intake and thus is a necessary condition for second language acquisition (Schmidt 1995). Languaging is a critical activity because it enables students to clarify and acquire important elements of the language that they "notice" in the input.

Two methods of corpus-aided discovery

Through classroom research I have found that students learn to use corpus-aided discovery effectively during weekly sessions when they have questions they want to **answer** that are meaningful to them (Huang 2008). In addition, after testing both guided and unguided approaches, I found it worked better to use a guided, discovery learning approach that followed a process of (1) exploring a corpus and keeping a language log, (2) languaging about the discoveries, and (3) presenting the results. This process can be implemented by following the two examples that are described next.

Example 1: Assigning questions

This approach is appropriate for both intermediate and advanced learners. To begin the process, the teacher assigns questions for individual corpus exploration that focus students' attention on specific aspects of linguistic features. Figure 1 shows a post-exploration language log excerpt from a student who did not know what a subject or an auxiliary verb was during the first few weeks of class. By week 6, after using the non-subscriber editions of the *WordbanksOnline* corpus and the *British National Corpus*, the student could articulate about (1) the use of genitive (*my*) and accusative (*me*) subjects functioning as complements, as in "My manager is relying on **my/me** being there to carry out the project," and (2) the use of nominative (*she*) and accusative (*her*) forms functioning as adjuncts, as in "I consulted with Leslie, **she/her** being the Chair of the department."

Figure 1: Language Log for Assigned Questions

This week I started with the *WordbanksOnline* English corpus, since I don't think I have ever started off with this particular corpus before. First I decided to try the gerund-participial as complement construction. The entry "IN+me+being" (IN being the abbreviation for a preposition) produced approximately 40 results, while "IN+my+being" only had 26 results in the corpus pop-up window. From the results mentioned above, it seems that the informal accusative *me* is currently used more frequently in the English language.

Next I entered the gerund-participial construction as an adjunct. There was one match for "NN+he+being" and zero matches for "NN+him+being" (NN being the abbreviation for a common noun). These results were consistent no matter which nominative and accusative pronoun I replaced *he* and *him* with. Although the results I received were few, they did verify research claiming that when a gerund-participial is in adjunct function, the accusative subject is acceptable, although it is much more informal than the nominative form.

In an effort to conduct further testing, I decided to test the two constructions in the *British National Corpus*. There were 117 solutions for the accusative subject phrase "me+being" and 97 solutions for the genitive subject phrase "my+being." Similar to the results from *WordbanksOnline*, these findings illustrate that the accusative subject is used more commonly in the gerund-participial as complement construction. Next I entered the gerund-participial as adjunct phrases. The nominative subjective query "he+being" had 46 solutions, while the accusative subject query "him+being" had a startling 204 solutions! These solutions, indeed, varied from the *WordbanksOnline* results. Based on the *British National Corpus* information, it seems the conclusion can be made that, although the accusative is markedly informal, it is still used in the gerund-participial as an adjunct construction in British English.

Figure 2 is a language log excerpt that demonstrates what another learner who had never used a computer for course assignments was able to accomplish within a few weeks after exploring the usage of *stranded* prepositions ("This is the article that the author was referring **to**") vs. *fronted* prepositions ("This is the article **to** which the author was referring").

Figure 2: Language Log for Assigned Questions

Stranded prepositions: Samples that I looked up in the corpora included (1) *give it to*, as in "Who did you give it to?"; (2) *put it in*, as in "What did you put it in?"; (3) *come from*, as in "Where did this come from?"; and (4) *referring to*, as in "What are you referring to?"

In checking the three corpora recommended, I see that stranded prepositions seem to be most often used in an interrogative situation. The corpora do not seem to have recorded many interrogative situations, even though I would expect to find this usage in most standard spoken English. I did find a few examples, as in this one from *WordbanksOnline*: "Now, where had that memory come from? He groped for more." I also found this example in MICASE: "Where'd Bollinger get the money to give to the athletic department? Where'd that come from?" The *British National Corpus* seemed to have the most obvious examples of stranded prepositions. For example: "Where does it come from?" and "If it is the meat of a buffalo, what precise part of the animal does it come from?"

Fronted prepositions: I used the following sample search phrases: (1) "to whom did"; (2) "in what did"; (3) "from where did"; and (4) "to what are you." Fronted prepositions seem not to be used as often as stranded ones. The sample "to whom did" seems to be the most commonly used of the prepositional phrases I tested. Each search found many examples of the use of *to* in fronted prepositional phrases.

My choice of "in what did"—as in "In what did you put the sausages?"—was found only in a couple of cases. When these examples were read, it was obvious that the original speakers were either overcorrecting or being extremely ungrammatical in their speech.

My preliminary conclusion is that stranded prepositions are currently used much more often than fronted ones.

Both of these students demonstrated their ability to observe, analyze, and generalize in their language logs. The purpose of using corpus data is not to model a target linguistic usage typically used by native English speakers but to promote "noticing" and develop a sharper awareness of features in spoken or written texts in different contexts. The exploration of the corpora promoted the "noticing" of specific grammatical features and indicates that these students achieved an understanding of complex constructions in a short period of time.

As students develop their capacities and competencies in using corpora—such as posing queries, analyzing patterns, and interpreting results—teachers can give them increasing levels of freedom to initiate their own questions for discovery. Some other examples of topics explored by learners include (1) the use of a plain verb form or a primary verb form in subordinate subjunctive clause constructions; (2) the use of subjunctive constructions in expressing unlikelihood or doubt; (3) the use of subjunctive mandative, should mandative, and covert mandative constructions (e.g., "It is vital that **he be** [subjunctive mandative] **/ should be** [should mandative] **/ is** [covert mandative] informed immediately"); and (4) the use of "different from" and "different than."

Example 2: Examining metadiscourse

This approach, suitable for highly advanced learners, implements corpus-aided discovery learning through the examination of *metadiscourse*, which is usually an introductory clause that comments on the discourse itself and that writers and speakers use to do such things as convey intentions ("In summary, …"), state conclusions ("Therefore, …"), and establish the degree of certainty in their statements ("Perhaps…"). For example, teachers can help students learn about how the degree of certainty is expressed by directing them to analyze the corpus data for linguistic devices such as the use of hedges (e.g., expressions like "as far as I know," indicating that the speaker is being cautious about the truth of a statement), emphatics (e.g., the use of certain adverbs such as *certainly*), and attributors or evidentials (e.g., citing credible sources). Some useful questions to facilitate this exercise and provide ideas for students to write about in their language logs include:

- How do such rhetorical devices affect the strength of the statements?
- How are personal pronouns used in relation to epistemic verbs (e.g., *think, believe, know*) that convey degrees of certainty?
- In what ways do these rhetorical devices strengthen the force of commitment to an argument or weaken a claim by hedging its generalizability?
- How do speakers and writers use personalized and impersonalized language to modify their assertions?

Finally, to reinforce what students have discovered, the teacher can ask students to evaluate additional research articles in terms of the probability or truth of the propositional content that a speaker or writer wishes to express (Vande Kopple 1985; Hyland 1998).

Languaging and presentation of discoveries

Weekly verbal group languaging sessions are critical to the guided, corpus-aided discovery methods described above. During such sessions, students are asked to verbalize their experience with the corpora and what they have written in their language logs; this languaging reinforces their "noticing" and helps them better understand the grammatical concepts and usages they have encountered while exploring the corpora.

The teacher also can set aside fifteen minutes for small group languaging sessions each week (in class or outside of class) to let students discuss their discoveries. Students can then present synopses of their key discoveries by posting them on a website as text or sound files, thus making them available to other students in the course and perhaps generating further discussion.

Benefits of corpus-aided discovery

When the teacher acts as a facilitator and guide, analyzing corpora supports active learning in the following ways:

- *Generativity*. Students actively use language to develop and generate knowledge in a community of learners through weekly languaging sessions.
- *Relevance*. Students integrate discourse and language on topics that interest them.
- *Engagement*. Students engage in and present discoveries resulting from individual and group efforts. Many find that unexpected and serendipitous discoveries motivate their learning. As one student explained in an exit survey: "I really enjoyed each exploration throughout this semester. They gave me a chance to gain a better understanding of specific grammatical concepts covered in class that I may have otherwise just accepted as confusing. I also learned a lot from the explorations where I obtained information contrary to what I had expected. I thought these exercises were a great way to stimulate interest in the class and an effective teaching method."
- *Autonomy*. Students develop independence and ownership of the discovery process while learning how to observe language and make generalizations, as is demonstrated through their language logs and languaging sessions.
- *Integration*. Through the ongoing discovery and sharing of ideas and concepts, students learn to notice and become critical of their own linguistic choices. Many students notice aspects of their own and others' language use that they had been unaware of, including realizations that they actually do use certain constructions they had previously denied using.

Conclusion

Recent technological improvements have made corpus-based learning methods that actively engage learners' ability to analyze language an increasingly workable option for teachers. Corpora can easily be adapted as sources of linguistic insight and as stimuli for active learning and student engagement. Even though more empirical research into the effectiveness of corpus-aided discovery learning is needed, it is supported by theory, and many instructors report that using corpora improves student interest and learning (Weber 2001; Foucou and Kübler 2000). However, corpus-learner interactions are not replacements for learner-to-learner and teacher-learner interactions. Teachers have a special role in corpus-aided learning and must facilitate access to the online corpora, help students pose appropriate questions, and ensure that the focus remains on the learning objective.

The integration of languaging sessions plays a critical role in corpus-aided discovery learning. Through languaging, verbalized thoughts, whether spoken in a group or written in language logs, become available as objects about which additional questions can be raised and possible answers can be explored. In other words, languaging becomes a process that creates a visible or audible product about which students can language further, and while languaging, the learners may reach a new, broader, or deeper understanding of language features and language use in various types of discourse. Above all, the research skills that students appropriate through this process will likely benefit them in the years to come.

References

Bernardini, S. 2001. 'Spoilt for choice': A learner explores general language corpora. In *Learning with corpora*, ed. G. Aston, 220–49. Houston: Athelstan.

Foucou, P., and N. Kübler. 2000. A web-based environment for teaching technical English. In *Rethinking language pedagogy from a corpus perspective: Papers from the third international conference on teaching and language corpora*, ed. L. Burnard and T. McEnery, 65–73. Frankfurt, Germany: Peter Lang.

Gavioli, L. 2001. The learner as researcher: Introducing corpus concordancing in the classroom. In *Learning with corpora*, ed. G. Aston, 108–37. Houston: Athelstan.

Huang, L. S. 2008. Languaging grammatical concepts using a corpus-based approach. Paper presented at the annual symposium of Teaching and Research Using Technology in the Humanities, University of Victoria, British Columbia.

Hunston, S. 2002. *Corpora in applied linguistics*. Cambridge University Press.

Hyland, K. 1998. Exploring corporate rhetoric: Metadiscourse in the CEO's letter. *Journal of Business Communication* 35 (2): 224–45.

Johns, T. 1994. From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In *Perspectives on pedagogical grammar*, ed. T. Odlin, 293–313. Cambridge University Press.

Kirschner, P. A., J. Sweller, and R. E. Clark. 2006. Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist* 41 (2): 75–86.

Lantolf, J. P., ed. 2000. *Sociocultural theory and second language learning*. Oxford University Press.

Lantolf, J. P. 2003. Intrapersonal communication and internalization in the second language classroom. In *Vygotsky's educational theory in cultural context*, ed. A. Kozulin, B. Gindis, V. S. Ageyev, and S. M. Miller, 349–70. Cambridge University Press.

Lantolf, J. P., and S. L. Thorne. 2006. *Sociocultural theory and the genesis of second language development*. Oxford University Press.

Lapkin, S., M. Swain, and M. Smith. 2002. Reformulation and the learning of French pronominal verbs in a Canadian French immersion context. *Modern Language Journal* 86 (4): 485–507.

Long, M. 1996. The role of the linguistic environment in second language acquisition. In *Handbook of second language acquisition*, ed. W. C. Ritchie and T. K. Bhatia, 413–68. San Diego: Academic Press.

Mayer, R. E. 2004. Should there be a three-strikes rule against pure discovery learning?: The case for guided methods of instruction. *American Psychologist* 59 (1): 14–19.

Pica, T. 1994. Research on negotiation: What does it reveal about second-language learning conditions, processes, and outcomes? *Language Learning* 44 (3): 493–527.

Schmidt, R. 1993. Awareness and second language acquisition. *Annual Review of Applied Linguistics* 13:206–26.

——. 1995. Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In *Attention and awareness in foreign language learning*, ed. R. Schmidt, 1–63. Honolulu: University of Hawaii at Manoa.

Sinclair, J. M. 2004. *How to use corpora in language teaching*. Amsterdam: John Benjamins.

Smagorinsky, P. 1998. Thinking and speech and protocol analysis. *Mind, Culture, and Activity* 5 (3): 157–77.

Swain, M. 1995. Three functions of output in second language learning. In *Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson*, ed. G. Cook and B. Seidlhofer, 125–44. Oxford University Press.

——. 2005. The output hypothesis: Theory and research. In *Handbook of research in second language teaching and learning*, ed. E. Hinkel, 471–84. Mahwah, NJ: Lawrence Erlbaum.

——. 2006. Languaging, agency and collaboration in advanced second language learning. In *Advanced language learning: The contributions of Halliday and Vygotsky*, ed. H. Byrnes, 95–108. London: Continuum.

Vande Kopple, W. J. 1985. Some exploratory discourse on metadiscourse. *College Composition and Communication* 36 (1): 82–93.

Vygotsky, L. S. 1986. *Thought and language*. ed. A. Kozulin. Cambridge: MIT Press.

——. 1987. *The collected works of L. S. Vygotsky, Volume 1: Problems of general psychology, including the volume thinking and speech*. ed. R. W. Rieber and A. S. Carton. New York: Plenum.

Weber, J. 2001. A concordance- and genre-informed approach to ESP essay writing. *ELT Journal* 55 (1): 14–20.

Wells, G. 2000. Dialogic inquiry in education: Building on the legacy of Vygotsky. In *Vygotskian perspectives on literacy research: Constructing meaning through collaborative inquiry*, ed. C. D. Lee and P. Smagorinsky, 51–85. Cambridge University Press.

LI-SHIH HUANG, an Assistant Professor in Applied Linguistics and a Learning and Teaching Scholar at the University of Victoria, has over eight years of instructional and curriculum design experience in EAP, ESL, and EFL at the university and graduate levels. She has received TESOL's Award for Excellence in the Development of Pedagogical Materials.