Genre and corpora in the English for academic writing class

Viviana Cortes, Iowa State University

Researchers and instructors in the field of English for Specific Purposes (ESP) and English for Academic Purposes (EAP) have developed an interest in the use of corpora (Flowerdew, 2002; Bernardini, 2004; Mishan, 2004; Swales, 2004). Several studies have stressed the advantages of informing classes with well-designed corpora (Tribble, 2002), while others have pointed out the need to pay attention to context in order to lessen the negative impact of the decontextualized nature of corpora used for pedagogical purposes (Widdowson, 1998; Aston, 1999). Flowerdew (2005), in her review of corpus-based and genre-based approaches to text analysis, highlights the importance of the identification of specific genres to be explored in the ESP/EAP class. In another study that reports pedagogical applications of language corpora, Lee and Swales (2006) emphasized the benefits of creating and analyzing corpora in a course designed to help international doctoral students to improve their academic skills.

This article will describe the design and implementation of a corpus-based and genre-based English for academic writing class that was created for international graduate students. The idea underlying the course design was to teach students to analyze a corpus of texts that are similar to ones they are expected to write. This analysis should help them recognize linguistic patterns and organizational conventions frequently used by published authors of research articles in their disciplines. Students' analyses of the corpus make use of techniques connected to Data-Driven Learning, or DDL (Johns, 1991), where students are guided to discover patterns in the language.

Activities that explore language learning as schema-based restructuring (Bernardini, 2004) have students use corpora to observe and analyze linguistic conventions that are well-established in their academic communities. Flowerdew (1993) suggested that when students have to create a text in the target language belonging to a genre they are not familiar with, they should examine similar instances of that genre to try to discover "typical lexico-grammatical and discourse features unavailable in dictionaries or grammar books" (Flowerdew, 1993: 312).

The objective of this new course is to better prepare graduate students to write a research article, a task that they may find challenging but which is necessary to their academic success. The course uses a twofold top-down/bottom-up approach to the analysis of the Research Article (RA). Students read current studies that analye the writing of RAs in different disciplines, particularly studies that use Swales's "Move-schema" (1981, 1990) for this analysis. Students then test these findings in the writing of their own disciplines by exploring a corpus of RAs they collect for the course.

This article will introduce a discussion of key issues that had to be considered prior to and during the design of the new course, offer a detailed description of the course design and implementation, and give examples of course activities and samples of students' work from the course.

Advanced academic writing: The state of the art

International graduate students who come to the United States to complete Master's or Doctoral programs of studies often are very proficient in English. At the university for which this new course was created, a minimum TOEFL score of 213 (230 for Economics) and high quantitative and verbal GRE scores (which may vary across programs of study) are required for students to be admitted into different

graduate academic programs. It is still under debate, however, whether these tests evaluate constructs directly connected to certain academic tasks. That is why at our university, all new international students whose first language is not English are tested for academic skills upon arrival in order to evaluate how good these students are when they complete tasks frequently performed in academic settings.

The battery of tests administered covers academic skills individually. Academic reading, writing, and listening are evaluated through activities that resemble those that students are expected to complete in their daily academic tasks, such as listening and comprehending academic lectures, engaging in academic discussions, writing expository and analytical papers, and reading and analyzing research reports.

Many international graduate students often show a high level of writing skills in English, but their writing sometimes needs some improvement in the use of academic writing conventions and organization. These students are placed in an advanced academic writing class, English 101D, which is a required course.

that a given section of the course could have twenty students who belonged to twenty different disciplines and academic programs made it impractical to create a discipline-specific course

In general, students take this class in their first or second semester at the university. Four sections of this course are offered every semester. Classes usually have up to twenty students from a wide variety of disciplines, such as Engineering, Business, Architecture, Statistics, Design, Biology, and Chemistry, to mention only a few. Students' first languages are also varied, including Japanese, Korean, Portuguese, Russian, Spanish, and Turkish, among many others.

The original course, taught by professors, instructors, and graduate assistants from the TESL/Applied Linguistics program, used to cover a wide variety of genres: formal letters and memos, article and book reviews, conference abstracts, paper proposals, and, most importantly, the research paper. These varied genre analyses resulted in a highly ambitious syllabus. The fourteen weeks of actual instruction, with two eighty-minute classes per week, did not provide enough time to investigate and master each genre. A large segment of the course was

devoted to the investigation of the research report, for which the program focused on the exercises presented on a textbook (Weissberg & Buker, 1990) as well as on specially-designed class materials.

The course was generally well-received by students. They often commented that even though the examples of research reports presented in the textbook were dated and mostly belonged to the humanities, they learned to pay special attention to certain aspects of academic writing. Some students, however, expressed concern about the course materials. Even though they felt the course was good and helped them gain new insights into the writing of research reports, many students still believed that in their disciplines, researchers "do not write like this." The fact that a given section of the course could have twenty students who belonged to twenty different disciplines and academic programs made it impractical to create a discipline-specific course.

The issue of discipline-specificity has been raised by several researchers such as Bhatia (2002), who stated that a genre often presents variation across disciplines. Differences in lexico-grammatical resources and rhetorical strategies are shown by different disciplines when expressing, for example, discipline-specific concepts, knowledge, and modes of conducting and reporting research. Bhatia's claim was consonant with the frequent concerns of the students taking this advanced academic writing class.

The design of a corpus-based/genre-based course

The core of the course has three important elements: a corpus made up of RAs, a user-friendly concordancing program, and a selection of reading materials extracted from studies in Applied Linguistics that report findings about the analysis of the different sections of the RA.

The first sections of corpus-based English 101D used a corpus of RAs specially collected for the class, about 500,000 words from ten disciplines. Journal articles were downloaded and stored elec-

tronically in the laboratory where classes were held. These articles had been deemed by professors in the disciplines to be good models of writing for students to analyze. After the course was taught using these corpora for several semesters, a survey was conducted which reported that students perceived the corpora to be too large to be analyzed in class and sometimes not very representative of their corresponding academic disciplines. Thus, a new approach that had students collect their own corpora was adopted.

During the first week of classes, students consult their advisors and professors for journals that can be accessed as full text through the university library electronic catalog[1] and that are seen as models of good writing. Then, instructors show students how to choose examples of experimental research articles and save them as whole files, trying to preserve as much of their format as possible. Students separate the the files into sections and store them in their own electronic materials folders. The number of papers in each corpus will vary according to the number of journals recommended, but students are instructed to keep no more than thirty texts.

A combined top-down/bottom-up approach

Many corpus-based pedagogical approaches have been criticized because they focus only on bottom-up procedures, i.e., the use of concordances to identify frequent lexico-grammatical features in a genre and to analyze the limited sentences provided by the concordancer. The result is an emphasis on word-level information. Flowerdew (2005) explained that the disadvantages that such a bottom-up approach may bring about can be overcome by analyzing whole texts and adding a top-down view. Rather than just pulling lexico-grammatical features out of context, the analysis would look at these features at the discourse level and note the characteristic patterns and sequences in which they occur.

Drawing on Mishan's (2004) distinction between inductive and deductive data-driven activities, the course uses inductive activities carried out by means of a concordancer specially designed for the course for its bottom-up approach. The top-down approach in the course, on the other hand, is based on deductive activities derived from selected readings

that create a framework of reference for comparison when using the corpus.

Software development: Word Search

The advantages of using concordancing software in the classroom have been highlighted in numerous studies (Dudley-Evans & St. John, 1998; Johns, 1988; Barlow, 2004; Sinclair, 2004). The use of the concordancer in this new course would allow students to work on a bottom-up approach to text analysis, looking for linguistic features that are frequently used by published authors in their disciplines. When deciding which concordancing software would be used in the class, three important factors were taken into consideration:

1. The program should be extremely user-friendly. Some commercially available concordancers yield so much information that the output can be a bit overwhelming for a course like this one.

2. The concordancer should provide large stretches of co-text around the searched item (around sixty words, thirty before and thirty after the searched word or expression). In this way, students will be able to have a clear picture of the use of any linguistic feature in context.

3. The program used in the new course should allow searches in a corpus of text files as well as in files saved in other formats (e.g., Word document files).

The program designed for the course, Word Search, is freeware designed by the researcher using Borland Delphi Studio 7. Word Search allows searches for words and expressions of up to four words as well as "wildcards." The wildcards allow the user to search both base forms (lemmas) and derived forms. For example, the wildcard "*" used with *introduce* as *introduc** would bring back not only the word *introduce*, but also *introduced*, *introduction*, and other related forms. Word Search allows users to search as many text or Word files as desired and to save their findings (searched items + co-texts) in a new file.

The program shows the number of matches (tokens) that could be found in the selected corpus, and a special window shows the files that the program is processing. The program also shows a menu
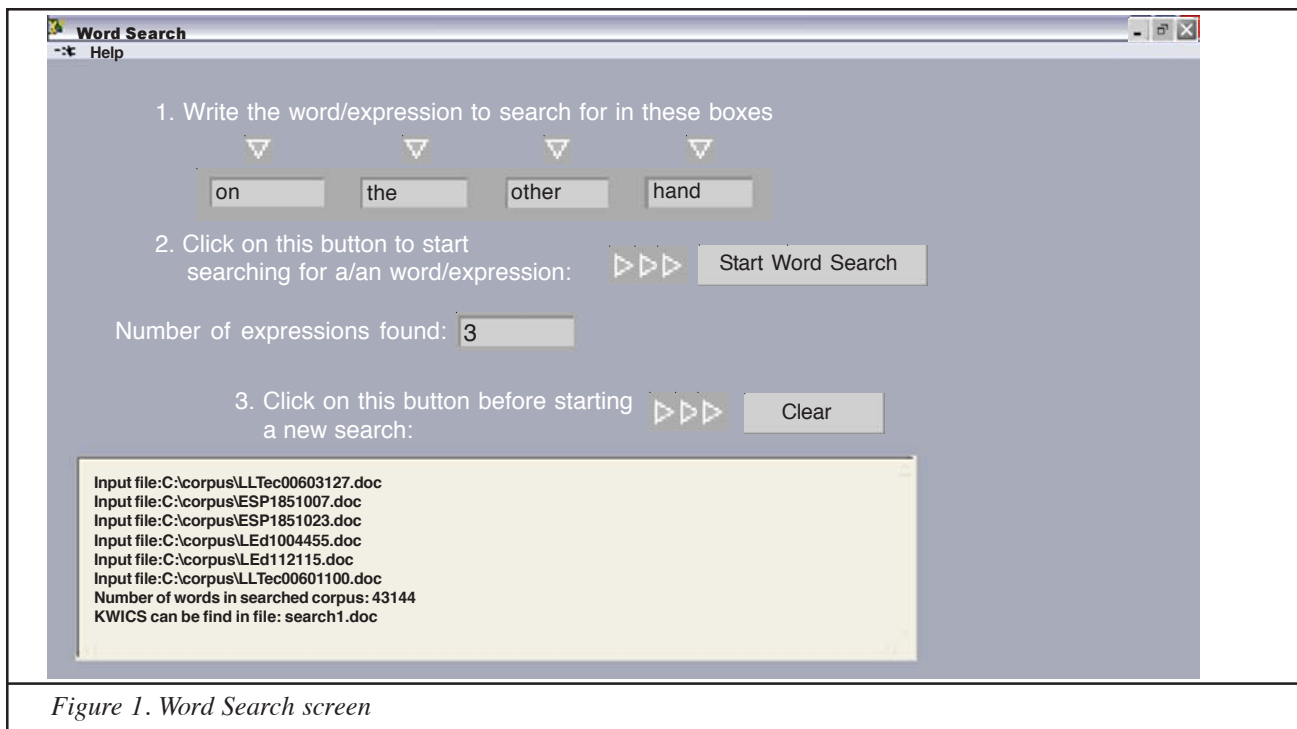
Figure 1. Word Search screen

with a help file that provides detailed steps on how to conduct a word search and how to retrieve the tokens in context. Figure 1 shows a snapshot of the Word Search screen after having searched for the expression *on the other hand*. The screen indicates the files the program opened to search for this expression, the number of tokens found, and the name of the file in which the identified hits can be found.

Word Search was piloted prior to the start of the course with a group of six graduate students. The result of this mini-study showed that it took these students between thirty and ninety minutes to become familiar with the corpora and the use of the concordancer. Thus, after students collect their corpora, they spend a class period working on corpus management activities to get acquainted with the concordancer and the methodology.

Analyzing the RA: Reading materials selection

In order to help students explore the corpus to formulate their own hypothesis on disciplinary writing, the course takes a top-down approach by means of deductive activities. Most of these activities are based on a selection of readings that contain excerpts from RAs in Applied Linguistics, some of them using corpus-based methodologies. The authors explore disciplinary writing and report their findings with examples taken from published written work. The course uses Swales' (1981, 1990) seminal work on Move-schema in introductions as a core reading, enabling students to note basic patterns in the flow of information in an introduction. Table 1 shows an analysis of introductions in academic papers, adapted from Swales (1981:22). The rest of the reading selections apply similar methodologies to investigate the other sections of the RA.

The reading selections are generally exploited by means of two types of exercises. First, students are expected to answer several reading comprehension questions in writing. These questions help students get deeply involved in the content of the readings and the findings of those research studies. The second type of exercises encourages students to test their hypotheses through the analysis of the specific sections of the articles in the corpus. The exercises in Table 2 correspond to these types of activities.

For some of the exercises, the use of the concordancer is suggested. For other exercises, students are instructed to skim, scan, or read through various articles or article sections on their computer screens in order to test whether the claims reported

Table 1. A Possible Structure for a Major Type of Article-Introductions

Move 1	Establishing the Field
	a. showing centrality
	• by interest
	• by importance
	• by topic-prominence
	• by standard procedure
	b. stating current knowledge
	c. ascribing key characteristics
Move 2	Summarizing Previous Research
	a. strong author-orientation
	b. weak author-orientation
	c. subject orientation
Move 3	Preparing for Present Research
	a. indicating a gap
	b. question-raising
	c. extending a finding
Move 4	Introducing Present Research
	a. giving the purpose
	b. describing present research
	• By *this/the present* signals
	• By move
	• By switching to first person pronoun

Adapted from Swales, 1981, p.22

Search) or the following verb in the past, but present perfect.

We can see that all authors in my discipline (Agronomy) referred to previous research following a subject orientation. All the verbs were in the present perfect. Some verbs were commonly used to report previous studies in my discipline such as: *identify, study, address, and conduct*. Therefore, two more verbs (*address and conduct*) can be added to the list of verbs mentioned by Swales: *suggest, propose, report, show, investigate, find, study, discuss, examine, develop, identify, refine, reveal, stress, summarize, support*. (Agronomy)

In addition, a final exercise for each section of the RA asks students to prepare a written report in which they present the findings of the corpus exploration they conducted in class and compare their findings with those reported in the reading selections.

Course implementation

The class is taught in a computer lab that has twenty computer stations, a projector, and a screen on which instructors show class presentations, introduce new materials, and model exercises. Upon arrival in the classroom, students are instructed to log

in the selected readings are reflected in the writing of their disciplines as represented in the corpus. In this way, both top-down and bottom-up analyses are conducted on the texts. The following excerpts have been extracted from the answers provided by a student in the course to the exercise presented in Table 2:

> The way the authors signal that they are about to start the summary of previous research of the introduction is pretty constant in my discipline, Agronomy. In four out of five introductions (because the fifth didn't have Move 2) the authors used the word 'research' or 'studies' followed by a verb in present perfect tense (4 examples are bellow this paragraph). None of them used the word 'first' (I used the program Word

Table 2. Sample exercises.

Writing introductions. Move 2: Description of Previous Literature (DPR) Exercise 2 – Exploring the corpus

A. "Summarizing previous research" in introductions: Looking for examples in your discipline

For this exercise, you will have to read several introduction sections of articles in your discipline from the RAC. You will also need to use Word Search.

Look for examples of ways in which the authors signaled that they are about to start the summary of previous research or the literature review section of the introduction.

B. How do authors in your discipline reference previous research?

Look for ways in which previous research is referenced in articles in your discipline. Do authors prefer a subject orientation or an author orientation?

Write a short response to these issues.

In addition, make a list of verbs of communication often used in your discipline when reporting previous research.

Table 3. Course schedule

Week 1:	Diagnostic Test. Introduction to the course and materials. Corpus collection I: contacting advisors. Identifying journals in electronic library.
Week 2:	Corpus Collection II: Downloading papers from electronic library. Corpus organization.
Week 3:	The research paper sections. RA Introductions. Moves 1 and 2.
Week 4:	RA Introductions. Moves 3 and 4.
Week 5:	Language focus exercises on citations. Group discussion. Preparatory work on final report for Introductions.
Week 6:	The Methods Section. Moves and communicative purposes.
Week 7:	Language focus exercises on materials description. Group discussion. Final report on the Methods Section.
Week 8:	Student Conferences.
Week 9:	Communicative Purposes in results sections. Moves in results sections.
Week 10:	Language focus for reporting results. Reporting Results. Final discussion and report on Results sections.
Week 11:	Discussions and Conclusions.Moves in discussion sections. Communicative purposes in discussions and conclusions.
Week 12:	Group discussion. Final report on Discussions and Conclusions preparation.
Week 13:	Writing RA Abstracts.Moves in Abstracts.
Week 14:	Language focus for abstract writing. Final report on Abstracts.
Week 15:	Reading week. Student Conferences.
Week 16:	Finals Week. Final project due.

on to the private environment created for the course. In this environment they can find a folder called Discussion, in which they save the folders containing their corpora and class materials. Class and home exercises are placed in a different folder called Dropbox. Once students 'drop' their exercises in that folder, only instructors can move the files back to their authors' folders; that is, students have no access to each others' files or even their own. Instructors can then correct those exercises in electronic form and save them back in the corresponding student's folder.

Week one is devoted to the administration of a diagnostic test (an evaluation given in order to ensure that students really need this class), instructor and student introductions, and an overview of the course design and class dynamics. Then, students start collecting their corpora, first asking for professors' suggestions as to which journals to focus on and then downloading and organizing articles in their folders. In the second week, after they finish downloading the articles for their corpora, they work on corpus exploration exercises that call for the use of the concordancer or more exploratory analysis of the texts in the corpus. (See Table 3, Course schedule.)

As of week three, after a brief overview of the overall organization of the RA, the course focuses on one section of the article at a time. The class generally starts with a wrap-up of contents presented to students in the previous class, which paves the way

for the introduction of the new class materials. Table 3 shows a progression of course content and written assignments. The instructor presents the highlights of the section of the RA under analysis and offers a detailed explanation. Students then read the corresponding work file that they copy and paste in their folders for the daily activities they will complete. Students go on to engage in the reading comprehension and corpus exploration activities.

At the end of each section of the RA, students work in small discussion groups in which they share their findings and the conclusions that they drew about the writing of that specific section in their disciplines. These discussions provide students with insights into the writing in disciplines other than their own and give them a better framework of comparison for their own analyses. Their final conclusions on the linguistic conventions, organization, and schema of communicative purposes of the sections of the RA in their disciplines are presented in a final report that students write at home and hand in to their course instructors.

In addition to the class activities and final reports, students meet with their instructors during the semester in private conferences. In these conferences, students and instructors discuss writing process issues, research methodology reports, and prospects for the final project. For the final project, students are encouraged to work on the writing of a research paper they might be required to complete

for one of their disciplinary courses. In this way, students have the chance to transfer the findings of their analysis to the writing of their own research reports. The final projects are completed following a drafting technique: students are allowed to re-write their drafts of the different sections of their final papers, taking into account instructors' feedback and corrections in order to improve their writing in their final drafts.

Further considerations and conclusions

All the sections of English 101D currently offered at our university are now corpus-based. Course evaluations for the new course design have been high. Students consistently say that they have acquired knowledge about the writing of RAs in their disciplines that was unknown to them before taking the course. Several questionnaires that students were asked to complete at the end of the initial academic terms revealed high levels of motivation. Students feel the course provided them with tools they can apply to their own writing of RAs and that it also gave them skills to analyze articles they read for their disciplinary courses. In addition, many students ask their instructors for a copy of Word Search to install in their home computers for further text analysis that they could conduct in the near future.

Instructors who previously taught the traditional sections of the course and had the chance to teach the computer-based class felt that the RAs written by students in the new course as a final assignment presented a better handling of academic writing conventions than those that were produced by students who took the former course. These RAs present, in most cases, an organization that reflects the Move-schema students were introduced to in the new course.

How much of the new knowledge on writing RAs that was imparted through this new course actually transfers to students' disciplinary writing is, however, difficult to verify. Longitudinal studies that follow the same students along their academic and professional careers, focusing on the development of their research report production, should be conducted in order to check whether students are making use of these newly-acquired skills throughout the varied research-oriented writing tasks they complete for their academic communities.

Notes

1. That all the texts in the corpus belong to journals to which the University Library subscribes is essential to conform to the Copyright Act (section 18 (f)(4)).

References

Aston, G. (1995). Corpora in language pedagogy: Matching theory and practice. In G. Cook & B. Seidlhofer (Eds.), *Principles and practice in Applied Linguistics* (pp. 257-270). Oxford: Oxford University Press.

Barlow, M. (2004). Software for corpus access and analysis. In J. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 205-221). Philadelphia: John Benjamins.

Bernardini, S. (2004). Corpora in the classroom: An overview and some reflections on future development. In J. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 15-36). Philadelphia: John Benjamins.

Bhatia, V. (2002). A generic view of academic discourse. In J. Flowerdew (Ed.), *Academic discourse* (pp. 21-39). London: Longman, Pearson Education.

Bordland Delphi Studio. (Version 7.0.) [Computer software]. (2004). Imprise Corporation.

Dudley-Evans, T. & St. John, M.G. (1998). *Developments in English for Specific Purposes*. Cambridge: Cambridge University Press.

Flowerdew, J. (1993). An educational, or process approach to the teaching of professional genres. *ELT Journal, 47,* 305-316.

Flowerdew, L. (2002). Corpus-based analyses in EAP. In J. Flowerdew (Ed.), *Academic discourse* (pp. 95-114). London: Longman, Pearson Education.

Flowerdew, L. (2005). An integration of corpus-based and genre-based approaches to test analysis in EAP/ESP: Countering criticisms against corpus-based methodologies. *English for Specific Purposes, 24*, 321-332.

Johns, T. (1988). Whence and whither classroom concordancing? In T. Bongoerts, P. de Haan, S. Lobbe, & H. Wekker (Eds.), *Computer applications in language learning* (pp. 9-27). Dordrecht, The Netherlands: Foris.

Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning. In T. Johns & P.

King (Eds.), *English Language Research Journal 4: Classroom Concordancing*. (pp. 1-16). Birmingham: Birmingham University.

Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available corpora to self-compiled corpora. *English for Specific Purposes, 25*, 56-75.

Mishan, F. (2004). Authenticating corpora for language learning: A problem and its solution. *ELT Journal, 58* (3), 219-227

Sinclair, J. (2004). New evidence, new priorities, new attitudes. In J. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 271-299). Philadelphia: John Benjamins.

Swales, J. (1981). *Aspects of article introductions*. Aston ESP Research Reports No. 1. Birmingham: The University of Aston.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Swales, J. (2004). *Research genres: Exploration and applications*. Cambridge: Cambridge University Press.

Tribble, C. (2002). Corpora and corpus analysis. In J. Flowerdew (Ed.), *Academic discourse* (pp. 131-149). London: Longman, Pearson Education.

Weissberg, R., & Buker, S. (1990). *Writing up research: Experimental research report writing for students of English*. Englewood Cliffs: Prentice Hall.

Widdowson, H.G. (1998). Context, community and authentic language. *TESOL Quarterly, 32*, 701-716.

Viviana Cortes is an Assistant Professor in the English Department at Iowa State University, where she teaches corpus-based English grammar, discourse analysis, and advanced academic writing. Her research interests are related to grammar and to the investigation of frequently-occurring word combinations, such as lexical bundles, in different academic registers.