A Case Study of Corpus-Informed Online Academic Writing for EFL Graduate Students

WEN-MING HSIEH
The Pennsylvania State University
HSIEH-CHIN LIOU
National Tsing Hua University

ABSTRACT
Research articles (RAs) have been recognized as a distinct genre in the English-using discourse community because of their unique writing conventions. Despite the great number of studies on the analysis of the textual or phrasal aspects of abstracts of RAs, few have been transformed into actual teaching materials for EFL graduate students. The purpose of this study is to extend the results of abstract analysis to the development of online course content for the explicit teaching of abstract writing in Applied Linguistics. Results of the comparison of patterns found in the corpora of 50 abstracts of RAs published in journals and 50 abstracts of papers given at conferences—written by novice, nonnative writers—to patterns in a learner corpus formed the core of the online course content, along with principles of L2 writing and CALL task design. An academic concordancer and a synchronous peer review tool were incorporated into the Moodle-based course. Formative evaluation of the unit on abstract writing was conducted with 35 students. The abstract-writing unit was then incorporated into an academic writing course in which the analysis of students’ writing samples demonstrated some improvement in their writing. The study suggests that corpus-informed online instruction that includes learner corpus insights can facilitate EFL academic writing.

KEYWORDS
English for Academic Purposes (EAP), Learner Corpus, Corpus-informed Online Materials Development, Research Article (RA), Abstract Writing

INTRODUCTION
In English as a foreign language (EFL) contexts in which English is taught only in schools as a subject and is not available as input outside the classroom, most learners are often very much unprepared to write acceptable research articles (RAs) for publication. However, both scholars and novice writers (or student writers in the case described here) face the need to publish either as a degree requirement or to advance their careers. In spite of the abundant text analysis on RAs, few have been transformed into actual teaching materials for academic writing to meet EFL learners’ specific needs.

Academic writing needs to be explicitly taught to those who want to participate in the academic community (Hyland, 2000; Lau, 2004; Swales, 1990). Simple exposure to RAs in the professional discipline does not guarantee acquisition of relevant writing conventions. Helping students “notice” and become aware of the different contexts of academic disciplines and discourse functions used in those disciplines is an important task for teachers (Cortes, 2004). Further, online teaching has become an attractive option for teachers to harness the power of advanced internet and computer technologies for academic-writing instruction.
The purpose of this study was to extend the research results of academic text analysis by developing research-informed online course materials for the explicit teaching of RA abstract writing for EFL graduate students in applied linguistics. To achieve this goal, three steps were taken. First, small corpora of abstracts of RAs published in applied linguistics journals and abstracts of papers given at local applied linguistics conferences—written by novice, nonnative writers—were compiled, and the patterns of information structures and phraseology used in those abstracts were identified. Second, a comparison of the published journal abstracts and the conference paper abstracts was made to determine which particular aspects of the abstracts seem to deserve teachers’ special attention. Third, based on the results of these analyses, we developed online course materials for explicit instruction on the selected aspects for EFL graduate students and collected formative evaluation data to refine the materials. We then integrated the online materials into a writing course and investigated students’ writing samples after they used the online materials. The web-based academic writing course was constructed on the Moodle platform enhanced by an academic concordancer and a synchronous peer review tool. The evidence obtained in the study suggests that corpus-informed online instructional materials demonstrate substantial potential for EFL learners to acquire necessary writing skills and knowledge to meet their professional needs.

STUDIES OF RA ABSTRACTS

A large number of studies conducted on EAP have contributed to our understanding of various aspects of the academic genre (e.g., Gledhill, 2000; Martin, 2003; Ruiying & Allison, 2003, 2004; Samraj, 2002; Thetela, 1997; Vassileva, 2001). Researchers who study the information structures of RA abstracts often use the term “move,” which can be one to several sentences, as the basic unit for conducting analyses. Dos Santos (1996, p. 485) defined move in the following way:

As genres are purposed, staged activities, the move was chosen as the unit of analysis. A move is to be considered as a genre stage which has a particular, minor communicative purpose to fulfill, which in turn serves the major communicative purpose of the genre.

Swales (1990) and Hyland (2000) have identified sets of RA abstract information structures such as the Introduction-Purpose-Methods-Products-Conclusion pattern. In addition to move analysis, other features of abstracts in RAs in several disciplines have been examined, including both hard and soft disciplines (Lau, 2004; Lores, 2004; Salager-Meyer, 1990, 1992; Dos Santos, 1996). Some researchers analyzed rhetorical structures of abstracts (Lau, 2004; Martin, 2003; Dos Santos, 1996), others investigated verb-tense and modality use in medical journal abstracts (Salager-Meyer, 1992), and still others analyzed the rhetorical organization and thematic structures of RA abstracts in linguistics (Lores, 2004).

Lau (2004) compared RA abstracts written by 50 EFL Ph.D. students in Taiwan to those written by 30 international scholars in journal articles in the life sciences in order to show the factors that can affect structural patterns of abstracts. Lau used a modified version of the abstract structural patterns proposed by Hyland—Introduction-Purpose-Methods-Results-Conclusion—for the analysis of his corpora, substituting Results for Product. After manual coding and analysis of the 80 abstracts, Lau’s results showed that abstracts written by the scholars contained the assumed five moves with the exception of Method. Student abstracts, on the other hand, did not contain all five moves, and the percentage of abstracts with the comprehensive five-move structure was relatively lower than that of the foreign scholars. He found three representative types in students’ shorter abstracts: (a) Background-Purpose-
Results-Conclusion, (b) Background-Results-Conclusion, and (c) Results-Conclusion. Lau provided three possible reasons for the differences. First, it is likely that the students were academically immature so that they failed to show their contribution to their field in a process of problematization which involves describing the situation, the problem, the response, and evaluation. Second, the students were unable to provide contextual information because of their linguistic inadequacy. Third, word limits for abstracts might have prompted the student authors to provide less background information and concentrate more on their findings. Lau suggested the need to familiarize students with various structural patterns. The five-move structural patterns and the problematization process (situation, problem, response, and evaluation) can be useful in the teaching of academic English.

Using an applied linguistics corpus, Dos Santos (1996) examined the features of 94 published RA abstracts and pointed out the apparent mismatch between recommendations in the technical writing textbooks and the ways in which writers compose their articles. Dos Santos found three major genre-specific features: (a) the size of textual space allocated for each move (move balance), (b) the blending of moves into the same statement (move embedding), and (c) the reversed sequence of moves (move reversal).

LEARNER CORPUS RESEARCH

Recent research on learner corpora has increased our understanding about the differences between the target reference corpora and actual learner behavior (Upton & Connor, 2001; Cortes, 2004; Granger, 2003; Hyland & Tse, 2005). Moreover, with the advancement of computer technology, specially designed computer programs are available to aid the traditional manual process of text analysis. Tools like collocation extractors and concordancers can contribute to the research process by helping to locate, search, and compute the frequency of particular patterns.

Only a few studies on learner corpora have been conducted in English for academic purposes (EAP). Upton and Connor (2001) illustrated the usefulness of a multilevel analysis of a computerized genre-specific learner corpus. They computed hand-tagged moves and frequency counts of linguistic features associated with aspects of politeness in business letter writing using a learner corpus and demonstrated the efficacy of a multilevel analysis of a genre-specific learner corpus. In another case, Cortes (2004) analyzed a particular type of frequent word combination—lexical bundles—with the help of computer programs. Cortes defined lexical bundles as “extended collocations, sequences of three or more words that statistically co-occur in a register” (p. 400) and found that lexical bundles have two features: idiomaticity and fixedness. They are idiomatic expressions, yet their meanings are transparent; and their fixedness is determined by the frequency with which the individual words co-occur. Cortes collected 291 published writings by writers and 95 writings by students for corpus analysis. She identified 4-word lexical bundles used by published authors of history and biology and then compared these bundles to those used in three different levels of students’ academic writing: undergraduate lower division, undergraduate upper division, and graduate. Her findings revealed that students rarely or never used many lexical bundles frequently used by published authors and that students at different levels used lexical bundles to convey functions different from those identified in published writings.

Hyland and Tse (2005) examined the frequencies, forms, and functions of evaluative that in 465 published research articles, master’s theses, and doctoral dissertations written by L2 students across six disciplines. Hyland and Tse defined evaluative that-clauses as “a grammatical structure in which a complement clause is embedded in a host super-ordinate
clause to complete its construction and to project the writer’s attitudes or ideas” (p. 124). One such example given by Hyland and Tse is “We believe that more attention should be given to evaluative language.” They used concordance programs and qualitative analysis software for the analysis of the texts. Overall, the widespread use of evaluative that in both expert and student corpora demonstrates that the use of such a construction provides academics with a rich means to present their evaluative statements. One interesting feature uncovered by this study is that student writers exercised considerable control over evaluative that-clauses even though there was no parallel structure in the students’ L1. In sum, the student writers “tended to express affective, as well as only epistemic, meaning, offered greater certainty in their statements, and were more reluctant to make use of their personal voice than their professional counterparts” (p. 137). Based on these results, the authors called for language teachers’ attention to raise students’ awareness of preferred patterns to express different meanings by explicitly teaching their use in real texts.

TECHNOLOGY-ENHANCED ACADEMIC WRITING

Only a few technology-enhanced language-teaching studies have targeted academic writing, especially with regard to the genre of research articles. Hegelheimer (2006) investigated the utility of a designed online grammar resource to improve advanced ESL learners’ writing. He examined nine students’ use of the resource by means of questionnaires, interviews, records of use, and writing samples. His study showed that the students knew how to use the resource and perceived it to be useful for reducing their errors in writing.

Lee and Swales (2006) described an experimental corpus-based EAP course for nonnative-speaking doctoral students. The students were required to compile, and make comparisons of, their own writing with RAs in their fields. This corpus-based approach was regarded as “decentering” because the approach allowed nonnative speakers to make discoveries and consult texts written by various writers instead of relying solely on a single native speaker’s point of view. Empowered by such an approach, the students were able to solve their language questions without the help of a native speaker.

There are several websites dedicated to EAP teaching and learning, and we mention five here because of their usefulness in learning to write academic English. Three websites—OWL (http://owl.english.purdue.edu) UEfAP (http://www.uefap.co.uk/index.htm), and CILL (http://elc.polyu.edu.hk/cill/eap)—provide general guidance and some resources for general academic English. Two other websites focus specifically on the writing of academic research articles. The OWN-NHS website (http://www.engl.polyu.edu.hk/ectr/html/ownintro.htm) was developed to provide faculty and students in the School of Nursing at the Hong Kong Polytechnic University with research-writing support in English (Sengupta, 2003). OWN-NHS has general information regarding top journals, specific information on language-related issues, and useful links. The website also has a journal library, a series of tasks, corpus findings, an annotated bibliography, and language explanations. The NCTU website (http://tech.nctu.edu.tw/webEST/sra/index.htm) not only provides resources for scientific research articles but also contains frequency-based scientific word lists categorized according to scientific research fields based on a corpus of scientific journal articles.

To summarize, the studies by Hegelheimer (2006) and Lee and Swales (2006) investigated the writing of nonnative student writers but did not address the needs of EFL graduate students or students in the field of Applied Linguistics. In addition, in spite of valuable resources in the five websites mentioned above, they do not provide sufficiently specific information about writing conventions for research papers nor contain adequate information on abstract writing. This gap formed the major motivation for the current project.
THE CURRENT STUDY

The study described here links systematic corpus analysis with online materials development for EFL graduate students in applied linguistics and presents the results of a complete instructional process with preliminary assessment data. The sections below describe the corpus analysis, the design of the online abstract-writing materials, the description of the online unit, the formative evaluation of the online materials, and the analysis of students’ writing samples after their use of the materials.

Corpus Compilation and Text Analysis

A total of 100 RA abstracts in the field of applied linguistics (50 journal paper abstracts and 50 conference paper abstracts) were collected for analysis of their rhetorical structures. The journal article abstract corpus (JPA) consisted of 50 research article abstracts selected from five well known international journals (total of 7,770 words): The Modern Language Journal, Language Learning, English for Specific Purposes, Studies in Second Language Acquisition, and the CALICO Journal. The conference paper abstract corpus (CPA) consisted of 50 conference paper abstracts selected from the published proceedings of local conferences (total of 9,233 words). The reason for using abstracts from local conferences was that there were no professional journals published in the country where the study was done. At the same time, however, publication of local conference proceedings has been a practice for more than 20 years in the country. Both the abstracts and the full articles in conference proceedings went through a peer review process that yielded a genre similar to that of journal articles.

We approached the analysis of the two corpora in two stages: (a) manual move tagging and frequency calculation and (b) computer analysis of the texts in the corpora for phraseology and collocations that was complemented by human judgment.

Move distribution

Based on previous literature on the analysis of abstracts (e.g., Lau, 2004; Lores, 2004; Salager-Meyer, 1990, 1992; Dos Santos, 1996), we devised a coding scheme with a six-move structure (Background-Gap-Purpose-Method-Results-Conclusion). Three graduate students and a professor in applied linguistics manually tagged the texts in the corpora. After several rounds of face-to-face discussion, we made spot checks across coders to ensure consistency on the 100 abstracts. We then counted the number of moves and calculated the percentage of the frequency of the moves in each corpus. For example, Background moves appeared 24 times in the 50 abstracts in the JPA corpus, that is, 48%. (It should be noted that the moves we identified in the study corresponded well to the Lau’s (2004) problematization process of situation, problem, response, and evaluation.)

The frequency of use of the Background and Gap moves in both JPA and CPA corpora were not as high as that in the Purpose, Method, Results, and Conclusion moves (see Table 1). The lower frequency of use of the Background moves indicates that the writers presented mainly the results and conclusion of their studies and seemed to assume that readers already had the relevant background knowledge of the issue being discussed. This practice also saved space for more detailed descriptions of the Purpose, Method, Results, and Conclusion moves. Even more extreme, the Gap move appeared in only 8% of the abstracts in both corpora. Thus, in general, it would seem that the Background and Gap moves can be viewed as optional moves, whereas the Purpose, Method, Results, and Conclusion moves should be viewed as obligatory moves.
Table 1
Distribution of Moves in JPA and CPA Corpora (n = 50 in Each Corpus)

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Background</th>
<th>Gap</th>
<th>Purpose</th>
<th>Method</th>
<th>Results</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>JPA</td>
<td>24 (48%)</td>
<td>4 (8%)</td>
<td>40 (80%)</td>
<td>38 (76%)</td>
<td>46 (92%)</td>
<td>42 (84%)</td>
</tr>
<tr>
<td>CPA</td>
<td>29 (58%)</td>
<td>4 (8%)</td>
<td>48 (96%)</td>
<td>48 (96%)</td>
<td>46 (92%)</td>
<td>36 (72%)</td>
</tr>
<tr>
<td>Total</td>
<td>53 (53%)</td>
<td>8 (8%)</td>
<td>88 (88%)</td>
<td>86 (86%)</td>
<td>92 (92%)</td>
<td>78 (79%)</td>
</tr>
</tbody>
</table>

Computer analysis of phraseology and collocation

Liu (2002) indicated that the most serious miscollocation type, based on her EFL learner corpus analysis in which the learners had the same L1 as those in the present study, was the verb-noun (VN) category; thus, we chose verb-noun collocations for analysis in our study. A total of 240 3-word VN collocations were identified by a computer program. From the 240 items, 54 collocations were selected manually according to the researchers’ pedagogical judgment. Among the nine most frequently used 3-word collocations, nearly half occurred in both the JPA and CPA corpora: *investigate effect of, examine/explore effect of, have impact/effect/influence on,* and *play role in.* These 3-word collocations appear to be among the most frequently used in applied linguistics, at least in the corpora analyzed here.

In addition to VN collocations, we also examined the most frequently used 4-word lexical bundles. We used the freeware concordance program, *AntConc* 3.0.1 (http://www.antlab.sci.waseda.ac.jp; Anthony, 2004), to identify 4-word lexical bundles. Since *AntConc* extracts any word that co-occurs with another word according to its frequency, the extracted results were sometimes sentence fragments (e.g., *of this study is, class hereafter class*) or repetitive word use pertaining to certain topics in the abstracts (e.g., *attitudes toward technology integrated*), manual selection was needed to screen irrelevant samples. As a result, a total of 93 4-word lexical bundles were identified in the JPA corpus, 35 of which were regarded useful according to the researchers’ pedagogical judgment. In contrast, 210 4-word lexical bundles were identified in the CPA corpus, of which 55 were chosen for later reference in the instructional materials development. The researchers also observed other commonly used phraseology in the moves in the JPA corpus (see Table 2) and adopted them for inclusion in the content of online materials because they seemed pedagogically important.

Table 2
Common Phraseology Used in the Moves in the JPA Corpus

<table>
<thead>
<tr>
<th>Move</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background</td>
<td>This article presents the results of ... several studies have shown ...</td>
</tr>
<tr>
<td>Gap</td>
<td>But, to date, very few ... ... however ...</td>
</tr>
</tbody>
</table>
Comparison of the JPA corpus and the CPA corpus

To understand the different patterns used by the experienced journal article authors and novice, nonnative writers, we compared move patterns in the JPA and CPA corpora. There were four problematic phenomena in the CPA corpus that attracted our attention (see Table 3): (a) inverted move sequence, (b) missing obligatory moves, (c) disproportional abstracts, and (d) outlining information in a move. Although similar phenomena were observed in both corpora, the CPA corpus generally contained a greater number of instances of them.

Table 3
Number of Four Problematic Phenomena in the JPA and CPA Corpora

<table>
<thead>
<tr>
<th>Phenomenon</th>
<th>JPA Corpus</th>
<th>CPA Corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inverted move sequences</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Missing obligatory moves</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Disproportional abstracts</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Outlining information in a move</td>
<td>7</td>
<td>11</td>
</tr>
</tbody>
</table>

_Inverted move sequence_. Most of the abstracts in the two corpora followed the logical sequence of moves, Background-Gap-Purpose-Method-Results-Conclusion. However, four abstracts in the CPA corpus did not and inverted the Background-Purpose moves (see Table 4).
Table 4
Inverted Move Sequences in the JPA and CPA Corpora

<table>
<thead>
<tr>
<th>Abstract</th>
<th>JPA corpus</th>
<th>CPA corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MLJ03LG</td>
<td>TA04A02</td>
</tr>
<tr>
<td></td>
<td>Background-Method-Purpose-Method-Results-Conclusion</td>
<td>Purpose-Background-Method-Results-Conclusion</td>
</tr>
<tr>
<td></td>
<td>MLJ04PK</td>
<td>TA04P30</td>
</tr>
<tr>
<td></td>
<td>Background-Gap-Purpose-Method-Results-Conclusion-Method-Conclusion</td>
<td>Purpose-Background-Method-Results-Conclusion</td>
</tr>
<tr>
<td></td>
<td>LL04LB</td>
<td>TA04P213</td>
</tr>
<tr>
<td></td>
<td>Background-Method-Purpose-Results</td>
<td>Purpose-Background-Method-Results-Conclusion</td>
</tr>
<tr>
<td></td>
<td>ESP04VC</td>
<td>TA01P223</td>
</tr>
<tr>
<td></td>
<td>Background-Purpose-Method-Background-Method-Results-Conclusion</td>
<td>Purpose-Background-Method-Results-Conclusion</td>
</tr>
<tr>
<td></td>
<td>ESP05HK</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Background-Gap-Background-Method-Results-Conclusion</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SLA04MV</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Background-Purpose-Background-Method-Results-Conclusion</td>
<td></td>
</tr>
</tbody>
</table>

The following is an example of inverted move sequences in an abstract in the CPA corpus (the symbol // marks each move code):

Conference paper abstract 1

//Purpose//The current study intends to investigate washback effects of the Basic Competence English Test for Junior High School Students (BCET) on EFL teaching in junior high school through questionnaire and interview with English teachers.

//Background//Since the BCET provides the criteria to admit junior high school students to senior high schools, its washback effects have become important issues to be addressed.

**Missing obligatory moves.** Two abstracts in each of the JPA and CPA corpora lacked more than one of the obligatory moves (Purpose, Method, Results, Conclusion). The two abstracts in JPA corpus that lacked obligatory moves included two and three moves in total: Results-Conclusion and Background-Purpose-Conclusion. Similarly, the two abstracts in the CPA corpus lacking obligatory moves included only two and three moves in total: Background-Purpose and Background-Gap-Purpose. The abstract below, which contains only the Background and Purpose moves, illustrates the problem.

Conference paper abstract 2

//Background//While Internet has been local college students’ major source of information, how to cultivate students’ critical thinking in analyzing and evaluating a sea of English information online has become an important issue for EFL teachers. Researchers have found that the real need in e-learning is teachers’ understanding of how students learn with computer to develop effective online pedagogy than merely implementing technology into EFL curriculum.

//Purpose//The purpose of the paper aims to explore teachers’ role as online facilitators in discussion forum and practical techniques to facilitate and maximize students and teacher interactions for optimum critical thinking in EFL Internet-supported classroom.
Disproportional abstracts. A disproportional abstract is one in which the author allocated too much of the space to a single move (e.g., Method or Results) and compressed the space for the remaining moves. The defining criterion for this phenomenon was a single move containing more than 50% of the total number of words in the abstract. For example, if the total number of words in an abstract was 149, and the number of words in the Method move was 87, then the percentage of the Method move of the entire abstract was 56% (84/149), and the abstract was marked as disproportional. Disproportional abstracts were found in both JPA and CPA corpora with four instances in the JPA and nine instances in the CPA. In the conference paper abstract below, the author devoted 121 out of 162 words (75%) to the Method move, 15 words (9%) to the Purpose move, and 26 words (16%) to the Conclusion move.

Conference paper abstract 3
//Purpose//This study examines the teaching of economy metaphors in Taiwanese EFL classroom through psycholinguistic experiments.
//Method//Two tasks are carried out towards 24 undergraduate students of National Taiwan University.
//Method//One half of the students are provided input of vocabulary items based on the Lexical Approach (Lewis, 1993, 1997); the other half are given additional input based on the Conceptual Mapping Model (Ahrens, 2001).
//Method//The CMM is itself a bottom-up, linguistic-based model which contributes to the variation of pedagogy as well as the systematic way of comparing metaphors cross linguistically.
//Method//In Task 1, which is a vocabulary test on 10 economic metaphors, subjects with the CMM input score higher in items of lower frequency.
//Method//Task 2 tests the similarities and differences of Mapping Principles cross-linguistically by analyzing the source domains selected by subjects in English-to-Chinese translation of metaphors.
//Conclusion//This research provides an alternative pedagogy for teaching metaphors in the EFL classroom – a method that underlines the Mapping Principle for target-source pairings in metaphor teaching.

Outlining information in a move. Outlining information occurred when an author used letters or numbers in parentheses to list information elements within a single move (e.g., a b c or 1 2 3). Outlining information occurred more often in the CPA corpus (11 out of 50 abstracts) than in the JPA corpus (7 out of 50 abstracts). In the CPA corpus, the greatest number of cases were found in the Result move (7), followed by Method move (2), Background move (1), and Purpose (1) move. One abstract in the CPA corpus used this strategy twice in the Purpose and Result moves. The sample below illustrates the use of outlining information in a move.

Conference paper abstract 4
//Results//The results show that: (1) Pre-junior high English learning exerted a long-term significantly positive impact on students’ English achievement in junior high. (2) The length of pre-junior English learning significantly influenced students’ English achievement of only the 1st semester of the seventh grade, but not the rest of the semesters investigated. (3) The instructor’s nationality, placement method and in-class activity significantly influenced students’ good English performance in junior high. (4) In-class activity and the instructor’s proportion of encouragement and blame were significantly related to students’ interest and motivation for English learning.
All seven instances of outlining information in the JPA corpus used letters to outline information. However, some authors also followed the more traditional strategy of using words (e.g., *furthermore* and *finally*) to list informational elements. The sample below illustrates this strategy.

Journal paper abstract 3

//Results//Results indicate that a developmental sequence may exist and that this sequence may not be linear, with some longer (i.e., two and three member) codas emerging before some singleton codas. *Furthermore*, it was found that production type (e.g., with epenthesis or feature change) differed significantly by coda type and was indicative of acquisition processes. *Finally*, linguistic environment (preceding and following) and grammatical conditioning were found to significantly affect coda production and development.

In sum, the contrast of the four phenomena listed in Table 3 above between the CPA and JPA corpora suggest some nonnativeness in the CPA corpus. We could infer that the writers of the conference paper abstracts were not as familiar with Western abstract-writing conventions as the writers of the journal article abstracts and did not use effective and sufficiently succinct language in their writing.

**Design Principles of the Online Materials**

The results of the corpus analyses provided strong pedagogical support for the development of the online learning materials to meet specific needs of EFL learners, particularly regarding the insights gained from the comparison of a reference corpus (the JPA corpus) and a corpus of novice, nonnative writers (the CPA corpus). Comparison of the two corpora served to identify acceptable rhetorical norms in the JPA and the possible problems of novice writers in the CPA.

We constructed an online abstract-writing unit for EFL students in our country that focused on the presentation of the overall organization of abstracts and effective language use and highlighted the four phenomena found in the analysis of the CPA corpus. Based on the verb-noun collocations and lexical bundles we identified in the corpus analysis, we grouped these expressions in accordance with the six moves in the abstracts and also presented the learners with more phrasal choices including reporting verbs (e.g., *argue, explain, point out*).

In addition to insights gained from the literature review and text analysis of the two corpora, we also followed some design principles of online materials gleaned from the EAP websites mentioned above. We further referred to the 10 methodological principles by Dougherty and Long (2003) for task-based language teaching and foreign language distance learning. Six of these 10 principles represented guidelines for our development of the abstract-writing unit: (a) the use of tasks, (b) learning by doing, (c) rich input, (d) inductive learning, (e) focus on form, and (f) learner feedback. In the online unit, tutorials on the rhetorical structure of abstracts included input enhancement (e.g., highlighting and font changes), input flooding, inductive learning, and focus on form (e.g., pop-up windows providing language explanations).

In L1 academic writing classrooms, Elbow (1991) contended that “students need metacognition and metadiscourse to help them understand just what these new intellectual practices are that they are being asked to learn” (p. 149). In other words, teachers, in one
way or another, need to help students to “think about” the discourse of their disciplinary community in order to participate in that community. In the long run, metacognitive teaching may also help learners develop confidence and become responsible for their writing. The forms and functions of academic language and the mapping between the two need to be explicitly taught in order for students to succeed (Salager-Meyer, 1992). This can be accomplished by using a concordancer. A concordancer allows students to see the set of contexts in which a word appears so they induce the rule for its common collocates or the sentence constructions that follow it. This practice can heighten learners’ awareness of the linguistic structures in the language they are using.

In the design of our abstract-writing unit, the corpus-based approach was a crucial element in the design of our unit, and we encouraged the students to make substantial use of a concordancer in academic corpora. In the Lee and Swales (2006) study, the authors acknowledged that the outcome of their course would have been completely different had they not worked with a highly motivated group of students who had already been acculturated in their disciplinary community. However, our unit targeted first-year MA students in an EFL context who had just entered the field of applied linguistics. It should be remembered that our students’ familiarity with the field, motivation, computer skills, and ability to work with quantitative data were probably not as high as those of the students in Lee and Swales study.

**Description of the Online Unit**

The online unit (http://formoosa.fl.nthu.edu.tw/moodle2) was developed within the Moodle course management system and included two additional tools: POWER for peer editing and CARE for concordancing (see Figure 1).

**Figure 1**

Overview of the Online Unit in Moodle
A module that used an existing software shell called POWER (Peer Online Writing & Editing Room; Liou & Chien, 2005) was integrated into the design to facilitate learners’ synchronous online peer review. This module combined the advantages of a chat room and a document management program. The interface for writing and editing in POWER resembled a combination of Word and a text-based virtual reality environment (e.g., a MOO) (see Figure 2). By providing a place for students to enter and edit texts, POWER was designed to benefit students both as writers and readers by expanding their sense of audience beyond that of the teacher and thereby create a greater sense of learner engagement.

Figure 2
Peer Editing in POWER

A concordancer designed for EAP, named CARE (Concordancer of Academic wRitten English), was included in the unit because it provided access points for authentic academic written texts for the students (Liou, Kuo, J. Chang, Chen, & C. Chang, 2005). The students were able to use the concordancer to look up academic passages by language functions and keyword(s). They could enter the name of an information structure (e.g., Purpose move) and keyword(s) and retrieve a list of examples. The citations in the output of the concordancer were arranged according to the length of sentences, readability measures, highlights of academic words (based on the list by Coxhead, 2006), and subject area (computer science or applied linguistics) (see Figure 3).
The CARE program was augmented with a record-keeping function to enable the teacher to see the entries and frequency of look-ups by students over time (see Table 5). This learning history was designed to help teachers (and the developers) understand how students use CARE and which entries they have difficulty with.
The online unit consisted of four major components: a prequestioning poll, web resources, lessons, and tasks. The prequestioning poll focused students’ attention on the overall organization of abstracts before they received explanations about the individual structures used in them. The web resources section contained links to other EAP websites for learners to do further research into their individual questions and problems.

The lessons section had a total of six lessons: four for text explanations with authentic examples of move structures and two for tutorials on the computer tools. The first two lessons covered the move structures using examples from the JPA corpus as models. The third lesson highlighted the four nonnative phenomena in both JPA and CPA corpora to raise students’ awareness of them. The fourth lesson introduced frequently used phraseology organized according to the six moves. The fifth and the sixth lessons familiarized the learners with the functions of CARE and POWER.

The tasks section contained two types of tasks and a total of three activities. The activities focused on the structural organization of abstracts and directed learners to re-sequence moves, match the move structures of an abstract, and analyze text samples. In addition to completing these three activities, learners posted their responses to the activities to an online discussion forum and discussed why and how they completed the activities with other learners. In the phraseology tasks, learners did gap-filling exercises for frequently used phraseology, substituted phraseology with other possible alternatives, and did their own concordancing of move structures or keywords with CARE.

The goal of focusing students’ attention on academic language was to raise their awareness of academic language conventions and demonstrate the use of the conventions in their own writing.

**Formative Evaluation of the Online Materials**

After the students completed the online learning unit, we conducted a formative evaluation to investigate their reactions to the unit with a view toward making improvements in the unit.
A questionnaire was used to elicit target learners’ attitudes about the usefulness of the unit. The questionnaire contained four sections: (a) the learners’ background information on academic writing, (b) the usefulness of the online materials, (c) the overall design of the learning unit, and (d) the new tools—POWER and CARE. The questionnaire consisted of 48 items on a 5-point Likert scale (ranging from strongly agree [= 5] to strongly disagree [= 1]) and a few open-ended questions. The means of students’ responses to the Likert-scale items was computed.

In the background section, the students reported different experiences and needs in writing academic articles. With respect to the usefulness of the online materials, the students rated the lessons section highest among the components of the unit, followed by web resources and activities sections. One of the highly rated aspects of the online materials was the comparison of experienced writers’ use of rhetorical features to that of the novice writers. The students attributed greater value to knowing how the published writers tended to write \((M = 4.20)\) than to knowing the common mistakes the novice writers tended to make \((M = 3.97)\). Most students gave positive responses to the open-ended questions at the end of the questionnaire such as “I learn how to prepare myself for a more professional academic writing style” and “I could have more choices of words to use [on the phraseology section].” Compared to the other components of the unit, students’ responses to the new computer tools were only moderately positive. They regarded \textit{POWER} \((M = 3.54)\) and \textit{CARE} \((M = 3.47)\) as useful and expressed willingness to continue using the tools in the future. The students generally acknowledged the usefulness of \textit{CARE} but wished to see more functions added to the tool in the future.

Based on the students’ feedback, three major changes in the unit were made:

1. The text explanations were shortened, and more text-highlighting devices such as different colors and font types were added.
2. Internal links were added to allow learners to choose the examples they wished to see, thereby avoiding overloading the students with all the texts at once.
3. The web resources were expanded and updated, and more specific instructions on how to use them were added.

\textbf{Analysis of Students’ Writing After Use of the Online Unit}

After the revisions to the online learning unit were made, the unit was subsequently incorporated into an academic-writing course designed to train first-year MA-TEFL students to write research articles \((N = 10,\) all of whom were EFL learners). In this course, the students completed the online abstract unit and then wrote an abstract for their own research articles in which they had already written the introduction, method, result, and discussion sections. During the abstract-writing process, students were required to submit a draft of their abstract, commented by the instructor, and then a final version. Thus, 10 drafts and 10 revisions of writing samples were collected. We followed the same method of analysis of the 20 student samples as in the previous analysis of the texts in the JPA and CPA corpora. We found that the move distributions between the drafts and final versions varied only slightly. The number of Gap moves decreased by one from the drafts (6 occurrences, 60%) to the final versions (5 occurrences, 50%). The other five moves occupied the same percentage in the drafts and the final versions: Background 8 (80%), Purpose 10 (100%), Method 9 (90%), Result 10 (100%), and Conclusion 9 (90%). As a result, only the students’ final versions were analyzed further and compared to the JPA and CPA data.
As mentioned above (see Table 1), the obligatory moves in abstracts in the field of Applied Linguistics seem to be the Purpose, Method, Results, and Conclusion moves. After completing the online learning unit, the students in the course seemed to have control over these basic information elements in their abstracts (see Table 6).

<table>
<thead>
<tr>
<th>Number of moves (%)</th>
<th>JPA</th>
<th>CPA</th>
<th>Learner corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background</td>
<td>24</td>
<td>29</td>
<td>8</td>
</tr>
<tr>
<td>Gap</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Purpose</td>
<td>40</td>
<td>48</td>
<td>10</td>
</tr>
<tr>
<td>Method</td>
<td>38</td>
<td>48</td>
<td>1</td>
</tr>
<tr>
<td>Results</td>
<td>46</td>
<td>46</td>
<td>10</td>
</tr>
<tr>
<td>Conclusion</td>
<td>42</td>
<td>36</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 6
Move Distributions in the JPA Corpus, the CPA Corpus, and the Abstracts Written by the Students

One interesting finding was that students in the course tended to use the Gap moves much more often than JPA and CPA writers. It might be that the instruction in the online unit attracted the students’ attention and that they regarded this move as useful to their writing. Another possible explanation is that the students were not so confident about “worthiness” of their research topics and, therefore, felt the need to justify their research by indicating the gap between their study and previous studies.

The four phenomena found in the CPA corpus (inverted move sequences, missing obligatory moves, disproportional abstracts, and outlining information in a move) were used as criteria to examine the drafts and final versions of the students’ writing. As shown in Table 7, analysis of the students’ drafts revealed three occurrences of inverted move sequences and two occurrences of outlining information in a move. In their final versions, the inverted move sequences decreased by one, but outlining information in a move increased by two, resulting in two inverted move sequences and four cases of outlining information in a move.

Table 7
Move Sequences in the Students’ Drafts and Final Versions

<table>
<thead>
<tr>
<th>Student</th>
<th>Draft</th>
<th>Final version</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Background-Purpose-Method-Results-Conclusion</td>
<td>Purpose-Method-Results-Conclusion</td>
</tr>
<tr>
<td>B</td>
<td>Purpose-Method-Results-Conclusion</td>
<td>Background-Gap-Purpose-Method-Results-Conclusion</td>
</tr>
<tr>
<td>C</td>
<td>Background-Gap-Purpose-Results-Conclusion</td>
<td>Background-Gap-Method-Purpose-Results-Conclusion</td>
</tr>
<tr>
<td>D</td>
<td>Purpose-Method-Results-Conclusion</td>
<td>Purpose-Method-Results-Conclusion</td>
</tr>
<tr>
<td>E</td>
<td>Background-Gap-Purpose-Method-Results-Conclusion</td>
<td>Background-Gap-Purpose-Method-Results-Conclusion</td>
</tr>
</tbody>
</table>
The analysis of the students’ writing is both encouraging and discouraging. On the one hand, no instances of missing obligatory moves or disproportional abstracts were found, implying that the online unit worked as intended. On the other hand, the presence of inverted move sequences in the final versions calls the effectiveness of the online unit partially into question because these moves persisted after instruction. For the outlining of information in a move, since there is no clear evidence that using this strategy constitutes an unconventional style, students in this class could have considered outlining as a useful strategy to make their abstracts more complete, especially when they had much to say in a limited space. This might explain the increased use of outlining in the final versions. It should be noted that the drafts contained both intrasentential and intersentential types of outlining but that the final versions contained only the intrasentential type of outlining. It is also possible that the students’ first language (Chinese) may have contributed to both phenomena, but this question requires more in-depth analysis than is possible here.

Taken altogether, the online abstract-writing unit seemed to have a positive effect on the students’ writing. It helped them to become aware of the structure of abstracts. There were no missing obligatory moves or disproportional abstracts in the students’ writing, but inverted move sequences continued, decreasing only by one in the final versions. The persistence of inverted move sequences could mean that the design of the online unit needs to be improved or that more instructional time needs to be devoted to this problem.

CONCLUSION AND IMPLICATIONS
This study has applied the results of a small-scale text analysis to the development of online learning materials designed specifically for EFL graduate students in applied linguistics to help them learn how to write abstracts for research articles. The study has demonstrated the feasibility of linking research results to instructional materials development. The text analysis examined authentic language samples from journal article abstracts and conference paper abstracts. Among other linguistic features, phenomena from novice writers provided examples to contrast with recognized rhetorical conventions in Western academic discourse. The EAP instructional unit included a web-based concordancer used with academic corpora and a peer-editing tool. Formative evaluation by 35 graduate students identified specific areas for improvement. The revised online abstract-writing unit was subsequently integrated into classroom instruction for a group of 10 graduate students who wrote abstracts after having
completed the online instructional unit. Analysis of the students’ abstracts indicated that the unit triggered awareness of the structure of research article abstracts and informational elements to be included in them. The fact that inverted move sequences continued to appear in students’ writing may indicate room for improvement in the online unit or suggest that EAP instructors need to devote more time to this problem.

This study has some limitations. The size of the reference corpora may have been too small to support the collection of representative collocations and lexical bundles. Also, the researcher’s pedagogical instinct in selecting useful phraseology must be viewed as a subjective procedure. Nevertheless, combining computer-generated results with human selection resulted in useful lists of frequently used phraseology and collocations. Second, the evaluation of the effectiveness of the online unit with only 10 graduate students must be considered as a preliminary evaluation. The study should be replicated in a more carefully controlled research design in order to support claims of the effectiveness of the online learning materials.

Taking these limitations into account, the comparison of the corpus of experienced writers and that of novice writers provided useful insights into specific areas requiring more instructional attention for EFL graduate students. The development process of the online abstract-writing unit resulted in a balance of up-to-date electronic references, editing tools for interactive exercises, and principled design features which can be used to enhance traditional classroom teaching of EAP writing. The study suggests that materials development, peer review, and promotion of inductive learning using a concordancer may facilitate learning in other EAP contexts.

This study started from the idea of comparing the corpus of experienced writers with that of novice writers in order to identify potential problem areas that could be transformed into teaching opportunities. It would be interesting if other researchers could collect additional writing samples from students to identify other common patterns of EFL learners’ writing and compare them to conventional writing samples. Presenting learners’ samples alongside conventional samples should raise students’ awareness about the differences between normal writing conventions and their own academic writing.

In the future, researchers should try to expand the size of the reference corpus (the JPA in our study) in order to create more generalizable results and to present greater varieties of lexical choices in the output of the concordancer. Since e-learning has become an established component in foreign language learning, ways in which to effectively combine e-learning with traditional classroom instruction for optimal learning should continue to hold EAP researchers’ attention well into the future.

**REFERENCES**


**ACKNOWLEDGMENTS**

We would like to thank the National Science Council (NSC 93-2524-S007-002) under the National Science and Technology Program for E-Learning, Taipei, Taiwan, for its financial support. Help from programmers, participating students, and involved research assistants is also gratefully acknowledged.

**AUTHORS’ BIODATA**

Wen-Ming Hsieh is a first-year Ph.D. student in the Instructional Systems Program at the Pennsylvania State University. Before coming to Penn State, she earned an MA-TEFL degree from the Department of Foreign Languages and Literature, National Tsing Hua University, Hsinchu, Taiwan. She previously worked as a graduate research assistant to develop instructional materials for two National Science Council research projects in Taiwan: the Web-based Academic English Course Design and Materials Development, and the Corpora and Natural Language Processing for Digital Learning of English. Her area of interest is integrating technology into language learning.

Hsien-Chin Liou is a professor of Department of Foreign Languages and Literature at National Tsing Hua University, Hsinchu, Taiwan. She has been conducting CALL research for the past 20 years and has published in the *CALICO Journal, CALL, System*, and *Asian Journal of English Language Teaching*, among others. Her current interest is the use of Web 2.0 technologies for teaching of writing. She has directed two national projects on using corpora for general English teaching and academic writing in college contexts in Taiwan.

**AUTHORS’ ADDRESSES**

Wen-Ming Hsieh
The Pennsylvania State University
Email: Wenmin.hsieh@gmail.com

Hsien-Chin Liou
Department of Foreign Languages and Literature
National Tsing Hua University,
Taiwan ROC
Email: hcliu@mx.nthu.edu.tw