

## A corpus-based approach to online materials development for writing research articles

Ching-Fen Chang\*, Chih-Hua Kuo

*Institute of TESOL, National Chiao Tung University, Taiwan*

### ARTICLE INFO

#### Keywords:

Corpus-based approach  
EAP  
Genre analysis  
Online materials development  
Research articles

### ABSTRACT

There has been increasing interest in the possible applications of corpora to both linguistic research and pedagogy. This study takes a corpus-based, genre-analytic approach to discipline-specific materials development. Combining corpus analysis with genre analysis makes it possible to develop teaching materials that are not only authentic but also research-supported. An RA corpus consisting of 60 research articles from three major journals of computer science was constructed. A word frequency list derived from the corpus was analyzed to develop a vocabulary profile for the genre. Move analysis was also conducted based on a self-developed coding scheme of rhetorical moves in the target genre. The move codes were tagged in the corpus texts so that individual moves and move patterns could be retrieved for the purpose of developing research-based online teaching materials for graduate students of computer science. Numerous examples of specialized vocabulary, grammatical usage, and move structures that showcase and characterize academic computer science discourse were used in not only lessons but also learning tasks, discussion topics, and online writing models. The paper ends with a discussion of the usefulness and effectiveness of the online RA writing materials, based on student feedback and assessments.

© 2011 Elsevier Ltd. All rights reserved.

### 1. Introduction

With the rapid development and globalization of science and technology, it is important for researchers to participate actively in the international academic discourse community. With English being the *de facto* lingua franca in the research world, however, academic writing poses a great challenge for non-native speakers of English (NNSs) to participate actively in the international academic discourse community. In the ‘publish-or-perish’ academic culture, unless the results and findings from research are properly written up in the form of research articles (RAs) and published in internationally renowned journals, research achievements will not be recognized. Academic competence, therefore, is characterized by not only professional knowledge but also writing proficiency in English, especially for the specific genre of RAs.

The pioneering work on genre analysis and RAs is Swales’ influential book (1990), which develops the key concepts of genre, discourse community, and task. It proposes a genre-based approach to academic discourse, particularly RAs. Its sequel (Swales, 2004) further elaborates on the theoretical and methodological issues of various research genres, including an chapter revisiting RAs. A cornucopia of studies have to date been conducted to explore the information structure and linguistic features of this genre, taking the genre-based approach.

\* Corresponding author. Address: 516, Assembly Building I, 1001 University Road, Hsinchu 300, Taiwan. Tel.: +886 3 5131402 (work), +886 3 5627993 (home); fax: +886 3 5739033 (work).

E-mail addresses: [cfchang311@gmail.com](mailto:cfchang311@gmail.com), [cfchang@mail.nctu.edu.tw](mailto:cfchang@mail.nctu.edu.tw) (C.-F. Chang).

The complexity and sophistication of this genre, however, has caused great difficulty for novice writers and non-native writers who often suffer from unfamiliarity with its writing style, a register represented at various linguistic levels, including lexico-grammatical features, rhetorical functions, writing skills, and generic structures. For instance, Flowerdew (1999) reveals that Hong Kong Chinese scholars often have less facility of expression and less rich vocabulary than their native English-speaking counterparts, and that they find it more difficult to make forceful claims.

In response to the needs and difficulties of writing for publication, English for Academic Purposes (EAP) programs have recently been increasingly prominent in universities worldwide. Key issues, such as theoretical premises, methodological approaches, and pedagogical issues, including curriculum design, materials development, and assessment, have also been highlighted in numerous studies (e.g., Bhatia, 1993; Flowerdew, 2002; Swales & Feak, 2004; Ventola & Mauranen, 1996). EAP research has produced fruitful results with respect to the genre of RAs at two levels: the macrostructure and the linguistic features. The former line of research usually takes a genre-analysis approach to the information structures or rhetorical functions of various RA sections; for example, the Create-A-Research-Space (CARS) model proposed by Swales (1990) is aimed to account for the rhetorical movement in RA introductions; in other words, how information is organized, in terms of moves and steps, to perform the rhetorical functions of this section. A large number of studies have been conducted along this line (e.g., Dudley-Evans, 1986; Hopkins & Dudley-Evans, 1988; Posteguillo, 1999; Swales, 1990). The latter focuses on the analysis of linguistic features in RAs such as tense, voice, modals, hedging, metadiscourse, reporting verbs, and personal pronouns (e.g., Butler, 1990; Hyland, 1998; Kuo, 1999; Tarone, Dwyer, Gillette, & Icke, 1998; Thompson & Ye, 1991) (see Swales (1990) and Swales (2004) for more references). Some other studies have also explored the possible association of lexico-grammatical features with the rhetorical functions of a particular section (e.g., Gledhill, 2000; Lim, 2006), or part genres like acknowledgements (Giannoni, 2002).

While the genre of RAs has a “shared set of communicative purposes” (Swales, 1990), scholars have found that RAs also reflect disciplinary cultures (e.g., Bhatia, 2002; Hyland, 2000). As Hyland (2000) indicates, “the conventions of writing are embedded in deeper epistemological frameworks that are frequently discipline specific” (p. 145). Several genre-based studies have investigated possible variations in RAs across disciplines (e.g., Anthony, 1999; Hyland, 2000; Nwogu, 1997; Posteguillo, 1999; Samraj, 2002). Among them, Anthony (1999) and Posteguillo (1999) are particularly concerned with RAs in computer science. Anthony (1999) reveals that the CARS model does not account for a number of important features in RA introductions in software engineering, such as definitions of terms, examples, and evaluations of the study. Posteguillo (1999) indicates that the central part of computer science RAs does not conform to the IMRD pattern. Bloor (1998), examining various genres in computer science such as manuals and email messages, demonstrates that discourse in this field has a high degree of flexibility.

Corpus-based studies, a burgeoning area, have recently been proposed for descriptive or pedagogical purposes in EAP (Ghadessy, Henry, & Roseberry, 2001; Sampson & McCarthy, 2004; Swales, 2002). They are mainly characterized by (1) the use of a large amount of authentic materials, (2) data-driven, probabilistic computational models, (3) automatic or semi-automatic text analysis, and (4) language use in context. As Flowerdew (2002) notes, findings from the compilation and analysis of specialized genre-based corpora can be used to inform pedagogy in the field of EAP. Hyland (2002) also emphasizes that “materials should be based on analyses of representative samples of the target discourse” (p. 113). Combining corpus analysis with genre analysis makes it possible to develop not only authentic but also research-supported learning materials based on the analysis of target-genre texts.

The development of web technologies and text analysis software offers the potential of utilizing corpus data in designing online EAP learning materials that are based on language learning theories. For example, Chapelle (2003) indicates that CALL offers the possibility of enhanced input which is critical for language learning. The concept of enhanced input is based on a cognitive approach to second language acquisition, holding that the likelihood of learners acquiring linguistic input increases if their attention is drawn to salient linguistic features (Robinson, 1995; Schmidt, 1990). Sharwood Smith (1991) posits two variables – elaboration and explicitness – for input enhancement. Sharwood Smith (1993) further gives sample strategies for input enhancement, including explicit discussion of the form, explicit or implicit error correction, and textual enhancement. Schmidt (1994) uses the term “noticing” to refer to the process of bringing a stimulus into focal attention, whether voluntarily or involuntarily. He argues that “noticing is the necessary and sufficient condition for the conversion of input to intake for learning” (p. 17). We, therefore, conceived that in developing online materials, particularly those about moves, of which most learners may have no prior knowledge, both textual and audio enhancement of moves as well as their rhetorical functions can be used to, hopefully, lead to enhanced input and increase student writers’ consciousness of move structures in the sections. Textual enhancement can be created, for instance, by marking identified moves on the text sample of each section. As Han, Park, and Combs (2008) indicate, the majority of the studies on textual enhancement employ more than one means to augment their effect. Therefore, an audio file explicitly explaining the move structure and linguistic features of the section may become an audio enhancement to draw learners’ attention to the salient features we want to emphasize.

A number of studies have explored related topics and provided insights for the potential incorporation of corpus and computer technologies into EAP research and pedagogy. For instance, Lee and Swales (2006) present a corpus-informed EAP course which attempts to achieve “technology-enhanced rhetorical consciousness-raising” (p. 72). Charles (2007) also reports that combining discourse analysis with corpus investigation “provides enriched input necessary for students to make the connection between general rhetorical purposes and specific lexico-grammatical choices” (p. 289). Other studies include Bianchi and Pazzaglia (2007), taking a metacognitive approach integrated with corpus tools to teach Italian undergraduates to write research articles in English; Bloch (2009), using an online concordancing program for teaching reporting verbs; and

Charles (2006), investigating phraseological patterns in citations. Despite these findings and technological applications, freely accessible academic writing websites, mostly constructed by universities such as Purdue OWL (<http://owl.english.purdue.edu/owl/>) and The Writer's Handbook constructed by the University of Wisconsin, Madison ([http://writing.wisc.edu/Handbook/SciRep\\_Intro.html](http://writing.wisc.edu/Handbook/SciRep_Intro.html)), are at the individual level. There seem few EAP websites providing a genre-based and corpus-informed courseware for writing research articles in a specific discipline.

This paper reports on efforts to investigate the information structure and linguistic features of the genre of RAs in computer science through the compilation and analysis of a target-genre and target-discipline corpus, and describes the resulting online EAP course and preliminary student feedback on its usefulness. Methodologically, the present study integrates corpus analysis with genre analysis. It proceeded through the following stages: corpus compilation, genre analysis, move tagging, text analysis, online materials development, and experimental teaching and assessment. For genre analysis, a coding scheme of rhetorical moves was developed for analyzing the information structure of RAs. Move tagging and concordancing were then conducted for the statistical analysis of moves and retrieval of examples. Linguistic features such as tense, modals and reporting verbs were also analyzed using the concordancing function of the computer software, *AntConc3.0.1* (Anthony, 2005). An EAP course website was subsequently constructed and online materials designed. Learning tools were incorporated into the materials and learning tasks. Finally, the online materials developed were pilot-tested in an academic writing course. Student feedback and learning assessment were also analyzed.

A specialized corpus provides invaluable resources for both research and pedagogy. The corpus-based approach to the development of online EAP materials is characterized by not only real-world professional practice but also enhanced learning input which raises learners' consciousness of distinctive generic features.

## 2. Corpus compilation

The corpus we compiled and analyzed consists of research articles in the field of computer science (CS), since our target learners are graduate students in this field. Three major computer science journals were selected as the sources of the text samples: *IEEE Transactions on Computers*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, and *Computational Linguistics*. The selection of the journals was based on the recommendations of the faculty members at the Department of Computer Science in the two universities where the online learning materials would be used. Twenty RAs were randomly selected from each journal, making 60 texts in total. These text samples were taken from issues ranging from 1996 to 2005, two articles per year from each of the three journals.

## 3. The coding scheme and genre analysis

The next step was a genre analysis of the RAs in the corpus. It was aimed at identifying all the rhetorical moves in computer science RAs. Although there have been many genre analytic studies of RAs, most have focused on a single section. The linking of sections and the schematic structure of the complete RA has often been neglected. Furthermore, little attention has been paid to variation across RAs of different research natures or in different disciplines. In other words, the information structure of RAs, as reflected in moves, may not be as 'neatly' organized as the popular IMRD model prescribes. For example, Posteguillo (1999) found that there is great variation in moves and sections of RAs in computer science and they do not completely conform to the results of existing genre studies of single sections. Our aims in performing genre analyses of moves, therefore, were to identify (1) major and optional moves in each section of the RA, (2) common move sequences in each section, and (3) move structure throughout the whole RA.

A coding scheme was specially developed for genre analysis (Liou, Chang, Kuo, Chen, & Chang, 2005). Although it was originally based on the research findings of important genre analysis of RAs (Bhatia, 1993; Dudley-Evans, 1986; Hopkins & Dudley-Evans, 1988; Peng, 1987; Swales, 1990; Thompson, 1993), modifications were made throughout the process of analysis to reflect the information structure of RAs in computer science. The concept of moves and steps was first introduced in Swales (1981) and later elaborated in Swales (1990, 2004) and other studies (e.g., Nwogu, 1997). A move in genre analysis is "a discursal or rhetorical unit that performs a coherent communicative function in a written or spoken discourse" (Swales, 2004, p. 228), and a step is a sub-unit of move. Both are functional, not formal, units. The identification of moves was determined, following Nwogu (1997), by inferring from context and by reference to linguistic clues in the discourse, such as an explicit expression in the text signalling the rhetorical function that a move performs; for example, *the purpose of the study is to...* can be used to identify a purpose move in a text. In addition, since the move structure was intended for pedagogical purposes, a model with a single level of moves, rather than two levels of moves and steps, was considered easier for students to learn and use. Thus, a move in our pedagogical model of move analysis can be defined as an information unit that performs a specific rhetorical function in research articles. For instance, the background move is a text segment that functions to provide background knowledge for the proposition of concern in the text. The moves do not necessarily occur in a fixed sequential order, and a move may occur in more than one section of a research article. The coding scheme for all the moves (Table 1) is expected to be realistic, and practical and able to reflect the move structure of RAs in computer science.

Four coders, the researchers in this study and two graduate assistants, conducted the move analysis. To ensure better inter-rater reliability of the analyses, a number of conceptual definitions were established, such as the definition of a move and

**Table 1**  
The coding scheme.

Moves/sections	A	I	M	R	D	C
B (background information)	AB	IB	MB			CB
P (purposes or major tasks)	AP	IP	MP	RP	DP	CP
M (methods or theories)	AM	IM	MM			CM
R (results)	AR	IR		RR		CR
D (explanations, implications, comparisons, limitations)			MD		DD	CD
C (partial or complete conclusions, evaluation)	AC	IC	MC	RC	DC	CC
L (literature review or reference to other studies)	AL	IL	ML	RL	DL	CL
O (local or global organization)		IO	MO	RO	DO	
G (gap or missing information)	AG	IG	MG			
F (reference to tables or figures)		IF	MF	RF		
J (justification and reasons)		IJ	MJ			
<i>Moves in individual sections</i>						
IQ (research questions)		IQ				
IV (values)		IV				
MS (populations, samples, or subjects)			MS			
MA (assumptions, conditions, criteria, or hypotheses)			MA			
MPo (procedure)			MPo			
MMI (equipment or materials)			MMI			
ME (definitions, variables, equations, or measurement)			ME			
MT (tests)			MT			
CF (recommendations or further research)						CF

A: Abstract; I: Introduction; M: Materials and Methods; R: Results; D: Discussions; C: Conclusions.

its boundary. In the first week of the analysis, each of the coders analyzed a set of five text samples. We interchanged the set of text samples in the second week. A meeting was then held to check inter-rater reliability. Afterwards, we met once each week to discuss the results and the problems of our analysis. Each rater raised questions about the parts of text which were difficult for move identification or demarcation. Discussion followed until consensus was reached. Moves in all 60 RAs were coded manually (on the hard copies for the purpose of discussion among coders); the move codes were later tagged in the electronic version of the RAs for further analysis.

The results of genre analysis revealed a number of distinctive features of computer science RAs, which influenced the design of our online materials. Firstly, Method, Results, and Discussions are often mixed in a number of sections between Introduction and Conclusions. This result accords with Posteguillo's finding (1999) that the central part of computer science RAs seems to depart from the IMRD pattern. However, his detailed move analysis included only introductions, results, and conclusions. In our analysis, it was found that the section succeeding the Introduction often indicates the research methodology and then in the following sections, related methodological descriptions or procedures are given in a more detailed way, very often mixed with results in the same sections and sometimes even accompanied by interpretations of results and comparisons with other studies. In addition, we found intra-disciplinary variation in the information structure of sections that involve method, procedure, and materials, mainly because research in the sub-areas of computer science can adopt very different research methodologies. For example, the procedure may focus on implementation, configuration, analysis, or experiment, and the materials may be computer models, architectures, or algorithms. Swales (2004) proposes "a cline with heavily clipped texts at one extreme and highly elaborated ones at the other" (p. 220) to explain possible variations in a Method section across disciplines. It is further indicated that when methodological innovativeness is the aspect the writers want to highlight, considerable elaboration can be expected. The method sections in computer science RAs are obviously inclined towards the elaborated end, and the complexity of research methodology in this discipline is reflected in the more-than-one method-related sections in a single RA and the mixing of method, results and discussions in some of them.

A second feature is related to the scope of moves. Some moves may be regarded as macro-moves which occur in more than one sections although they may be differently realized linguistically and have different rhetorical functions in different sections. For example, "referring to other studies" (L) can occur in Introductions (IL), Methods (ML), Results (RL), Discussions (DL), and Conclusions (CL). This move, when occurring in the Introduction, may perform the rhetorical function of reviewing literature and be realized linguistically by using appropriate citation forms; when it occurs in Methods, it may refer to the method adopted in another study. Differently, reference to other studies can serve the purpose of comparing the research results with those of another study in Results or Discussions. The present study revealed a few such cross-sectional macro-moves, such as 'describing purposes' (P), 'indicating theories or methods' (M), or 'reporting results' (R), in addition to 'referring to other studies' (L), indicated above. In the corpus, such a move is tagged differently in different sections; for instance, L was tagged as IL in the Introduction and ML in Methods. This allows learners to access and observe its functions and linguistic realizations in various sections by using the concordancer provided by our courseware.

Also, probably reflecting the organizational feature of computer science RAs, as indicated earlier, headings of the sections between the Introduction and Conclusions are rarely entitled 'Method,' or 'Results.' Instead, topic-related words, such as

'Architecture,' 'Image Processing,' or 'System Model and Framework,' are more commonly used for these sections. However, 'Introduction' and 'Conclusions' are still used by a majority of authors of RAs in the corpus.

Finally, the analysis of move bundles in individual sections, using computer software *AntConc3.0.1*, revealed that both 3-move and 4-move bundles have low frequencies. For example, in Introduction, which has a total of 714 occurrences of moves, only six 3-move bundles (such as IB-IL-IG and IO-IB-IL) have a frequency equal to or higher than ten, and only two 4-move bundles have such a frequency. This suggests that moves in computer science RAs are not often organized in a fixed way.

The results of the genre analysis were incorporated in the design and structure of our online course and the organization of the learning materials. For instance, we decided to combine Results with Discussions to form one section. Also, consciousness-raising tasks were designed to show learners the relationships between sections and moves.

#### 4. Move tagging

After the analysis of rhetorical moves in our RA samples, move tagging was performed. The text samples obtained in the PDF form were converted into pure text files, and unnecessary graphics and non-text words deleted. Move codes were then added as tags in the electronic texts (Fig. 1), using the agreed coding scheme. The tags were positioned at the beginning of each move. Move tagging of all text samples serves the following two functions: statistical analysis of moves, and retrieval of all occurrences of specific moves via concordance for manual analyses. The former can help us identify the frequency and range of moves or move patterns in each section, which are in turn used in the learning materials; the latter can provide authentic move examples, also for pedagogical purposes.

#### 5. Text analysis

In the next stage, text analysis was performed. The analysis targets three fundamental aspects of writing research articles: field-specific vocabulary, lexico-grammatical features, and linguistic features related to the communicative purposes of each sections such as making citations in Introduction. A CS wordlist was first created from the word frequency analysis of the corpus for the purpose of providing learners with high-frequency, field-specific words in writing a research paper. The CS wordlist consists of 335 word families (1230 word forms) in total. The wordlist was constructed through the following steps. First, word forms that had a frequency of 47 or higher were selected, since these word forms constituted 80% of the whole corpus. These words were then compared against the General Service List (GSL). Words in the GSL were deleted from the wordlist, as they are mostly common or function words. However, a few words which are in the GSL but have technical meanings (that is, words which are "recognizably specific to a particular field or discipline" (Nation, 2001, p. 198)) or sub-technical meanings and special usages (that is, words which have the majority of uses with a specific meaning in a particular field or discipline (Nation, 2001, p. 199), Category 3) in the field of computer science, such as *map*, *figure*, *given*, or

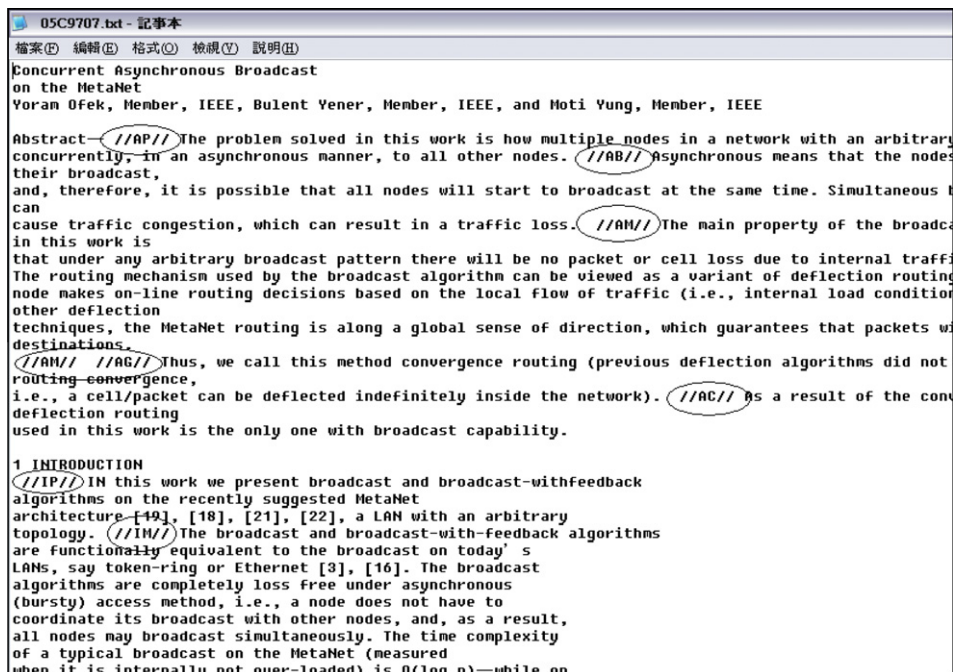


Fig. 1. A text sample with move tags.





Fig. 2. CS wordlist in combination with a web concordancer.

frequency, were retained in our list. The wordlist was provided on the website as a vocabulary resource and linked to the concordancer we developed in an earlier study (Kuo, 2008) so that learners could search example sentences showing the usages and contexts of the CS vocabulary from the corpus (Fig. 2).

In addition to the CS wordlist, specific lexico-grammatical features were also identified based on corpus analysis, such as final participial phrases, modal verbs, and active/passive voice. For example, it was found that final participial phrases perform a wide variety of rhetorical functions in computer science RAs, showing the logical relationship of not only cause and result, but also reason, means, purpose, and condition. Moreover, since EAP research has revealed that phraseology, or formulaic expressions (stretches of conventionalized language use), in academic writing are often problematic for non-native writers (Cowie, 1998; Gledhill, 2000; Howarth, 1996), we identified such lexical phrases that can be linked to the communicative purposes of a section through the analysis of lexical bundles (using Word Bundles in *AntConc*) in each section. For example, *the purpose of the study* indicates the purpose of research in the Introduction, and *as shown in Fig. (x)* refers to data and results presented in graphics, mostly used in Results.

With respect to linguistic features related to the communicative purposes of each section, we focused on not only those that had been indicated in existing research, such as citations in the Introduction and qualifications in presenting results (e.g., Swales, 1990; Swales & Feak, 2004) but those directly derived from the corpus as well, such as compounds in Methods and comparative structures in Discussions. The linguistic realizations of a feature were usually identified through an inductive analysis of their occurrences and sectional contexts in the corpus. For example, as EAP research had indicated that making citations is essential in reviewing literature in the Introduction, we used the subcorpus of Introductions to analyze citations. Citations were identified via concordancing, using either the citation number or parenthesis for retrieval. They were then manually classified into various types and forms. Reporting verbs, which can express RA writers' attitudes towards the reviewed literature, were also examined. Typical examples of each type and form of citations were then retrieved from the subcorpus for the purpose of designing materials.

The results of text analysis not only provide research support for the selection of specific linguistic items in individual sections but also ensure that authentic language use data are included in the online materials.

## 6. Online EAP course design

Based on the results of both the genre analysis and text analysis of the EAP corpus, the online content was developed. The whole procedure consisted of three phases: constructing a website, designing course structure, and creating online materials. The EAP course was constructed on *Moodle* (Modular Object-Oriented Dynamic Learning Environments) (Fig. 3), which is a cost-effective tool for presenting instructional materials on the web. In addition, *Moodle* provides the functions of record-keeping and tracking. Teachers can easily obtain complete reports of the performance of individual students, or all students, for a specific learning activity. In an online writing course, these features help teachers analyze or compare students' online writing during the course and track the development of their writing ability.

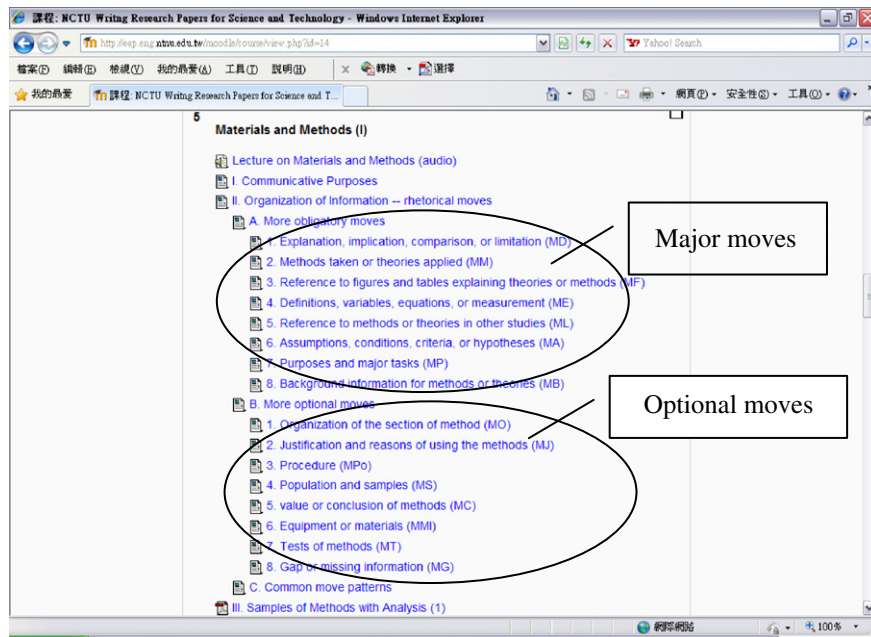


Fig. 3. The online EAP course on Moodle (Major/optional moves in Materials and Methods).

The online EAP platform developed in this study was aimed at graduate students who need to write for scholarly publications. The platform was thus organized around the major sections of the research article, namely, Introduction, Materials and Methods, Results and Discussions, Conclusions, and Abstract. Each section was further divided into two units: information structure and language use. The information structure unit presents the rhetorical moves and common move patterns of each section. It also provides an explanatory audio file in Chinese, lecture notes on the communicative purposes of the section, a sample of the section marked with moves, and a discussion topic. The language use unit focuses on the writing skills as well as the linguistic features of the section, such as how to make citations in the Introduction section. This unit is also characterized by online writing aids and learning resources. Since we wanted to provide a supportive online writing environment, various writing aids, such as an online dictionary, a collocation builder and a concordancer, were incorporated into the website. Learning resources provide students with links to related web resources. In both *information structure* and *language use* units, practical and cognitively demanding homework and learning tasks were included to help learners review related learning points and practise what they have learned.

To help learners who are not familiar with the academic writing style and the genre of research articles, the first two units deal with the fundamentals of academic writing and scientific research articles, serving the functions of orientation and overview. The last part of the online courseware, the appendixes, provides useful references for learners who may need help with the basics, such as the use of articles or compounds.

## 7. Online EAP materials development

The theoretical premise for each section of the RA being addressed through two lesson units, one focusing on information structure and the other on language use, is that the target users, who are non-native EAP graduate students, often have problems with both these aspects. With respect to the former, genre studies have suggested that a genre-based approach to the information structure of RAs is an effective approach to improving genre awareness (e.g., Berkenkotter & Huckin, 1995; Fortanet, Posteguillo, Palmer, & Coll, 1998). For the online materials, we not only extracted examples of individual moves from the corpus but also analyzed the frequency and range of these moves so that we could suggest major and optional moves in each section (Fig. 3). Furthermore, common move patterns were derived from the analysis of move sequences (mostly groups of three or four moves) in the corpus and examples of these patterns were included in the materials to show the possible sequencing of the moves.

To put the idea of input enhancement into practice, in each section, we included a move analysis sample of the section. It is a text sample in its original PDF form marked with appropriate move codes (Fig. 4). Also, an audio instruction was recorded, using the learners' first language, and uploaded. It provides a concise explanation of the information structure and linguistic features in the section. Both serve to show explicitly the focus of a section and to reinforce learners' awareness and understanding of the moves and linguistic features in the section. In addition to the examples provided, learners may also refer to the coding scheme and use the web-based concordancer to search for more move examples from the corpus and, by themselves, induce linguistic realizations of any move by.

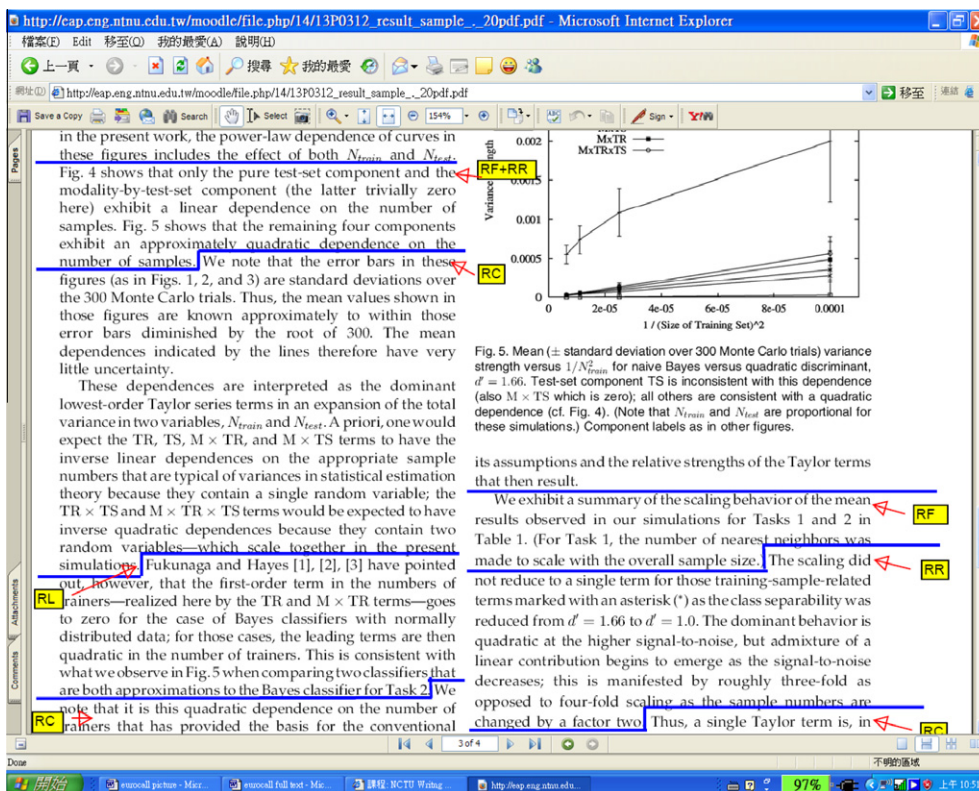


Fig. 4. A move analysis sample of Results.

With respect to language use, writing skills and linguistic features were included in each section. Writing skills focus on specific skills related to the writing of a particular section. For instance, interpreting graphics is essential in the Results section of RAs in computer science, which is characterized by the presence of many graphics presenting research data and re-



Fig. 5. MAKE – the moves and keywords engine.



sults. Linguistic features refer to lexico-grammatical features also characterizing the particular section, such as impersonal structures in the Materials and Methods section. Examples of linguistic realizations of the skills and features were retrieved from the corpus and presented with explanations.

Each section consists of the following modules: learning tasks, discussion, homework, learning resources, and online writing. These allow learners to check their learning, interact with the teacher and peer learners, self-evaluate, access related web resources, and practice writing. Many of the tasks and activities involve the use of the corpus data. In particular, they are based on the concept of enhanced input, as explained earlier, and are informed by the results of our corpus analysis. For example, a learning task for the practice of tenses is specially geared towards the special usages of tenses in the RA in comparison to the usages of tenses in general English. The online writing practice task links up with an online writing system featuring a supportive writing environment that provides peer review, writing tools and resources, and an e-portfolio (Kuo, 2008).

In addition to the course content, the online EAP writing course is supplemented by a number of powerful tools—a web-based concordancer, a collocation tool, and a self-developed moves and keywords engine (*MAKE*), as shown in Figure 5. The purpose of these tools is to facilitate non-native students' writing process and writing development as L2 writing research has revealed that in the process of composing and revising, L2 writers may need to deal with lexico-grammatical problems (Johnson, 1985; Martin-Betancourt, 1986). The tools provided involve learners in the active analysis of words and structure usage in EAP and are based on authentic data from the corpus.

## 8. Experimental teaching and assessment

The online EAP materials were pilot-tested in an EAP writing course. A hybrid model was adopted; in other words, classroom teaching was combined with online learning. We decided to do this because we wanted to closely observe students' online learning behavior and solicit their responses to the various parts of the online courseware. Twenty-three graduate students enrolled in the course.

At the beginning of the semester, orientation to the online course was given. In particular, the design principles of the online materials were explained, including the concepts of moves, move analysis, corpus-based materials, and writing tools. During the semester, online materials were incorporated into classroom teaching, learning activities, and homework. Students were required to complete use the homework and learning tasks on the EAP website and use the discussion and learning resources it also contained. Learning activities, such as searching for move examples or extracting and inducing usages of words were designed.

Since the online materials were based on a genre-analysis approach to the organization of RAs, with which the students might be unfamiliar, the class teacher, one of the researchers, spent half of each class period (50 minutes) explaining the rhetorical functions of moves and move patterns. The writing skills and linguistic features characterizing each major section were also emphasized, as the English proficiency of most of the students was only at the intermediate level. However, the students had an immediate need of writing RAs and theses/dissertations; therefore, they had high motivation and welcomed the cognitively demanding tasks.

A survey was carried out at the end of the semester, using a questionnaire of 16 questions with a Likert scale of five, to collect students' feedback to the online learning materials as well as other course components. The questionnaire consisted of four sections: online materials, online activities, resources and tools, and learning platform. Analysis of the 17 collected questionnaires (Table 2) showed that over 70% of the students were positive about the usefulness of the online materials. With respect to the learning activities, around 80% of the students were satisfied with the online activities. A further examination of the four types of online activities showed that learning tasks and online writing were rated more useful than homework and discussion. The design of the learning platform, however, did not receive positive feedback. It may be due to the rather dull, plain-text format provided by *Moodle*, the platform we used.

Assessment of the effectiveness of learning using the online materials was also performed. Two writing tasks were designed and pre-learning and post-learning assessments were carried out (see Appendix A). The first assessment task focused on the move structure in the Introduction section. It was aimed at evaluating whether the students could organize related information to make a proper link between past research and the writer's own study, a major rhetorical function of the Introduction section. The second writing task was targeted at data commentary, an important skill for reporting research results illustrated in graphics. The purpose of the task was to evaluate whether the students could acquire this writing skill, after learning about it, and apply it to a new task (see Appendix B for a student writing sample).

Since the first task was designed for assessing the effectiveness of the online materials and the extent to which the application of appropriate move structure could improve the overall quality of writing, we did not give numerical scores but qualitatively compared and analysed of each student's pre- and post-learning writing. In total, 20 students' pre- and post-learning writings tasks was analyzed. The results revealed that 13 of the 20 pre-learning writing tasks (65%) had inappropriate organization of information, mostly disorganised move structures and a lack of transitions or links; in contrast, all post-learning writing showed appropriate sequencing of information representing the various moves. In addition, a majority of the students used appropriate expressions to signpost the rhetorical functions involved, such as *however* to indicate IG (gap), *the purposes of this study* to indicate IP (purpose), and *It is hoped that* to indicate IV (value). The improvement in students' post-learning works suggests that the proposed genre-based approach to online materials has led to students' increased consciousness of moves, their rhetorical functions and possible linguistic realizations. The textual and

**Table 2**  
Students' perception of the online EAP materials.

Item	SA/A		Neutral		SD/D	
	n	%	n	%	n	%
<i>I. Online materials</i>						
1. The units on general academic writing and research articles provide useful orientation	15	88	2	12	0	0
2. The audio files in Chinese are helpful	13	76	4	24	0	0
3. The analyses and examples of moves are useful	13	76	4	24	0	0
4. The common move patterns are useful	13	77	3	18	1	6
5. The authentic sample tagged with move structure is useful	12	71	3	18	2	12
6. The explanation and examples of writing skills are useful	17	100	0	0	0	0
7. The explanation and examples of linguistic features are useful	16	85	0	0	1	6
<i>II. Online activities</i>						
8. The learning tasks provide useful practice	14	88	1	6	1	6
9. The homework provides useful practice	13	76	3	18	1	6
10. The discussion topic provides chance to think over important issues in writing	13	67	4	24	0	0
11. Online writing is a good design for developing my writing ability	15	89	0	0	2	12
<i>III. Resources and tools</i>						
12. The learning resources provide additional and useful information	14	83	2	12	1	6
13. The web concordancer is useful to find more examples of words and moves	13	77	3	18	1	6
14. The wordlists is useful for learning scientific and academic vocabulary	14	88	1	6	1	6
<i>IV. Learning platform</i>						
15. Moodle is easy and friendly	12	71	4	24	1	6
16. The web page design on Moodle is appealing to me	8	47	8	47	1	6

Note: SA/A refers to strongly agree/agree; SD/D refers to strongly disagree/disagree.

audio enhancement may contribute to this effect, since both moves and move-signposting expressions were marked on the move samples and the rhetorical functions of moves were emphasized in the audio files.

A trait-based scoring method was adopted for the second assessment task, using 'information structure' and 'language use' in the Results section as two relevant traits. With the scores from these two traits, the essays of each student received a total score ranging from 1 to 6 (see Appendix B). After the scoring of each student's writing, statistical analysis was performed. The mean scores (with standard deviations) of the pre-learning and post-learning tasks of the 20 students were 4.15 (1.11) and 4.50 (0.81), respectively, suggesting some improvement after the lessons, although the result of a paired sample *t*-test shows no statistical difference ( $t(20) = -1.142, p = 0.133$ ). This is probably due to the fact that the pre-learning task was assigned as homework while the students completed the post-learning task in the classroom. The time constraint may had an impact on the quality of the writing tasks.

Both the survey and the assessment, though limited in terms of scope and depth, provide valuable information about student attitudes and their online learning behaviors, particularly about how we can improve the corpus-based online EAP courseware.

## 9. Discussion and conclusions

Advances in computer technology and the Internet have had a marked impact on many aspects of language learning. This EAP study started from a compilation of a corpus of 60 research articles in computer science. It then conducted genre and text analyses of moves and lexico-grammatical features of this genre, and finally developed online materials and carried out experimental teaching and pre- and post-learning assessments.

At the various stages of the study, a number of issues arose and attracted our attention. Firstly, unexpected differences in move structures from the results of previous genre analytic studies of RAs were found; such disciplinary variation, we believe, can be identified only through careful examination of a target corpus. This points to the usefulness of a corpus-based approach to EAP pedagogy. Differences in moves and move patterns can reflect different ways of claiming and constructing knowledge, which can reveal disciplinary research cultures.

In addition, analyses relating lexico-grammatical features of a genre, like RAs, to its information structure, particularly different communicative purposes or rhetorical functions of the sections, are linguistically illuminating and pedagogically useful. Results from such analyses provide data and information for contextualized learning materials. Since RAs are a conventionalized genre, non-native EAP learners need to be explicitly taught these generically distinctive features.

Nevertheless, at the present stage, automatic analysis of text features is still limited. More intelligent and powerful tools can provide much help for researchers, not only saving considerable time but also validating existing hypotheses of the linguistic descriptions of a genre. Specifically, the powerful extension of the computer's role in writing instruction may be achieved through the implementation of a supportive writing environment which provides helpful resources and friendly tools. These can contribute to the accomplishment of the goals of learners as researchers and autonomous learners. Nevertheless, we may still need to guide students towards making the most of these resources and tools.

Finally, corpus-based research also offers teachers a wealth of suggestions on online learning design, such as consciousness-raising tasks highlighting specific information structures and language forms revealed by corpus analysis. The corpus-based approach to EAP enables us to inform EAP pedagogy. As indicated by Lee and Swales (2006), “the closer the participants could come to their discipline-specific written discourses, the more engaged with the texts they became and the more time they were willing to spend on them. . .” (p. 71). A target corpus can provide abundant and invaluable resources for both teachers and students, facilitating the development of learning materials and increasing the motivation of learners to study the target genre. Most importantly, the online EAP courseware promotes active learning with research-supported materials which are based on real-world language use data. The present study reports an attempt to construct a supportive writing environment that incorporated helpful writing aids and tools. For future research, more focus may be placed on the learning behaviors or the writing process of learners using online resources or tools in relation to learning effectiveness.

The study is not without limitations, of course. The corpus on which the learning materials are based is not large enough to promise representativeness. The results of the move analyses thus need to be treated with caution. The size of the participants in our experimental course is too small to ensure a proper assessment of both learner feedback and learning effects. However, for us, the whole process of the study, though time-consuming, was a valuable exploratory experience in the utilization of educational technology in language pedagogy.

## Acknowledgements

We would like to thank the editor and the two anonymous reviewers for their valuable comments on earlier versions of the manuscript. We would also like to acknowledge the financial support by National Science Council, Taiwan, Republic of China (NSC93-2524-S-007-002).

## Appendix A

### A.1. Prompt for pre-learning task 1

Writing task: Complete the part of the introduction section after the literature review, using the information provided. Achievement and Mobility

Movement of families from one area to another is an accepted part of modern life (US Census Population Reports, 1974). The influence of this mobility on school achievement has been the focus of several studies. Bourke and Naylor (1971), in an early review of the literature, found that 11 previous studies reported no effect of mobility on academic achievement, while 12 studies found lower achievement. Goebel (1975) ascertained that the rate of mobility was not a significant factor in determining either short- or long-term academic performance. Benson, Haycraft, and Weigel (1979) studying sixth graders determined mobility to be negatively related to achievement. Likewise, Abramson (1974) and Schaller (1976) both reported that mobile students had lower academic performance.

1. We need to consider the factor of adjustment.
2. The findings in the studies are inconsistent.
3. The significance of this study is as follows: clarify the inconsistency among previous studies; achieve a better understanding of the relationship between academic performance and mobility.
4. This study explores the relationship between mobility and classroom adjustment.
5. This study also investigates how mobility affects school achievement.

### A.2. Prompt for post-learning task 1

Writing task: Complete the part (IG, IP, IV) of the introduction section after the literature review, using the information provided.

(The same passage used in the pre-learning task)

IG: gap

1. Findings inconsistent.
2. The need to consider the factor of adjustment.

IP: purpose of the study

1. First, the relationship between mobility and classroom adjustment.
2. Then, how mobility affects school achievement.

IV: significance of this study

1. To clarify the inconsistency among previous studies.
2. To achieve a better understanding of the relationship between academic performance and mobility.

## Appendix B

### B.1. A sample of essay for post-learning task 2

Table 7.2. Community hearing test data.

Age	Number Tested	Overall Percent	Number Passed	Number Failed	Percent Passed	Percent Failed
0–19	26	6.5%	26	0	100.0%	0.0%
20–29	89	22.1%	89	0	100.0%	0.0%
30–39	85	21.1%	79	6	92.9%	7.1%
40–49	64	15.9%	59	5	92.2%	7.8%
50–59	52	12.9%	33	19	63.5%	36.5%
60–69	62	15.4%	31	31	50.0%	50.0%
70–79	20	5.0%	7	13	35.0%	65.0%
80–89	5	1.2%	0	5	0.0%	100.0%
Total #	403	100%	324	79		

The community hearing test data is presented in Table 7.2. It showed that the ratio of people passed the hearing test was negatively related to the age. People in age group 0–19 and 20–29 all passed the test. In older age group, the passed rate decreased. No one in the age group 80–89 could pass the test. In addition, the passed rate had a dramatically drop between age group 40–49 and age group 50–59 (from 92.2% to 63.5%). This suggested that people after 50 years old should pay more attention to their hearing ability.

## Appendix C

### C.1. Scoring rubric for task 2

Score	Information structure	Language use
3	All three major information elements contained and well-organized	Excellent use of language to interpret data shown in the table
2	Information elements incomplete or inadequate interpretation of data	Inconsistent language use in interpreting data shown in the table
1	Information not presented in a recognizable way or only raw data presented	Low readability and poor vocabulary and grammar

## References

- Anthony, L. (1999). Writing research article introductions in software engineering: How accurate is the standard model? *IEEE Transactions of Professional Communication*, 42, 38–46.
- Anthony, L. (2005). AntConc, Version 3.0.1. Retrieved from <<http://www.antlab.sci.waseda.ac.jp/software.html>> [computer software].
- Berkenkotter, C., & Huckin, T. N. (1995). *Genre knowledge in disciplinary communication: Cognition/culture/power*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bhatia, V. K. (1993). *Analyzing genres: Language use of professional settings*. London: Longman.
- Bhatia, V. K. (2002). Professional discourse: Towards a multi-dimensional approach and shared practice. In C. C. Candlin (Ed.), *Research and practice in professional discourse* (pp. 39–60). Hong Kong: City University of Hong Kong Press.
- Bianchi, F., & Pazzaglia, R. (2007). Student writing of research articles in a foreign language: Metacognition and corpora. In R. Fachinetti (Ed.), *Corpus linguistics 25 years on* (pp. 259–287). Amsterdam: Rodopi.



- Bloch, J. (2009). The design of an online concordancing program for teaching about reporting verbs. *Language Learning and Technology*, 13(1), 59–78.
- Bloor, M. (1998). Lexical and grammatical choices in innovative language use in computer science. In A. Sánchez-Macarro & R. Carter (Eds.), *Linguistic choice across genres: Variation in spoken and written English* (pp. 155–171). Amsterdam: John Benjamins.
- Butler, C. S. (1990). Qualifications in science: Modal meanings in scientific texts. In W. Nash (Ed.), *The writing scholar: Studies in academic discourse* (pp. 137–170). Newbury Park, CA: Sage.
- Chapelle, C. A. (2003). *English language learning and technology*. Amsterdam: John Benjamins.
- Charles, M. (2006). Phraseological patterns in reporting clauses in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes*, 25(4), 310–331.
- Charles, M. (2007). Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions. *Journal of English for Academic Purposes*, 6(4), 289–302.
- Cowie, A. P. (1998). *Phraseology: Theory, analysis, and applications*. Oxford, England: Clarendon Press.
- Dudley-Evans, T. (1986). Genre analysis: An investigation of the introduction and discussion sections of M.Sc. dissertations. In M. Couthard (Ed.), *Talking about text*. Birmingham, England: English Language Research, Birmingham University.
- Fortanet, I., Posteguillo, S., Palmer, J. C., & Coll, J. F. (Eds.). (1998). *Genre studies in English for academic purposes*. Castelló, Spain: Servei de Publicacions-Universitat Jaume I, Castelló, Spain.
- Flowerdew, J. (1999). Problems in writing for scholarly publication in English: The case of Hong Kong. *Journal of Second Language Writing*, 8, 243–264.
- Flowerdew, J. (Ed.). (2002). *Academic discourse*. London: Pearson Education.
- Ghadessy, M., Henry, A., & Roseberry, R. L. (Eds.). (2001). *Small corpus studies and ELT*. Amsterdam: John Benjamins.
- Giannoni, D. S. (2002). Worlds of gratitude: A contrastive study of acknowledgement texts in English and Italian research articles. *Applied Linguistics*, 23(1), 131–156.
- Gledhill, C. (2000). The discourse functions of collocation in research article introductions. *English for Specific Purposes*, 19(2), 115–135.
- Han, Z., Park, E. S., & Combs, C. (2008). Textual enhancement of input: Issues and possibilities. *Applied Linguistics*, 29(4), 597–618.
- Hopkins, A., & Dudley-Evans, T. (1988). A genre-based investigation of the discussion sections in articles and dissertations. *English for Specific Purposes*, 7(2), 113–122.
- Howarth, P. (1996). *Phraseology in English academic writing*. Tübingen, Germany: Max Niemeyer Verlag.
- Hyland, K. (1998). *Hedging in scientific research articles*. Amsterdam: John Benjamins.
- Hyland, K. (2000). *Disciplinary discourse: Social interactions in academic writing*. London: Longman.
- Hyland, K. (2002). *Teaching and researching writing*. New York, NY: Longman.
- Johnson, C. (1985). *The composing process of six ESL students*. (Unpublished doctoral dissertation) Illinois State University, Normal, Illinois.
- Kuo, C.-H. (1999). The use of personal pronouns: Role relationships in scientific journal articles. *English for Specific Purposes*, 18(2), 121–138.
- Kuo, C.-H. (2008). Designing an online writing system: Learning with support. *RELC Journal*, 39(3), 285–299.
- Lee, D., & Swales, J. M. (2006). A corpus-based course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes*, 25(1), 56–75.
- Lim, J. M. H. (2006). Method sections of management research articles: A pedagogically motivated qualitative study. *English for Specific Purposes*, 25(4), 282–309.
- Liou, H. -C., Chang, J. -S., Kuo, C. -H., Chen, H., & Chang, C. -F. (2005). Web-based academic English course design and materials development. In *Proceedings of the fourteenth international symposium on English teaching* (pp. 452–463). Taipei, Taiwan: Crane Publishing.
- Martin-Betancourt, M. (1986). *The composing processes of Puerto Rican college students of English as a second language*. (Unpublished doctoral dissertation), Fordham University, New York, NY.
- Nation, P. (2001). *Learning vocabulary in another language*. Cambridge, England: Cambridge University Press.
- Nwogu, K. N. (1997). The medical research paper: Structure and functions. *English for Specific Purposes*, 16(2), 119–138.
- Peng, J. (1987). Organizational features in chemical engineering research articles. *ELR Journal*, 1, 79–116.
- Posteguillo, S. (1999). The schematic structure of computer science research articles. *English for Specific Purposes*, 18(2), 139–160.
- Robinson, P. (1995). Review article: Attention, memory and the “noticing” hypothesis. *Language Learning*, 45, 285–331.
- Sampson, G., & McCarthy, D. (Eds.). (2004). *Corpus linguistics: Readings in a widening discipline*. London: Continuum.
- Samraj, B. (2002). Introductions in research articles: Variations across disciplines. *English for Specific Purposes*, 21(1), 1–17.
- Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158.
- Schmidt, R. W. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review*, 11, 11–26.
- Sharwood Smith, M. (1991). Speaking to many minds: On the relevance of different types of language information for the L2 learner. *Second Language Research*, 7, 118–132.
- Sharwood Smith, M. (1993). Input enhancement in instructed SLA. *Studies in Second Language Acquisition*, 15, 165–179.
- Swales, J. M. (1981). *Aspects of article introductions*. Birmingham, England: The University of Aston, Language Studies Unit.
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge, England: Cambridge University Press.
- Swales, J. M. (2002). Integrated and fragmented worlds: EAP materials and corpus linguistics. In J. Flowerdew (Ed.), *Academic discourse* (pp. 150–164). Harlow, England: Pearson Education.
- Swales, J. M. (2004). *Research genres: Explorations and applications*. Cambridge, England: Cambridge University Press.
- Swales, J. M., & Feak, C. B. (2004). *Academic writing for graduate students* (2nd ed.). Ann Arbor, MI: The University of Michigan Press.
- Tarone, E., Dwyer, S., Gillette, S., & Icke, V. (1998). On the use of the passive and active voice in astrophysics journal papers: With extensions to other languages and other fields. *English for Specific Purposes*, 17(1), 113–132.
- Thompson, D. K. (1993). Arguing for experimental ‘facts’ in science: A study of research article results sections in biochemistry. *Written Communication*, 10(1), 106–128.
- Thompson, G., & Ye, Y. (1991). Evaluation of the reporting verbs used in academic papers. *Applied Linguistics*, 12, 365–382.
- Ventola, E., & Mauranen, A. (Eds.). (1996). *Academic writing: Intercultural and textual issues*. Amsterdam: John Benjamins.

**Ching-Fen Chang** is an assistant professor in the Institute of TESOL at National Chiao Tung University, Taiwan. Her research interests include computer-assisted language learning, corpus-based studies, second language writing, and sociocultural theories in second language teaching and learning.

**Chih-Hua Kuo** is an emeritus professor from the Institute of TESOL at National Chiao Tung University, Taiwan. She has published research articles in *English for Specific Purposes*, *IEEE Transactions on Professional Communication* and *RELC Journal*, among others. Her primary research interests include EST/EAP, discourse analysis, and corpus-based studies.