

CHAPTER 2

REVIEW OF LITERATURE

This chapter discusses the relevant literature in the field of corpus studies and English for Specific Purposes (ESP) writing research. Four areas are reviewed, including genre analysis in ESP writing, previous studies of the structure of research articles, corpus-based lexical analysis and lexical bundles.

2.1 Genre Analysis in ESP Writing

Swales (1990) provided a comprehensive definition of “genre” as a class of communicative events, the members of which share the same set of communicative purposes. This rationale shapes the schematic structure of the discourse and influences and constrains choice of content and style. Following Swales’ definition, Bhatia (1993) made a further elaboration and stated that each genre is an instance of a successful achievement of a specific communicative purpose using conventionalized knowledge of linguistic and discoursal resources. In other words, genre is a socially recognized, highly structured and communicative discoursal event or activity which aims to fulfill a particular communicative purpose in a certain community.

According to Swales (1990), genre analysis is the study of how language is used within a particular setting. It focuses on issues such as rhetorical styles and

discourse types and relates research in sociolinguistics, text linguistics and discourse analysis to the study of specialist areas of language. Henry and Roseberry (2001) stated that the general aim of genre analysis is to identify the moves and strategies of a genre, the allowable order of the moves, and the key linguistic features. The next step is to explain why these features were chosen by expert users of the genre to achieve their communicative purpose.

A number of applied linguists are interested in ESP genre analysis of written and spoken discourse in academic and professional settings. Swales (1981, 1990) and Bhatia (1993, 1997) represented ESP genre analysis. It began with Swales' pioneering works (Swales, 1981, 1990) on the Introduction section of RAs, and many researchers have used structural move analysis to explore the generic patterns in genres such as academic RAs (Hopkins & Dudley-Evans, 1988) magazines and newspapers (Nwogu, 1997), public reports (Harvey, 1995), letters of application (Henry & Roseberry, 2001), and dissertation acknowledgements (Hyland, 2004).

The organization of texts consists of moves and each move contains one or more steps. "Move analysis, as articulated by Swales, represents academic RAs in terms of hierarchically organized text made up of distinct sections; each section can be subdivided into moves, and each move can be broken into steps" (Kanoksilapham, 2005, p.271). The recognition of "move" and "step" is central to ESP genre analysis. Many researchers gave different meanings to the recognition of "move" and "step", such as, "moves are discriminative elements of generic structure"

(Bhatia, 1993, p.30). Nwogu (1997) further specified the definition of “move” as a text segment made up of a bundle of linguistic features which give the segment a uniform orientation and signal the content of discourse in it. Yang and Allison (2003) stated that a move is a semantic unit of text achieving a unified purpose in ESP genre analysis. And the concept of “move” captures the function and purpose of a segment of text at a more general level; while “step” provides a more detailed rhetorical means of realizing the function of a move. The set of steps for a move is the set of rhetorical choices most commonly available to RA authors to realize a certain purpose. The order of steps presented in each move only shows a preferred sequence for the choices to occur when in combination.

In the ESP tradition, Swales (1981, 1990) advocated a “move-step” model to analyze a specific genre in a functional perspective. Swales provided a four-move model (Establishing field--Summarizing previous research--Preparing for present research--Introducing present research) to analyze RA introduction and revised it into a three-move analysis of CARS (Create A Research Space) model (1981, p.30), as illustrated in the following:

The CARS Model (Swales, 1990, p.141)

Move 1 Establishing territory

- Step 1 Claiming centrality and / or
- Step 2 Making topic generalization(s) and / or
- Step 3 Reviewing items of previous research

Move 2 Establishing a niche

Step 1 A Counter-claiming or

Step 1 B Indicating a gap or

Step 1 C Question-raising or

Step 2 Continuing a tradition

Move 3 Occupying the niche

Step 1 A Outlining purposes or

Step 1 B Announcing present research

Step 2 Announcing principal findings

Step 3 Indicating RA structure

2.2 Previous Studies of the Structure of Research Articles

Recently, a number of studies have been done in the area of writing in academic and research settings for specific purposes. RAs, the central mechanism for the exchange of information, received attention in genre analysis. A lot of researchers made an effort to report the discourse structure of RAs from individual sections to complete IMRD sections. Table 2.1 summarizes previous studies of the structure of individual sections.

Table 2.1 Previous Studies of the Structure of Individual Sections

Author(s)	Year	Section	Field	Findings
Samraj	2002	Introduction	Wildlife Behavior/ Conservation Biology	<i>Step Review of Literature</i> can be identified in 3 moves.
Samraj	2008	Introduction	Biology/Philosophy /Linguistics	The usage of the first person pronoun were found in the third move.
Ozturk	2007	Introduction	SLA/SLW	M1-M2-M3(predominant) M1-M2-M1-M3 (40%) M1-M3 (30%)
Lim	2006	Methods	Business/Management	3 moves and 12 steps
Brett	1994	Results	Sociology	3 major moves
Holmes	1997	Discussion	History/Political /Science/Sociology	8 moves

In the past few years, some research studies examined discourse structures of various sections of RAs, as well as patterns of use of linguistic features. For instance, Samraj (2002) reported an analysis of RA Introductions from two related fields, Wildlife Behavior and Conservation Biology, using Swales' (1990) model. Twelve RAs from these two fields, all published in 1995, were randomly selected from two journals which were central in these fields. Three moves were identified in her work. The results revealed disciplinary variation in the structure of this genre and some similarities in the patterns proposed. The review of literature can be found in all three moves both in Wildlife Behavior and Conservation Biology, and they are not only limited to Move 1. But it served different rhetorical functions in each move: presenting background information in Move 1, Step 2, elaborating on the gap in research in Move 2, Step 1 and specification of the goal of the study in Move 3, Step

1. These results indicated that a deeper exploration of Swales' (1990) model was needed to explain the structures found in the introductions analyzed.

Samraj (2008) investigated the discourse structure of masters' theses across three different fields of biology, philosophy and linguistics with a focus on the Introduction section. She found three moves in the Introduction section of Biology. She also examined the use of the first person pronoun in all three disciplines. Her findings revealed that most usages of the first person pronoun were found in the third move of the Introduction section where the writers occupy the niche they have created.

Ozturk (2007) compared the move structure of research article introductions between two sub-disciplines of applied linguistics, namely second language acquisition and second language writing research. Twenty RAs were analyzed in terms of Swales' (1990) model. The findings suggested that most of the RAs followed the sequence of Move 1 - Move 2 - Move 3 in the field of second language acquisition. Moreover, two different move structures were almost equally predominant in the organization of RA Introductions (M1-M2-M1-M3, 40% in the field of second language acquisition, 40%; M1-M3, 30% in the field of second language writing).

Lim (2006) analyzed the Methods section in 20 articles from two business management journals and identified 3 moves and 12 steps in the Methods section. The three moves include *describing the data collection procedures*, *delineating procedure/s for measuring variables* and *elucidating data analysis procedure/s*. He

took the following references in his study to identify the move boundaries: 1) linguistic features in the text to indicate the internal boundaries, 2) more obvious markers such as typographical features. He described the different characteristics in different moves. For example, Move 1, Step 1 contains noun phrases and verb phrases. According to Bruce (2008), to date, only Lim (2006) provided a very detailed move-and-step analysis linked to linguistic features, following the ESP approach to genre.

Lim (2006) described the close relations between writer's communicative purposes and the linguistics features used and also investigated which results of analysis might be effectively and commonly used in the Methods sections of management RAs. The findings also revealed the pedagogical significance of the relation between linguistics features and language content. It also suggested that writing courses should meet the needs of students who have difficulties in linking linguistic features with communicative functions in their writing.

Brett (1994) examined the Results sections of 20 sociology RAs, using Swales' (1990) model and identified 3 major moves: metatextual, presentation and comment moves. He described each of the three moves in terms of function, lexis, and grammatical form.

Holmes (1997) analyzed the Discussion section of 30 History, Political Science, and Sociology RAs. His findings revealed no completely obligatory move. He modified the framework of Hopkins and Dudley-Evans's (1988) model which consisted of 11 moves.

Their moves are: 1) background information, 2) statement of results, 3) (un)expected outcome, 4) reference to previous research (comparison), 5) explanation of unsatisfactory result, 6) exemplification, 7) deduction, 8) hypothesis, 9) reference to previous research (support), 10) recommendation, and 11) justification. It seemed that Move 2: *statement of results* is the most common. Holmes made a modification of their framework and added one new move, namely, *outlining parallel or subsequent developments*. He deleted Move 6 *exemplification*, Move 7: *deduction*, Move 8: *hypothesis* and Move 11: *justification*, while combining Move 4: *reference to previous research (comparison)* and Move 9: *reference to previous research (support)* into one move, namely, *reference to previous research*. Therefore, his new model contains 8 moves.

Most previous studies followed Swale's (1990) model as a framework. However, very few of them have investigated the complete rhetorical moves of RAs. Table 2.2 summarizes the previous studies on the structure of complete IMRD sections.

Table 2.2 Previous Studies of the Structure of Complete IMRD sections

Author(s)	Year	Number of Move	Field
Nwogu	1997	11 moves	Medical Science
Posteguillo	1999	14	Computer Science
Yang and Allison	2003	20 moves	Applied Linguistics
Kanoksilapatham	2005	15 moves	Biochemistry

Nwogu (1997) studied the organization of information in medical research papers using Swales' (1981, 1990) genre-analysis model. The corpus used in his

study consisted of 30 RAs from 5 journals. After the preliminary analysis carried out on the 30 initial texts, 15 were randomly selected for detailed analysis. An eleven-move schema was identified: 3 moves for the Introduction section, 3 moves for the Methods section, 2 moves for the Results section and 3 moves for the Discussion section. Nine of the moves were found to be “normally required” and two were “optional”. Each schema was found to embody “constituent elements” and to be characterized by distinct linguistic features. The study provides insights into the nature of discourse organization in this genre of written discourse. The structure is shown as follows:

Introduction Section

Move 1: Presenting background information

Move 2: Reviewing related research

Move 3: Presenting new research

Method Section

Move 4: Describing data-collection procedure

Move 5: Describing experimental procedures

Move 6: Describing data-analysis procedures

Results Section

Move 7: Indicating consistent observation

Move 8: Indicating non-consistent observations

Discussion Section

Move 9: Highlighting overall research outcome

Move 10: Explaining specific research outcomes

Move 11: Stating research conclusions

Posteguillo (1999) presented a linguistic description of the schematic organization of RAs in the field of Computer Science. Forty articles from three different academic journals in computing research were analyzed and 14 moves were found. The results indicated that the IMRD pattern could not be applied to RAs in Computer Science systematically and there was no clearly identifiable methods section in computing RAs. The section, following Introduction, was conventionally termed as “Methods” but the computer engineers avoided using this term. This section was divided into sub-sections and termed as Preliminaries, Algorithms, or Analysis of a Problem. Posteguillo (1999) used Swales’ (1990) model as a framework for the Introduction section, took Brett’s (1994) model as reference in the Results section and followed Swales’ (1990) model as a framework for the Discussion section. Swales (1990) identified 3 moves in the Introduction section and he considered that the first two steps *Claiming centrality* and *Making topic generalizations* in Move 1 were optional while Step 3 *Reviewing items of previous research* was obligatory. But in Posteguillo’s (1999) study, Step 3 in Move 1 was optional. Cooper (1985) suggested that the possible reason for this might be the relative newness of this discipline.

Yang and Allison (2003) presented the main lines of a genre analysis of the macrostructures of RAs in Applied Linguistics using Swales' (1990) model. The corpus used in their study consisted of 40 RAs. After reviewing the literature on relevant aspects of RA structure and its functions, 20 RAs were analyzed. This study reported a systematic genre analysis of the sections of Results, Discussion, Conclusion and Pedagogic Implications in RAs and showed that these sections tend to relate to one another. Altogether, 20 moves were identified in their framework: 6 moves for the Results section, 7 moves for the Discussion section, 3 moves for the Conclusion section and 4 moves for the Pedagogic Implications section. The findings showed that the format of RAs in applied linguistics is different from the conventional IMRD framework.

Kanoksilapatham (2005) identified 15 moves of the whole structure of biochemistry research papers using Swales' (1990) model in her study: 3 moves for the Introduction section, 4 for the Method section, 4 for the Results section, and 4 for the Discussion section. To ensure the representativeness of her sample in biochemistry writing, the top five journals in biochemistry were selected. Twelve articles were randomly selected from each journal. On the basis of her move analysis, a two-level rhetorical structure (moves and steps) was proposed. Her framework was applied in the present study.

In conclusion, previous studies have examined the discourse structure of RAs in different disciplines. Most of them used Swales' (1990) model as a framework for

their studies. However, many previous studies focused on move analysis of a single IMRD section. Very few studies have investigated all sections of RAs. Nevertheless, we should be cautious with these findings. First, the discourse structure was not analyzed on the basis of a representative corpus because of limitations of corpus size. For example, Nwogu's (1997) corpus consisted of 30 RAs from 5 journals recommended by medical practitioners; Posteguillo's (1999) corpus consisted of 40 RAs from 3 journals recommended by subject teachers. Therefore, the sampling might reflect the subjectivity of those who recommended them. Furthermore, different genres (such as clinical reports and experimental articles) were included in the same corpus (e.g. Williams, 1996).

2.3 Corpus-based Lexical Analysis

Corpus means a body or collection of texts in any form, but in the eyes of linguists, corpus means a collection of computer-readable texts compiled using clear design criteria. In other words, corpus linguistics means the study of language through corpus-based or corpus-driven research.

Conrad (1999) stated that corpus-based research has three important characteristics. Firstly, a corpus is used in corpus-based studies but the results of an analysis might vary among corpora which use different design criteria. So representative samples might not be included if the size of the corpora is too small. Secondly, computers are needed for analysis in corpus-based studies. Analyzing and counting items can be done automatically with a computer program and interpreted by a

human user. It is important to code items clearly if a feature is ambiguous. Finally, both quantitative analysis and qualitative analysis use are included in corpus-based studies. While determining the major patterns in language use is allowed by quantitative analyses, the functional interpretations describe the communicative functions, which are related to the major patterns that are found by qualitative analyses.

Sinclair (1997) indicated that corpus linguistics is simply the study of language through corpus-based research, but it differs from traditional linguistics in its insistence on the systematic study of authentic examples of language in use. In this case, corpus approach is often referred to as a “lexical approach”. Francis and Sinclair (1994) stated that irregular lexis is often viewed as secondary as opposed to the use of systematic syntactic rules in traditional linguistics.

With the development of computer technology, the way of seeing and studying language has changed. Computer technology has already revealed quite unsuspected patterns of phenomena, through the interpretation of the language use, which traditional descriptive frameworks are normally not able to account for (Sinclair, 1991). A good feature of computers is their huge ability to count and counting allows researchers to obtain the recurring patterns in text. Qualitative evidence about how language works is acquired by using large samples of corpora. Sinclair (1991) argued that we need huge quantities of text in order to collect a large number of instances and arrive at representative patterns by selecting the most typical of these instances. Sinclair believes accurately specifying the established phrase of a language is a part of the long term task in modern lexical research.

The corpus approach to language illustrates the meeting of lexis and grammar. In 1961, Jones and Sinclair used an early corpus of spoken language, which aimed to find evidence of lexical rather than grammatical organization to reveal associations between the words. Halliday is one of the earliest researchers to relate the study of lexis and grammar. He proposed the combination of lexis and grammar, “which is referred to in everyday speech as the ‘wording’ and technically, it is called lexico-grammar” (Halliday, 1992, p.63).

‘Lexis’ and ‘grammar’ are complementary perspectives names, like the synoptic and dynamic perspectives on a semiotic process; or wave and particle as complementary theories of light; each explaining different aspects of a single complex phenomenon. Given this concept of lexico-grammar, it does not make sense to condone relative frequency in lexis but deny its validity in grammar (Halliday, 1991). Lexical study based on corpora is, in essence, empirical, which shares the common features of a corpus-based approach. Biber, Conrad and Reppen (1998) indicate that the essential characteristics of corpus-based analyses include qualitative, functional interpretations of quantitative findings. A corpus-based approach investigates language use rather than studies language structure. Previous studies have been limited by rhetorical moves; and very few studies have emphasized lexico-grammatical linguistic features. Therefore, by combining the strengths of both qualitative and quantitative corpus analysis tools, the present study provides a more comprehensive description of discourse structure linked with specific linguistic features.

2.4 Lexical Bundles

The term “lexical bundles” was first suggested by Biber, Johansson, Leech, Conrad, and Finegan (1999) who gave a definition of lexical bundles from the perspective of corpus-based research. Lexical bundles are combinations of three or more words which are identified empirically in a corpus of natural language and show a statistical tendency to re-occur. They can be regarded as extended collocations, such as, as a result of, on the other hand, in the case of, the context of, and it is likely to.

Many bundles are not structural units, and not expressions which would be recognized as idioms or other fixed phrases. And a difference between idioms and lexical bundles has been proposed. Saeed (2003) defined an idiom as words collocated together which happen to become fossilized, becoming fixed over time. This collocation - words commonly used in a group - changes the definition of each of the words that exist, such as, kick the bucket (die). However, most lexical bundles are not so fixed and always structurally incomplete. They are far from being simple expressions. An important feature of bundles is their variability across different genres. Biber (2006), for instance, discovered that the spoken genre of classroom teaching uses about twice as many different bundles as conversations and about four times as many as textbooks. Similarly, Cortes (2004), Scott and Tribble (2006) and Hyland (2008), also found systematic differences between genres, with bundles typical of published academic prose being far less common in writing by second language students.

Although lexical bundles are neither idiomatic nor structurally complete, they are important as building blocks in discourse. Lexical bundles provide a kind of pragmatic ‘head’ for larger phrases and clauses, where they function as discourse frames for the expression of new information. That means the lexical bundle is used to express stance or textual meanings, while the remainders of the phrase/clauses express new propositional information that has been framed by the lexical bundle. In this way, lexical bundles provide interpretive frames for the developing discourse, as in the following examples: (Biber, Conrad & Cortes 2004, pp. 392-393)

Now, we want to talk about getting out sample mean...

Today, we are going to talk about testing hypotheses.

If you look at the answers that are given, there’s only two answers that have one big M...

Three major functional categories of lexical bundles can be found in Biber et al.’s (2004) work: stance expressions, discourse organizers and referential expressions. Stance bundles express attitudes or assessments of certainty that frame some other proposition. There are five functional sub-categories of stance bundles: epistemic, desire, obligation, intention/prediction, and ability.

Epistemic lexical bundles: (expressing some degree of certainty)

I don’t know what the voltage is here.

There was irony in the fact that the Russian Revolution [...] proclaimed itself to be Marxist...

Desire bundles: (discussing possible future activities, often talking on an indirect directive function or identifying possible actions)

I don't want to deliver bad news to her.

I want you to take out a piece of paper and jot some notes down...

Obligation (directive) bundles: (directions to listener/reader to complete their actions or expressing predictions of future events that do not involve any participation of the speaker.)

All you have to do is work on it.

Intention/prediction bundles: (expressing the speaker's own intention to perform some future action)

Right now what we're going to take a look at are ones that are [...] positive and beneficial.

Ability bundles: (expressing ability)

I want you to be able to name and define those four curriculum category[sic].

Discourse organizing bundles indicate the overall discourse structure and signal the informational status of statements. Discourse organizing bundles serve three major sub-functions: topic introduction, topic elaboration / clarification, and referential identification / focus.

Topic introduction bundles: (providing obvious signals that a new topic is being introduced)

What I want to do is quickly run through the exercise...

Topic elaboration / clarification bundles: (providing additional explanation or clarification)

It has to do with the START talks, with the Russians.

Identification / focus bundles: (focusing on the noun phrase following the bundle as especially important or being often used after a lengthy explanation emphasizing or summarizing the main point.)

For those of you who came late I have the, uh, the quiz.

Finally, **referential bundles** identify an entity or single out some particular attribute of an entity as especially important. Three major sub-categories are distinguished: imprecision indicators, specification of attributes, and time/place/text reference.

Imprecision bundles: (indicating imprecise reference not necessarily exact or additional references of the same type that could be provided)

I think really we now have what about, six weeks left in class or something like that.

Bundles specifying attributes: (identifying the specific attributes of the head noun)

It creates a little bit of wealth.

These figures give an idea of the size of the ethnological community in Russia.

... students must define and constantly refine the nature of the problem ...

Time/place/text-deixis bundles: (referring to a place, time, or text deixis depending on the particular text)

Children in the United States are not formally employed in farm work...

She's in that ... uh ... office down there ... at the end of the hall ...

As shown in Figure 4.4 ...

For each of these primary discourse functions, taxonomies of sub-categories associated with more specific functions and meanings are developed following a primarily inductive approach. That is, concordance listings of individual bundles are generated to examine the use of each bundle in its discourse context. The bundles are then assigned to groups, and discourse functions for the groups are identified by looking at common traits in the functions of different bundles. Previous theoretical studies of the discourse functions of linguistic features help inform this final synthesis process.

Many linguists have studied frequent word combinations. Lewis (2000) presented an edited volume with many innovative ways of teaching collocations. There are very few studies, however, which have focused on the results relevant to the teaching of lexical bundles or formulaic sequences. Jones and Haywood (2004) described an in-depth study among undergraduate students on an EAP course to reveal their progress in the production of formulaic language in academic discourse after a 10-week instruction period. They indicated that in the case of university students' EAP classes, the use of formulaic sequences can help these students to express the complex ideas in an economical way, to mark different stages in their discourse, and to show the necessary level of formality. These authors also emphasized that the absence of formulaic sequence in students' academic writing may result in inadequate writing. Cortes (2004) compared the use of lexical bundles, defined as three or more

word combinations, between published academic papers and students' writing performance in the disciplines of history and biology. The first part of this study focused on the use of lexical bundles in published academic writing. The most frequent 4-word lexical bundles in her study were identified and classified structurally and functionally. The second part concentrated on the use of those target bundles in students' writing. The findings revealed that university students rarely used these target bundles in their writing compared with published authors in these disciplines.

In sum, lexical bundles are high-frequency word combinations that are used to perform certain functions in discourse and they are important as building blocks in discourse. The over 3-word lexical bundles were identified and focused on in this study, which reflect the realization of move boundaries. To my knowledge, there is no study identifying lexical bundles which are the realization of move boundaries in Agricultural Science writing.

2.5 Summary

This chapter offers a review of literature mainly from two aspects: corpus studies and ESP writing research. A definition of genre and genre analysis is provided. Next, previous studies on the structure of RAs are reviewed. In terms of specific linguistic features of discourse structure, corpus-based lexical analysis and studies of lexical bundles are reviewed. The methodology of identifying moves and lexical bundles is presented in the next chapter.