



当代外国语言文学研究文库

# LINGUISTICS

## 学术文本的短语型式与意义

——语料库数据驱动研究

A Corpus-Driven Study of Phraseological Patternings and  
Meanings in Academic Texts

李晶洁 著



上海交通大学出版社  
SHANGHAI JIAO TONG UNIVERSITY PRESS

当代外国语言文学研究文库

# 学术文本的短语型式与意义 ——语料库数据驱动研究

A Corpus-Driven Study of Phraseological  
Patternings and Meanings in Academic Texts

李晶洁 著

上海交通大学出版社

## 内 容 提 要

本书是中央高校基本科研业务费专项资金自由项目(11D11402)和重点追加项目(11D11412)的一部分。

本书基于 NEW-JDEST 语料库证据,系统地描述与概括学术英语文本中的短语序列,及其显著的形式、意义和功能特征。在数据处理上,本文从多词序列的内部粘着力着手建立新的方法实现计算机自动提取连续短语序列,并以 NEW-JDEST 语料库为测量对象,检验该方法的有效性。在语篇功能分析层面上,本书参照 Halliday 的语言元功能理论,提出短语序列语篇功能的三维模型——经验、立场和组织。这三类语篇功能并非互补而是相互独立,从不同的角度描述学术文本的短语学特征。

本书可供短语学、语料库语言学、以及学术语篇等领域的研究人员使用。

### 图书在版编目(CIP)数据

学术文本的短语型式与意义:英文/李晶洁著. —上海:  
上海交通大学出版社,2012

(当代语言学文库)

ISBN 978-7-313-08193-3

I. 学... II. 李... III. 英语—短语—研究  
IV. H314.3

中国版本图书馆 CIP 数据核字(2012)第 068558 号

### 学术文本的短语型式与意义

——语料库数据驱动研究

李晶洁 著

上海交通大学出版社出版发行

(上海市番禺路 951 号 邮政编码 200030)

电话:64071208 出版人:韩建民

常熟市梅李印刷有限公司 印刷 全国新华书店经销  
开本:787mm×960mm 1/16 印张:12 字数:240 千字

2012 年 5 月第 1 版 2012 年 5 月第 1 次印刷

印数:1~2030

ISBN 978-7-313-08193-3/H 定价:32.00 元

---

版权所有 侵权必究

告读者:如发现本书有印装质量问题请与印刷厂质量科联系  
联系电话:0512-52661481

## Acknowledgements

I would like to express my sincere appreciation to all those who have contributed to making this thesis a reality.

First and foremost, I am forever indebted to my supervisor, Professor Wei Naixing, who provided professional guidance, thought — provoking comments, and intellectual stimulation throughout the research process. If not for his supervision and encouragement, this thesis would not have come to fruition. Special thanks also go to Professor Hu Quansheng, Professor Hu Kaibao, Professor Wang Zhenhua, Professor Wang Tongshun, Professor Yang Huizhong... for their valuable suggestions for the research design and constructive comments on an earlier manuscript.

I would like to extend my gratitude to Jennifer Herriman and Michael Barlow, who generously offered me their PDF version of recent publications, and Professor Fan Fengxiang, my mentor during my MA study in Dalian Maritime University, who ushered me into the discipline of Corpus Linguistics.

A deep sense of gratitude goes to my husband, Hu Wenjie for his continuous support and encouragement during the writing of this thesis, especially in times of difficulty and frustration. I am also grateful to my parents, who have sacrificed too much for my education, and my sister for her care and love.

## Abstract

This thesis is a systematic corpus-driven study of phraseological sequences (henceforth PSs) in English academic texts, with a view to characterizing their most prominent formal and functional features. Theoretically set in Firthian linguistics, this study is an elaboration of the Firthian contextual theory of meaning, the Sinclairian extended units of meaning, and the Hallidayan metafunctions of language.

A new computing method is developed for extracting contiguous PSs by measuring their internal associations. NEW-JDEST corpus serves as the testbed. The new method involves three main elements:

(1) Further developing the concept of *pseudo-bigram transformation* (Silva & Lopes 1999) according to the 1st-order Markov model. In this way we may extend the use of current statistics-based measures, which are proposed for extracting bigrams, to the computing of  $n$ -grams, where  $n \geq 2$ .

(2) Constructing a new normalizing algorithm of *probability-weighted average* for refining the current statistics-based measures, enhancing the precision and recall of PSs extracted by these measures.

(3) Identifying the size of each PS by integrating the methods of “frequency threshold” and “local maxima”.

This research evaluates the new computing method by comparing its effectiveness in identifying PSs with that of other known measures and software (e. g. Traditional entropy, Wordsmith Tools 4. 0). Data of the sampling test suggest that the new method enhances the extraction precision to a greater extent (79. 8%). The extracted data thus better reflect the semantic and structural characteristics of PSs.

Within Halliday’s framework of the metafunctions of language, this study generalizes a three-dimensional model of discourse functions for the extracted PSs: experiential, stance, and organizational. The three prime functions are not complementary but are used to describe a PS from three parallel perspectives.

Experiential PSs form the largest category of phraseology in academic texts. PSs of this kind are the main conveyor of information, characterizing a high density of information in academic texts. Structurally, the majority of these PSs do not constitute full clauses but sequences of one or more (single or multiple)

clause constituents. Semantically, the most common experiential PSs are assigned to five main categories: specifying processes and actions; identifying entities, notions and activities; specifying attributes; expressing time and space; and expressing vagueness.

Lexico-grammatical marking of stance is distinguished between phrasal-level and clause-level PSs. The analysis of clause-level PSs is based on the differentiation of three kinds of structure: overt-subject, it-extraposition and null-subject. Within the systemic framework, stance it-extraposition is characterized by expressing four types of modality: epistemic, i. e. the expression of some degree of possibility and certainty; deontic, i. e. the expression of obligation and inclination; dynamic, i. e. the expression of potentiality and will; and evaluative, i. e. the expression of attitude, opinion and evaluation. Frequent occurrences of it-extraposed PSs reflect that although researchers try not to use subjective and affective locutions overtly in academic texts where objectivity and scientificness are essentially required, they still manage to reveal their attitudes and stands through somewhat covert means.

Organizational PSs are described at three levels of discourse functions: discourse acts, meta-discourse, and general logical marking. Drawing on corpus evidence, we investigate six main types of discourse acts: focusing, presenting views or facts, presenting results, reporting, exemplifying, and text deixis. We also discuss recurrent meta-discourse PSs for “outlining purposes” or “announcing present research” (Step 1 of Move 3 in the CARS model), for the purpose of characterizing their most common patterns and functions in Introduction and Abstract of research articles. NEW-JDEST data demonstrate that both discourse acts and meta-discourse have their own phraseological patternings in academic texts. Some PSs are so frequently used to perform concrete discourse acts or meta-discourse that in a sense, the most typical of them become signals of the concrete functions.

All the data and discussion in this study suggest that co-selection is at the very heart of choosing language forms to realize experiential, stance, and organizational meanings in academic text production. Conventionality prevails into every aspect of language use and co-selection is going on at multiple levels. Specifically, NEW-JDEST corpus evidence points to four types of co-selection in academic texts: the co-selection of lexis and lexis, the co-selection of lexis and grammar, the co-selection of PSs and research topics, and the co-selection of PSs and discourse structures. These co-selective relationships have provided strong indications that lexis, grammar and meaning are essentially an integrated whole,

a fact which has not received due attention in current language descriptions, as well as in linguistic theorizing.

The present study may yield unique insights and implications in the following aspects: the construction of the new extraction method, the re-identification of phraseologies, the elaboration of co-selection at multiple levels, the remodeling of traditional linguistic description, the categorization of discourse functions of PSs, and the EAP teaching in China.

**Key Words:** phraseological patternings, new computing method, experiential phraseological sequences, stance phraseological sequences, organizational phraseological sequences



## 摘 要

本文基于 NEW-JDEST 语料库证据,系统地描述与概括学术英语文本中的短语序列,及其显著的形式、意义和功能特征。本研究以弗斯语言学为理论框架,是对 Firth 的语境论、Sinclair 的扩展意义单位,以及 Halliday 的语言元功能的进一步探讨与实证研究。

在数据处理上,本文从多词序列的内部粘着力着手建立新的方法,实现计算机自动提取连续短语序列,并以 NEW-JDEST 语料库为测量对象,检验该方法的有效性。新的提取方法包括三方面内容:

(1) 以一阶马尔可夫模型为基点,发展“假拟二元序列转化”理论,将现有的两词序列统计方法应用于多词序列的计算和提取。

(2) 建立新的标准化算法“概率均值加权法”,对现有的统计手段进行调整,提高短语序列识别的准确率和召回率。

(3) 采用“频数阈值”和“局部最大值”相结合的方法,确定短语序列的长度。

本文将新方法提取的短语序列与现有的统计手段(例如传统熵手段,Wordsmith 4.0 软件)提取的数据进行抽样检验,结果表明,新方法能够更加有效地识别短语序列,准确率达到 79.8%,提取的数据更能够体现序列内部的语义与结构特征。

在语篇功能分析层面上,本研究参照 Halliday 的语言元功能理论,提出短语序列语篇功能的三维模型——经验(experiential)、立场(stance)和组织(organizational)。这三类语篇功能并非互补而是相互独立、从不同的角度描述学术文本的短语学特征。

实现经验功能的短语序列在学术文本中出现频数最多。这些序列是信息传达的主要载体,体现了学术文本高信息密度的特点。它们通常不具备完整的分句结构而是充当一种或多种分句成分。语义上,经验序列主要实现五类命题意义:表达过程和行动;指称实体、概念和活动;表达属性意义;表达时间和空间意义;表达模糊概念。

词汇—语法立场功能主要由短语层级序列和分句层级序列实现,后者是本文研究的重点。结构上,分句层级的序列可分为显性主语型式、it-外置型式和零主语型式。在系统功能理论框架下,it-外置型式立场在学术文本中表现为 4 类情态意义:认知(表示某种程度的可能与确定)、义务(表示义务与趋向)、动力(表示潜力和意愿)和评价(表示态度、观点与评价)。it-外置序列的高频出现说明,虽然研究的



科学性和客观性要求学术文本中避免出现明显的主观情感标识,研究者仍然使用大量的隐性手段来表达其立场和态度,潜移默化地影响读者。

组织序列的语篇功能体现为三个层面:实施语篇行为、组织篇章结构、表达一般逻辑语义关系。基于语料库证据,本文描述了学术文本中的六类突显语篇行为,分别是聚焦、呈述观点或事实、呈述结果、报道、例示、篇章指示。本文还讨论了CARS模型语步3中用于实现“概述目的”和“宣布研究特点”的元语篇短语序列,并分析了它们在引语和摘要中的复现型式与意义。数据显示,学术文本的语篇行为和篇章结构都有其各自的典型词语实现方式,某些序列被频繁地用来实施具体的语篇组织功能。某种意义上,典型的复现序列成为了具体功能的标志。

所有的数据与讨论都表明,共选(co-selection)是学术作者选择语言形式来实现经验、立场和组织意义的核心机制。规约性渗透至语言使用的不同方面,体现为多个层面上的形式共选。NEW-JDEST证据指向学术文本中的四类共选关系:词汇与词汇的共选、词汇与语法的共选、短语序列与话题的共选,以及短语序列与语篇结构的共选。这些共选关系反映出词汇、语法与意义的统一性。目前这一性质在语言描述和语言学理论中尚未得到足够的重视。

另外,本文在短语序列的自动提取方法、意义单位的重新界定、共选关系的细化、传统语言描述理论、短语序列的语篇组织模式,以及中国EAP教学等方面都有一定的价值和启示。

**关键词:** 短语型式、新的短语序列提取方法、经验短语序列、立场短语序列、组织短语序列

## List of Abbreviations and Signals

EAP:	English for Academic Purposes
EH:	Expected entropy value
EH/H:	Entropy ratio
FEI:	Fixed Expressions and Idioms
G:	Glue value or internal association value
H:	Observed entropy value
MM:	Markov Model
JDEST:	Shanghai Jiaotong Daxue English for Science and Technology Corpus
MI:	Mutual Information
New-MI:	Refined MI by using the new computing method
New-(EH/H):	Refined entropy ratio by using the new computing method
NEW-JDEST:	An expanded new version of the first generation of JDEST
PS:	Phraseological Sequence

# CONTENTS

<b>Chapter 1 Introduction</b>	(1)
1.1 Background of this study	(1)
1.2 Objectives of this study	(3)
1.3 Significance of this study	(4)
1.4 Terminological issues	(6)
1.5 Organization of this thesis	(7)
<b>Chapter 2 Theoretical Background</b>	(8)
2.1 Disentangling the notions of phraseology	(8)
2.2 Major theoretical framework	(17)
2.3 Previous studies of phraseology	(21)
2.4 Summary	(35)
<b>Chapter 3 Research Design and Methodology</b>	(36)
3.1 Analytical framework	(36)
3.2 Corpus to be used	(40)
3.3 Instruments, tools, work procedures	(42)
3.4 Summary	(43)
<b>Chapter 4 A New Computing Method for Extracting Contiguous PSs</b>	(44)
4.1 Previous methods for extracting PSs	(44)
4.2 Features of current statistics-based measures	(45)
4.3 Procedures for data extraction	(50)
4.4 Pseudo-bigram transformation	(52)
4.5 A new algorithm for normalization: probability-weighted average	(55)
4.6 Evaluation	(63)
4.7 Summary	(66)
<b>Chapter 5 Overall Frequency Distributions of PSs</b>	(68)
5.1 Distributions of PSs of various lengths	(68)
5.2 Distributions of PSs of different grammatical types	(69)
5.3 Distributions of PSs of primary discourse functions	(70)
5.4 Summary	(74)
<b>Chapter 6 Experiential PSs</b>	(75)
6.1 Notions of experiential PSs	(75)
6.2 Semantic categories of experiential PSs	(75)

6.3	Experiential PSs for specifying processes and actions .....	(77)
6.4	Experiential PSs for identifying entities, notions and activities ...	(80)
6.5	Experiential PSs for specifying attributes .....	(84)
6.6	Experiential PSs for specifying time and space .....	(88)
6.7	Vagueness expressions .....	(89)
6.8	Discussion and summary .....	(92)
<b>Chapter 7</b>	<b>Stance PSs</b> .....	(94)
7.1	Notions of stance PSs .....	(94)
7.2	Lexico-grammatical marking of stance .....	(94)
7.3	Structural categories of stance PSs .....	(95)
7.4	Adverbial stance PSs .....	(97)
7.5	Overt subject stance PSs .....	(99)
7.6	It-extraposed stance PSs .....	(102)
7.7	Null subject stance PSs .....	(112)
7.8	Discussion and summary .....	(115)
<b>Chapter 8</b>	<b>Organizational PSs</b> .....	(118)
8.1	Notions of organizational PSs .....	(118)
8.2	Functional categories of organizational PSs .....	(119)
8.3	Discourse acts PSs .....	(120)
8.4	Meta-discourse PSs .....	(141)
8.5	General logical signals .....	(150)
8.6	Discussion and summary .....	(153)
<b>Chapter 9</b>	<b>Conclusions and Implications</b> .....	(155)
9.1	Summary of major findings .....	(155)
9.2	Implications .....	(160)
9.3	Limitations of this study .....	(161)
References .....		(163)
Appendices .....		(174)

# Chapter 1 Introduction

## 1.1 Background of this study

Long regarded as a peripheral issue, phraseology is now winning the place it deserves in linguistic theory. The unwieldy terminology used to refer to the different types of phraseological sequences (henceforth PSs) is a direct manifestation of the wide range of theoretical frameworks and lines of research in which phraseological studies are conducted. Interest in the analysis of phraseology is accompanied by a growing awareness of the prevalence of PSs in language use and a wider recognition of the crucial role they play in many different fields of linguistics as well as in language acquisition, teaching, learning, natural language processing, etc.

The major and rapid expansion of the field of phraseology has brought about the co-existence of two research paradigms — theory-driven paradigm and corpus-driven paradigm.

The theory-driven research paradigm is at the foundation of a view that restricts the scope of phraseology to a specific subset of linguistically defined PSs. Representatives of this paradigm include, among others, Bolinger (1976), Fillmore (1979), Pawley and Syder (1983), Cowie (1988), Burger (1998), Nattinger and DeCarrico (1992), and Wray (2002). Bolinger (1976) is among the very first to attach importance to the conventional and formulaic nature of linguistic prefabrications. Cowie (1988) and Burger (1998) see phraseology as a continuum along which word combinations are situated, with the most semantically opaque and formally fixed ones at one end and the most transparent and variable ones at the other. Pawley and Syder (1983) and Nattinger and DeCarrico (1992) are two important early pieces of work that set the stage for this paradigm, and Wray's more recent work (2002) has done much to advance the field.

The theory-driven studies have several features in common: their point of departure is a theoretical model; their principal criterion for identifying a sequence as phraseological is the so-called "psychological salience"; and they largely focus on structurally well-formed and semantically idiomatic expressions

(Wei 2009: 272). In this paradigm, the most idiomatic units, whose meanings cannot be derived from the meanings of the constituents, are often presented as the “core”. This can be seen clearly in Glaser (1998: 126) when he makes the statement that “idioms form the majority and may be regarded as the prototype of the phraseological unit”.

The corpus-driven research paradigm, by contrast, has highlighted the role of “preferred ways of saying things” as key register markers and has contributed significantly to widening the scope of phraseology (Granger & Paquot 2008: 35). Representatives of this paradigm include Sinclair (1991, 2004, 2006), Biber (1999, 2004, 2006), Altenberg (1998), Tognini-Bonelli (2001), Moon (1997, 1998), Hunston (2000, 2008), Wei (2004, 2008), Granger and Meunier (2008), to name a few. Sinclair (1991) has established the corpus-driven paradigm, formulating the famous “idiom principle” and the notion of “extended unit of meaning” and describing related theoretical and methodological considerations. Renouf and Sinclair (1991) have shed light on the important roles of collocational frameworks. Hunston (2008) further develops the theory of pattern and meaning, with a notion of “semantic sequence”. Biber (2006) has brought to the fore the important features of lexical bundles across spoken and written registers of university language. Altenberg’s study (1998) is a milestone of the corpus-driven paradigm, which has pointed to the ubiquity of recurrent word combinations, their formal flexibility and pragmatic conventionality.

The corpus-driven research of phraseology has far-reaching significances for exploring linguistic choices and meaning realization in authentic language use. The major criterion for identifying PSs is probabilistic or frequency information rather than subjective psychological salience (Wei 2009: 272). Thus, many PSs that were considered as peripheral or falling outside the limits of phraseology in theory-driven studies have become central in the corpus-driven model as they have revealed themselves to be pervasive in language, while many of the most restricted units (e. g. idioms, proverbs) have proved to be highly infrequent (Moon 1998).

The overwhelming majority of corpus-driven studies (as introduced above) have been designed for describing the phraseology in general English texts. But there seems a lack of systematic studies of PSs in academic texts, whose characteristic meanings and functions are primarily realized by lexico-grammatical sequences with distinct features.

Most research so far on academic phraseology concentrates on the register of academic speech. For example, a growing number of studies on the Michigan

Corpus of Academic Spoken English relate in some way to the topic of phraseology; namely, Mauranen's studies of metalanguage (2001) and formulae (2003), Poos and Simpson's (2002) study of the hedges *kind of* and *sort of*, and Swales' (2001) article on *point* and *thing*. Some researchers have studied the phraseology in academic texts (Cortes et al. 2002; Cortes 2002; Hasselgard 2009), but studies of their kind are restricted to a small number of PSs, drawing on the database to elucidate their meanings and functions. Academic texts, however, are rich in discourse structures, informational and stance expressions, some of which overlap with other spoken and written registers and others of which seem particularly characteristic of academic prose. The description of a few individual PSs may not uncover the overall traits of phraseology. Therefore, this thesis aims to conduct a systematic study of phraseology in English academic texts, driven by fairly large quantities of the NEW-JDEST data.

## 1.2 Objectives of this study

Based on the NEW-JDEST data, this project sets out to conduct a systematic study of phraseology in academic texts, with a view to characterizing their salient patterns, meanings, and functions. Specifically, four research issues are considered in detail.

First, while the notion of phraseology is very widespread, different authors have defined it differently, sometimes not providing a clear-cut definition, or conflating several terms that many scholars prefer to distinguish. For an effective description of phraseology, this study attempts to propose an operational definition of PSs within the corpus-driven paradigm.

Second, current software and statistics-based approaches to the retrieval of PSs are mostly restricted to measuring the associations between two individual words. This research, on the other hand, attempts to construct a new method for computing the associations within multi-word sequences ( $n$ -grams, where  $n \geq 2$ ).

Third, quantitative features of phraseology will be described briefly, such as the frequency of varied types of PSs and the evenness of distribution of particular PSs.

The distributional facts raise the fourth set of research issues relating to the patterns and meanings of PSs; the extent to which internal structures contribute to the overall meaning; the specific functions that recurrent PSs realize in discourse; and the ways in which we can explain the co-selection of salient



patterns, meanings and functions. These issues are correlated and need to be tackled.

### 1.3 Significance of this study

Compared with the related literature, this study is significant in four aspects.

In the first place, this study will introduce a new computing method for extracting contiguous PSs from corpora. NEW-JDEST will serve as the test data.

The validity of a corpus-driven research of phraseology depends crucially upon the precision of extracting PSs. Many parameters may influence the outcome of extraction. They include the corpus size, the length of sequences, the use of filters such as a frequency threshold or a stopword list, and the statistical measure used (e. g. Entropy, MI). With different methods, the profile of extracted PSs may be different. Current corpus concordance software (e. g. WordSmith Tools 4. 0, ParaConc, Powergrab) and statistics-based measures (e. g. Entropy, MI, Log-Likelihood,  $\Phi^2$ , Dice, T test) are mostly restricted to calculating the associations between two individual words or bigrams, using different algorithms. The extraction of multi-word sequences, however, mainly rely either on the identification of physical co-occurrences of word forms, the defining of frequency thresholds, or the use of complex linguistic filters and language specific morpho-syntactic rules (e. g. Dias et al. 2000; Orliac 2008; Bretana et al. 2008), while neglecting the internal associations within sequences (e. g. Cluster function of Wordsmith, *N*-gram of Powergrab). As a consequence, the extracted data include a considerable amount of disturbing segments with incomplete structure and meaning (e. g. *in the*, *of a*, *to a*), and the extraction precision of PSs is thus very low. Sag et al (2002: 1) even regards the technique of extracting PSs as “a pain in the neck of NLP”. For this reason, the present study will construct a new normalizing algorithm of *probability-weighted average* for refining the current statistics-based measures, enhancing the precision and recall of PSs extracted by these measures.

Second, at the level of linguistic description, Firth’s contextual theory of meaning and Sinclair’s extended units of meaning constitute the main theoretical framework and research stand of this study. In the traditional linguistic theory, a word is described as the basic unit of meaning. However, recent corpus-driven studies have shown that language users employ more collocations, free word

combinations, and different kinds of lexico-grammatical sequences to express meaning under normal circumstances. Sequences of this kind are by no means isolated units, but stretch out to the surrounding context through various types of co-selection. Meaning is realized in context, and concept is defined through the communication within a discourse community. A good expression will be recognized, accepted and repeated by the community and hence occur with a higher probability. According to such information as contextual meaning and probability of occurrence, we are able to study and identify a large number of PSs that fall outside the limits of traditional description of units of meaning. By doing so, we hope to provide more insights into the mechanism of meaning realization in language use and enrich or even reconstruct the descriptive system of units of meaning.

Third, at the level of discourse functions, this research permits the prevalence of multiple-functioning of PSs in academic texts. A variety of schemes of classification have been proposed with varying focuses (Aijmer 1996; Biber et al. 1999; Cowie 1988; Erman & Warren 2000; Howarth 1998; Krashen & Scarcella 1978; Moon 1998; Nattinger & DeCarrico 1992; Yorio 1980; Glaser 1998; Lewis 1993; Granger & Paquot 2008), and a number of distinct categories of PSs have been identified. In most models of their kind, classifications are complementary, i. e. each PS is assigned to one, and only one function. However, the corpus analysis suggests that a number of PSs serve two or more discourse functions in the database, that is, they contribute to their texts in two or more ways. Moreover, cross-functioning operates instantially and relates to the behavior of individual PSs in particular contexts, foregrounding or thematizing the selection. Therefore, this research follows Biber's practice (2006) to list each PS under its primary functional category. Potentially multi-functional PSs are examined in concordance listings and classified in terms of their most common usage.

Fourth, at the pedagogical level, EAP (English for Academic Purposes) teaching practice in China is more or less word-based. The subsequent problem is that although some advanced learners of EAP may acquire a vocabulary as large as, or even larger than that of native speakers, they still have difficulty with nativelike collocation and idiomaticity. Many grammatical sentences generated by language learners sound unnatural and foreign to a discourse community (Pawley & Syder 1983; Granger 1998; Howarth 1998). Swales (1990) points out that every genre of EAP and ESP (English for Special Purposes) has its own phraseology, and learning to be effective in the genre involves learning this.