

EVALUATING THE IMPACT OF A PRESESSIONAL ENGLISH FOR ACADEMIC
PURPOSES PROGRAMME: A CORPUS BASED STUDY

by

STEPHEN ISSITT

A thesis submitted to the University of Birmingham for the degree of DOCTOR OF
PHILOSOPHY

Department of English Language and Applied Linguistics
School of English, Drama and American and Canadian Studies
College of Arts and Law
University of Birmingham
August 2016

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

ABSTRACT

This thesis investigates the impact of an intensive programme of English for academic purposes upon the second language writing development of postgraduate students at the University of Birmingham. The study uses a 300,000 word corpus (EAPCORP) of essays from the beginning and end of the programme covering two separate years, in order to identify and measure written linguistic feature development. A multidimensional investigative approach underpins both of the two main analytical tools applied to the EAPCORP, with the basic premise that it is possible to identify register differences between different types of language by the assemblage and analysis of a large number of textual features. Firstly, Coh-Metrix is a programme employing a range of algorithms applied to a series of data bases to analyse the linguistic structure of texts. Secondly, MAT (Multidimensional Analysis Tagger) employs algorithms developed by Douglas Biber and uses an automated text tagger. The analyses suggest strongly that there has been progression from the initial production of a high frequency of features characteristic of speech to that more typical of academic writing. The results emphasise the importance of well-designed EAP programmes especially in uncertain economic contexts.

CONTENTS LISTING

TABLE OF CONTENTS

Chapter One.

Introduction	1
1.1 Aims of the thesis	1
1.2 Background of the thesis	1
1.3 Outline of the thesis	6

Chapter Two.

Issues in Second Language Writing Development	10
2.1 Introduction	10
2.2 Conceptualising writing development	10
2.3 A conceptual paradigm	13
2.4 Research on writing development	17
2.4.1 Studies with a lexico-grammatical emphasis	18
2.4.2 Studies with a cognitive/developmental emphasis	28
2.4.3 Corpus-based studies	36
2.4.4 COH-Matrix analyses of L2 writing	42
2.4.5 Multidimensional analysis of L2 writing	45
2.4.6 The need for future research	52
2.5 Conclusion	54

Chapter Three.

The empirical contents of the research	56
3.1 Introduction	56
3.2 The nature of the programmes	57
3.3 Creating the pilot corpus	60
3.4 Identifying linguistic features for pilot analysis	63
3.5 Results of pilot analysis	66
3.6 Revisiting the initial conceptual framework	71
3.7 EAPCORP	73
3.8 Conclusion	74

Chapter Four.

The COH-Metrix analysis of EAPCORP	75
4.1 Introduction	75
4.2 Coh-Metrix: components and indices	75
4.2.1 The principal components	76
4.2.2 The indices	78
4.3 The procedure	84
4.4 The results of the Coh-Metrix analysis	86
4.4.1 Discussion of increases in individual indices	93
4.4.2 Discussion of decreases in individual indices	98
4.4.3 Feature movements across programmes	105
4.4.4 Feature movement for all courses	106
4.4.5 Feature movement for 7 out of 8 courses	106
4.5 Summary and conclusion	108

Chapter Five.

MAT analysis of EAPCORP	110
5.1 Introduction	110
5.2 A multidimensional approach	110
5.3 The Multidimensional analysis tagger: MAT	115
5.3.1 The individual variables	115
5.4 The procedure	119
5.5 Results of MAT analysis	122
5.5.1 Results of quantitative analysis	122
5.5.2 Qualitative analysis of the MAT findings	131
5.5.2.1 MAT variables with significant increase (60% or above) for all programmes	132
5.5.2.2 MAT variables with significant increase for 7 out of 8 programmes	135
5.5.2.3 MAT variable with less clearly identifiable upward movement	136
5.5.2.4 MAT variables with significant decrease (60% or above) for all programmes	137
5.5.2.5 MAT variables with significant decrease (60% or above) for 7 out of 8 programmes	140
5.5.2.6 MAT variables with less identifiable downward movement	142
5.5.2.7 MAT variables with contradictory findings	143
5.5.2.8 MAT variables which are infrequently recorded	144
5.6 Factors and dimensions	146
5.6.1 EAPCORP MAT dimensions	146
5.6.2 Movement along the dimensions	152

5.6.3 Dimensions which have indicated movement for all (8/8) courses	154
5.6.4 Dimensions which have shown movement for 7 out of 8 cohorts	155
5.6.5 Dimensions which have indicated movement for 6 out of 8 cohorts	158
5.6.6 Dimensions which have indicated movement for 4 out of 8 cohorts	160
5.7 Summary and conclusions	160

Chapter Six.

Summary and synthesis of findings 162

6.1 Introduction	162
6.2 Considering and synchronising the evidence	162
6.3 A spoken to written perspective	166
6.3.1 At word and phrase level	167
6.3.2 At sentence level	175
6.3.3 Beyond sentence level	177
6.4 A simple to complex perspective: beyond sentence-level feature movement	179
6.5 An inaccurate to accurate perspective	180
6.6 Inconclusive and contrary findings	182
6.7 Summary	183

Chapter Seven.

Implications of the study 184

7.1 Introduction	184
7.2 Implications for syllabus and materials design	184
7.3 Implications for assessment	191

7.4 Implications for the EAP profession	195
7.5 Further issues	199
7.6 Suggestions for further research	201

LIST OF FIGURES AND TABLES

Chapter Two.

Issues in Second Language Writing Development

Figure 2.1 Marking criteria for University of Birmingham Preessional Academic Research Paper (ARP)	14
Figure 2.2 Marking criteria for ARP (lowest end of scale)	16
Figure 2.3 Measures of L2 development (based on Wolfe-Quintero, Inagaki and Kim, 1998)	27
Figure 2.4 Second language writing proficiency levels (Verspoor, Schmid and Xu, 2012, p.258)	31
Figure 2.5 Indicators of language levels (Spinner, 2011)	32
Figure 2.6 Weighted clause values (Wigglesworth and Foster, 2008)	34
Figure 2.7 Noun phrase modification features (from Parkinson and Musgrave, 2014)	40-41
Figure 2.8 Tagged linguistic features of TWE essays (Grant and Ginther, 2000, p.130)	48-49

Chapter Three.

The empirical contents of the research

Table 3.1 Top 20 most frequent words for 15 week 2009 (22 matched scripts)	66-67
Table 3.2 Top 20 most frequent words for 15 week 2010 (35 matched scripts)	67
Table 3.3 Top 20 most frequent words for 10 week 2010 (86 matched scripts)	68
Table 3.4 Top 20 most frequent words for 6 week 2010 (40 matched scripts)	68
Table 3.5 Mean sentence length, feature movement <i>(increase or decrease of mean sentence length)</i>	70
Table 3.6 Lexical densities (percentage of lexical words nouns, verbs adjectives, adverbs) feature movement	70

Table 3.7 Error-free units, feature movement (increase or decrease of error free units)	70
Table 3.8 Additional feature movement for 2009	72
Table 3.9 The EAPCORP. Number of matching pairs	74

Chapter Four.

The Coh-Metrix analysis of EAPCORP

Figure 4.1 Coh Metrix indeces	79-84
Table 4.1 Score range for programmes	85
Table 4.2 Full results of the Coh-Metrix analysis	86-90
Table 4.3 (a) Coh-Metrix feature analysis-twenty features which have shown the most increase.	91
Table 4.3 (b) Coh-Metrix feature analysis-twenty features which have shown the most decrease.	92
Table 4.4 Mean sentence length for all courses	93
Table 4.5 Average word length all courses	95
Table 4.6 Average number of modifiers per noun phrase	96
Table 4.7 Incidence of first person singular pronouns	101
Table 4.8 L2 readability scores	105
Table 4.9 Indeces showing movement (+) or (-) for all 8 cohorts. 263 pairs	106
Table 4.10 Indeces showing movement for 7 cohorts (15, 10, 6 week programmes plus 20 week either 2010 or 2012 and excluding indeces 90 and 105) 254/5 pairs	107-8

Chapter Five.

MAT analysis of EAPCORP

Figure 5.1. Multidimensional features (Biber, 1988, pp.77-78)	111-113
Figure 5.2. Multidimensional analysis variables	116-119
Table 5.1 Score range for the programmes	120
Table 5.2 MAT feature analysis	123-129
Table 5.3 (a) MAT feature analysis.	130
Movements from the beginnings to the ends of the programmes, twenty features which have shown the most increase.	
Table 5.3 (b) MAT feature analysis.	131
Movements from the beginning to the end of the programmes, twenty features showing the most decrease. p.126	
Figure 5.3 (a) Dimension one.	147
<i>Involved production versus informational production</i> (Biber, 1988, p.128)	
Figure 5.3 (b) Dimension two.	
<i>Narrative versus non- narrative discourse</i> (Biber, 1988, p.136)	148
Figure 5.3 (c) Dimension three.	
<i>Explicit versus situation-dependent reference</i> (Biber, 1988, p.143)	149
Figure 5.3 (d) Dimension four.	150
<i>Overt expression of persuasion</i> (Biber, 1988, p.149)	
Figure 5.3 (e) Dimension five.	151
<i>Abstract versus non- abstract information</i> (Biber, 1988, p.152)	

Figure 5.3 (f) Dimension 6 (on-line)	152
<i>Informational elaboration</i> (Biber, 1988, p.155)	
Table 5.4 (a) scores and movements for 6 dimensions 2010	153
Table 5.4 (b) scores and movements for 6 dimensions 2012	153
Table 5.4 (c) movements for 6 dimensions 2010 and 2012	154

Chapter Six.

Summary and synthesis of findings

Table 6.1 EAPCORP Matching pairs per course	165-166
Table 6.2 Talk to writing feature movement at word and phrase level (1)	169-170
Table 6.3 Talk to writing feature movement at word and phrase level (2)	170
<i>personal pronouns</i>	
Table 6.4 Talk to writing feature movement at word and phrase level (3)	172
<i>noun orientation</i>	
Table 6.5 Talk to writing feature movement at word and phrase level (5)	174
<i>verbs and tenses</i>	
Table 6.6 Talk to writing feature movement (6)	176
<i>sentence length and sentence complexity</i>	
Table 6.7 Talk to writing feature movement (7) narrativity Coh-Metrix	177
Table 6.8 Talk to writing dimension movement for 2010 and 2012	178
Table 6.9 Simple to complex feature movement, (1) <i>readability</i>	179
Table 6.10 All course feature movement full stops	181
Table 6.11 Lexical density indicators	182

CHAPTER ONE

INTRODUCTION

1.1 Aims of the thesis

This thesis has two broad aims. The first is to examine a range of measurable improvements detectable in the written academic English of students on short, intensive English for Academic Purposes programmes at a British university and the second is to consider the implications of these linguistic feature developments for course design and teaching.

1.2 Background of the thesis

The initial motivation for this thesis is practical and stems from concern at the current situation regarding EAP provision at HE institutions in Anglophone countries. A growing number of universities, both pre and post 1992 (the date when former polytechnics were given university charters), have effectively outsourced their EFL and EAP provision to commercial organisations and English language service providers. For example, the company INTO publicises partnerships with several UK HE institutions including the universities of Exeter, Newcastle, Manchester, Queen's Belfast, Glasgow Caledonian, City of London and UEA (INTO 2015). Another private sector provider, Kaplan, advertises "*guaranteed progression* (my italics) onto approved degree courses" (Kaplan International 2016) at the universities of York, Aberdeen, Nottingham Trent, Glasgow and Liverpool. As a consequence, some academic EAP units are being shut down or downgraded as service

facilities. This has in some cases resulted in lower professional status, less favourable contracts and the imposition of non-academic line management for professional EAP teachers. Another outcome has been that teaching is increasingly seen as an activity not necessarily informed by, or related to, research. There is also a concern that the quality of teaching may be bypassed as an issue with private companies seeking access to public funding, unencumbered by the employment of relatively expensive professional staff. The status of EAP teachers and the related issues of professionalism in a changing HE environment is discussed by Riley-Jones (2012), who, drawing on the work of Benesch (2001), offers the view that, “the relationship between academe and professionalism is an ambiguous one ... between EAP, academe and professionalism may be considered even more so” (Riley-Jones 2012, p1). Riley Jones, in contributing to a BALEAP discussion thread, draws attention to the withdrawal of government funding in certain aspects of higher education and the requirement of institutions to stand on their own feet (and by association support themselves financially). He also highlights the newly instituted and perceived support function attributed to EAP and language units and the associated reduction in professional status within their respective institutions.

The University and College Union (UCU) has also drawn attention to the privatisation of UK Higher Education in the wider context, and in a detailed document (UCU 2010a, p 3) points to the “devaluation of UK Higher Education as universities chase a reputation governed by student preferences rather than academic standards” and “a growing private sector, aggressively marketing a poorer product at higher prices to vulnerable people to satisfy shareholders” (UCU 2010a, p iii). As the UCU is a trade union and likely to use polemical language in this context, there may well be a tendency on their part to overstate the danger to educational standards. It is nevertheless fair to say that there is a real concern about the maintenance of educational levels within the higher education sector. While not

wishing to state that privatisation is necessarily a bad thing or to denigrate privatised EAP or EFL operations, I would hope to place the issue of maintenance of educational standards at the centre of debate regarding this important and prescient topic.

From the perspective of EAP professionals, in addition to awareness of what could be considered as an attack on their livelihoods, there is a concern that less academically informed programmes, implemented by teachers with lower qualifications and levels of skill, are in many ways less capable of delivering effective EAP teaching. For example, another UCU document draws attention to the lower levels of qualification and experience required of teachers at certain Kaplan controlled language units (UCU 2010 b). The report refers to concerns by staff that, among others, there was insufficient rigour in checking entry qualifications and that unsuitable and unfocused teaching materials were being used. This position, which to a certain extent could be characterised as enlightened self-interest, bears a wider economic significance given that the English language industry is of crucial importance to the UK's national economy, both in financial and cultural terms. English language courses are a substantial and growing component of the UK's GDP, with £10.7 billion of export earnings in 2011/12 including £4.9 billion off campus expenditure (Universities UK 2014). From this perspective, it is important to have as much information as possible about a significant export earner and to make explicit the importance of maintaining a high quality educational service.

This simple idea can be augmented after consideration of the nature of the English language teaching profession. It is possible to be an English teacher in some contexts, such as certain language schools in the UK or abroad, with minimal or no qualifications at all, whereas in the British university setting, it is expected that a teacher of English should possess a degree, higher degree and specific teaching qualifications of a high level (British Council Accreditation 2015) and the maintenance of these professional qualifications may

well be of benefit to the Higher Educational sector as a whole. It is a competitive world, not least in education, and if international students might have in the past accepted whatever they were taught in the UK with little criticism, the situation now is very different. Students have a choice of destinations for their English programmes, not only in English speaking countries such as the UK, USA and Australia, but also in countries such as Holland, Sweden, Denmark and Germany where English medium study is not uncommon and often cheaper, with the result that all those connected with the English language profession need to be aware of possible threats to their integrity, viability and existence. It is therefore possible to say that the motivation for the construction of this thesis takes into account institutional, national and international perspectives. It does not adopt a position that privatisation is wrong and that private companies have no contribution to make to the educational arena in general and to language learning in specific contexts and as such it is apolitical, but it does seek to problematise an *unexamined* incursion into a public, educational realm by a series of private sector organisations.

The contention that highly trained and academically oriented EAP professionals do and can make a measurable difference in terms of student performance, however, is a subject of controversy. At the time of the initiation of this thesis, much academic research tended to support a position that the likelihood of real improvement in written competence over the short time frames characteristic of university preessional programmes was not possible in any significant sense. The research in this area was quite extensive but could be described as somewhat piecemeal and often did not focus specifically on EAP, with the result that there was a need for large scale empirical research. It was therefore the initial aim of this thesis to investigate whether it was in fact possible to identify any measurable improvements in student writing over the short timeframe of a summer preessional EAP programme. However, while my own research has been in progress there has been an increasing focus

on this very subject of measurable improvements in students' written English over the period of short, intensive language programmes, and several studies have suggested that it is, in fact possible to discern differences in written output over this time span.

Nevertheless, and as I will argue in more detail later in this thesis, the evidence provided by these studies is far from conclusive and often limited in range, and thus there remains a clear need for further research in this area. The data for the research reported in this thesis are drawn from students taking preessional programmes at the University of Birmingham, UK. These courses, which take place every year, are of 20, 15, 10 and 6 weeks duration and serve as entry pathways to a variety of postgraduate degrees across the entire university. With approximately 500 students embarking on these courses every academic year, a large amount of data concerning student progress is available for examination. The analysis that forms the investigative core of this thesis will be conducted by the application of computational tools and methods to a corpus of student essays. This corpus is derived from an extensive series of pre and post course writing samples, produced by the students with the overall aim of assessing and measuring increases in productive linguistic written capacity. I will argue that the results of this analysis do provide clear empirical evidence in support of the claim that professionally run and academically informed EAP programmes make a measurable positive difference to students' academic writing abilities. Of course, this is not to say that academically oriented EAP provision is thus superior to that provided by outsourced commercial organizations such as INTO or Kaplan. In order to make such a claim, it would be necessary to carry out an identical corpus-based study of work written by students on an outsourced, commercially run preessional EAP programme, and to compare the results of this study with those obtained for the academically oriented programme. It would be well beyond the scope of this thesis to carry out such a comparison. What the thesis does do, however, is provide a methodology and a set of benchmark results for an

academically oriented programme that allows such a comparison to be made. In this respect, the thesis may be seen as laying down a challenge to private sector EAP providers: can they match or even exceed the gains in student writing proficiency identified by the research reported here?

To summarise, the main aim of this thesis is to ask: do professional EAP teachers make a difference? Does professional (i.e. academically informed) instruction lead to improvements in students' academic writing skills within the short timeframe of an intensive summer programme? However, this is not the only aim of the thesis. The findings of the thesis will clearly be of practical pedagogic relevance as well as being of theoretical and professional-political interest. Accordingly, it will be an additional aim of the study to consider the possible implications of the research reported in this thesis for EAP teaching, course design and assessment.

1.3 Outline of the thesis

In Chapter Two, I review research on writing development and show that until relatively recently, this body of work has been generally pessimistic about the possibility of significant language change occurring as a result of explicit instruction in the short term. I then go on to describe some very recent studies which have employed corpus methodology to suggest that there are certain characteristics of academic writing which may indeed show development over the limited time frame of a preessional EAP programme.

Chapter Three describes a pilot study that was conducted in order to establish the viability of the larger scale corpus-based project as envisaged by this thesis. This pilot study involved the analysis of a small corpus of matched pre- and post-course writing samples taken from University of Birmingham preessional EAP students. Wordsmith Tools 5 (Scott,

2008) was used to investigate some predetermined features which were considered to be indicative of development. These features included mean word and sentence length, error free units and lexical density indicators. The word frequency component of Wordsmith Tools was also used to investigate frequencies and rankings of some high frequency lexical items from the corpus. The findings of the pilot study were very positive and seemed to contradict the pessimistic view in much of the SLA literature regarding the likelihood of students making rapid gains in second language writing development. This indicated that a larger scale corpus-based project would indeed be worth conducting. Chapter Three concludes by introducing the English for Academic Purposes Corpus (EAPCORP), the main analysis corpus of the thesis, which incorporates a far greater number of matched data samples than the pilot study.

Chapter Four presents the first of two large scale computer assisted analyses of EAPCORP. The analysis reported in this chapter is based on the Coh-Metrix software developed by McNamara *et al.* (2005; 2011; 2013; 2014). Coh-Metrix is a quantitative tool that supplies a detailed, multifaceted set of quantitative data across a range of dimensions or aspects of writing development. The range of features is elaborated, together with its macro or principal components and its application to the EAPCORP is outlined with the subsequent results reported.

In Chapter Five, I report on the second analysis that was carried out on EAPCORP. This analysis uses the Multidimensional Tagger software developed by Andrea Nini (Nini, 2014) as an implementation of multidimensional analysis as pioneered by Douglas Biber (1988). This second analysis offers an opportunity to triangulate the results of the Coh-Metrix analysis as reported in Chapter Three. Multidimensional profiles are obtained for the pre-and post-course data sets for each cohort (20, 15, 10 and 6 week courses) using the MAT software. The chapter concludes by arguing that changes in these profiles constitute evidence

of student writing development.

Chapter Six summarises the findings of the two main research Chapters (four and five) and identifies areas of agreement across a range of linguistic features, as well as noting areas where there is no discernible movement. It is argued that the results of the Coh-Metrix and MAT analyses are highly complementary, and that, taken together, they provide strong empirical grounds for claiming that the student writers represented by EAPCORP make measurable and quantifiable improvements in their academic writing skills during the course of the preessional EAP programmes in which they are enrolled.

The seventh and final Chapter of the thesis considers the broader implications of these feature movements for syllabus design, especially regarding preessional EAP courses, itemising the characteristics which could be considered worthy of inclusion in such programmes by virtue of their prevalence and tractability. The identification of tractable linguistic features is considered as bearing positive implications for course syllabus design, and a case is made for the consideration of the results of this study when planning an EAP programme which emphasises writing development. The chapter also considers the implications of the research for the status and professionalism of English language teachers and the significance for external validation organisations such as the British Council. It suggests that all EAP teachers who work in universities should be *research informed*. They should keep abreast of recent research in their field, attend and present at conferences and indeed carry out investigations themselves into the efficacy of their teaching and course programmes. The position of the current study is that this is to be expected in reputable institutions and to be of benefit to universities, teachers and students alike, not only because of its intrinsic worth but also as a defence against those who would devalue high quality English language programmes, preferring to offer rudimentary courses with unqualified staff, probably paid at unprofessional rates. To reiterate, the standpoint of this study is that

it is important that EAP and language teaching in general is *not seen as something anyone can do* with minimal preparation and minimal awareness of the issues involved, one of the purposes of the thesis being to defend the profession of EAP teaching specifically, and by association, the teaching profession in general.

CHAPTER TWO

ISSUES IN SECOND LANGUAGE WRITING DEVELOPMENT

2.1 Introduction

The purpose of this chapter is to conceptualise the topic of second language writing development and to outline a theoretical framework within which the complex issue of measuring student progression can be placed. Particular emphasis will be placed on previous research which has a direct bearing on the question of whether measurable improvement in student written production can be achieved over a relatively short time frame such as that of a 6 to 20 week preessional programme.

As we will see, the generally pessimistic view promoted by earlier classroom-based studies and SLA theory is now starting to be challenged by recent corpus-based research, which seems to offer greater grounds for optimism regarding the possibility of real improvement over such a short time period, as well as a more sophisticated and robust set of methodological options for studying EAP learner writing than were available to researchers in the past.

2.2 Conceptualising writing development

Many, if not most, English for Academic Purposes (EAP) teachers believe that their students' writing does improve as a direct result of the classroom instruction that they

receive on intensive preessional programmes. It is reasonable to assume that this belief is largely based on the teachers' own direct observations of their students' written work during the course of their studies. To illustrate this, consider the following two writing samples, which were produced by the same student at the beginning and the end of a 6 week course of intensive EAP tuition. Each sample consists of the opening two sentences of an essay that the student was required to write, under timed examination conditions, on the subject of the differences between the UK and China:

Sample 1 (beginning of 6 week programme)

I come from China which also old and beautiful like United Kingdom. We appreciate or history culture and arts well known all round world.

Sample 2 (end of 6 week programme)

It has been said that if you want to understand a county's culture deeply, you need to study there for a period. Culture and motivation for improving my professional skill are the two main cause for me to decide to study in English.

When these two samples were presented at a professional interest meeting of the British Association of Lecturers in English for Academic Purposes (BALEAP) (Issitt, 2008), there was unanimity as to which script was better and general consent amongst the audience that progress had indeed been made. Sample 2 was widely regarded as 'better' than Sample 1, which was judged to be quite rudimentary although the meaning was considered to be reasonably clear. Such judgements are commonly made by teachers and may be informal in character although often forming the basis of highly formal assessments and reporting upon students' progress. What is less certain, however, are the empirical bases for these informal judgements. This is not to say that such assumptions are necessarily wrong; on the contrary,

in the EAP context, an impressionistic observation such as “this writing has improved” may in fact be largely correct. What is lacking, is a clear and detailed definition of what it is that has improved, and of what aspects of writing may be conceptualized in terms of ‘improvement’ at all.

If it can be shown that certain text characteristics exhibit significant differences after a period of intense language instruction, then the concepts of improvement, development and progress can be given more substantial, quantitatively referenced support. To give an initial example of an observable linguistic feature, the capacity to write longer sentences could be said to be congruent with writing development and this may well be characteristic of maturing, well taught writers. While this is not necessarily true, as successful academic writing may, sometimes at least, consist of shorter, less rambling sentences, it could be the case that, as a student gains knowledge of the grammar of written academic English, their ability to produce longer sentences may result. Alternatively, or additionally, it could be proposed that the production of more long words (which tend to be rarer and more specialized in their meaning) may also be expected from a student exposed to an intensive course of EAP study. Another example might be changes in the number and types of error produced by students over the course of their studies, although this is a contested area in applied linguistics (see, for example, Brown 1987; Seidlhofer 2003; Thewissen 2015). Teachers are expected to correct their students’ written work and to invoke a reflection upon what has been produced which is “wrong”, grammatically or lexically, for example. It might therefore be expected that after a course of study, the number of errors would be lower, and a simple pre- and post-course error count could offer an indication of progress. Regarding specific lexico-grammatical characteristics, it might be expected that a feature such as the passive voice would show an *increase* in use after a defined period of study or that first person pronouns would show a *decrease*. We might also anticipate movements in other

specific features such as nominalisations, use of relatives, increased noun modification, and so on almost ad infinitum. The question to be addressed here is: which of the many possible language features that could be measured should be included in an analysis of preessional EAP student writing development, and on what theoretical basis?

2.3 A conceptual paradigm

A clear starting point for the analytical framework to be developed for this thesis is provided by Shaw's and Liu's (1998) study into the nature of second language writing development. This study, which focuses upon the characteristics of written English development over a short time frame, proposes that development can be conceptualized along three clines: inaccurate to accurate, simple to complex and spoken to written. Students' early work, for example, may contain many errors which if evaluated might attract low accuracy scores on assessment schedules. Similarly, a student's initial offering may be marked by a conspicuous simplicity, characterised by an accretion of short grammatically unrefined, almost childlike sentences with very little cohesion between them. Again, allocation of scores to such a contribution would probably be in the low band of any marking scheme (for example IELTS and Cambridge main suite TEFL exams). Students may also produce writing at the beginning of their programmes which can be described as more "oral" in character and that after instruction a more generally "written" style takes shape. Gilquin and Paquot (2007), for example, identifies a distinct tendency among non-native speakers to use the *spoken rather than the written form of English* e.g. using say as a reporting verb and amplifying adverbs thus creating a more informal academic voice (the study also associates this characteristic with novice writers either L1 or L2 and considers that it may be part of the process of acquiring an academic voice which needs to be lost in speech and gained in

writing)

These dimensions, inaccurate to accurate, simple to complex and spoken to written, can be seen as explicitly or implicitly framing, or at least informing assessments of EAP student writing proficiency. To illustrate, Figure 2.1 is an extract from the University of Birmingham Preessional Academic Research Paper (ARP) marking criteria:

Figure 2.1 Marking criteria for Academic Research Paper

BAND 1 Mark range 16-20 (max 20)

Organisation (text structure and cohesion)

The text is a recognisable, well-knit whole, with a consistent development of ideas and propositions throughout. Successful introduction and conclusion. Alternative points of view are presented and developed logically with a clear progression of ideas that are relevant and plentiful. Sophisticated, native-speaker-like use of linking between sections makes the text flow seamlessly.

Range (sentence structure, word choice and cohesion)

A wide range of sentence structures and word choice. The message can be followed effortlessly, and cohesive devices within and/or between sentences are skilfully managed so that no attention is attracted. At the bottom end, occasional slips or minor infelicities are tolerable, but there is nothing revealing serious ignorance.

Accuracy (grammar, word choice, spelling, punctuation)

Standards of grammar, word choice, word formation, spelling and punctuation are consistently of a very high level. Language is error-free, fully appropriate to the task with a compellingly communicative style. At the bottom end, occasional slips or minor infelicities are tolerable, but there is nothing revealing serious ignorance.

Register (formal language and academic conventions)

The writer is in total control over information from outside sources in a variety of ways. The text displays full and appropriate documentation and especially effective presentation. Use of language is consistently appropriate to the academic task, content & intention. Makes wholly appropriate use of source texts; bibliography follows conventions. At the lower end, occasional slips or minor infelicities of style or register mishaps are tolerable, but there is nothing revealing serious ignorance

These descriptors appear to reflect Shaw and Liu's typology to a certain extent. There is reference to accuracy (for example "language is error-free"), complexity ("a wide range of sentence structures") and the adoption of conventional or formal written as opposed to informal spoken style ("bibliography follows conventions"), and these bear an implicit acceptance that students' work can be described as having improved or otherwise. If a student's work was adjudged to be in the above band, then the lowest mark would probably be 64 (4 x 16) and the highest 80 (4 x 20). A score of 64 marks from 80 would represent an adequate level of written competence according to these criteria, with only minor infelicities of accuracy combined with good use of cohesive devices and good awareness of register representing at least a satisfactory level of writing. The four dimensions which frame the marks awarded (organisation, range, accuracy and register) and the descriptors themselves are wholly concerned with English and academic conventions with no reference to content except in an organisational sense. This is a deliberate policy which reflects the view that the course is an English language programme which is not engaged in teaching the content of the various subject disciplines that students will eventually go on to study for their postgraduate degree programmes. This language-directed set of criteria reflects the concern that students are able, by the conclusion of their programmes, to perform at a level of linguistic competence which will allow them to function as students at postgraduate level through the medium of English. Failure to reach the relevant levels of competence (and this varies, depending on the target subject) means that a student will be classified as unfit for postgraduate study, as the following descriptors from the lowest end of the scale as shown in Figure 2.2 illustrate:

Figure 2.2 Marking criteria for ARP

BAND 5 Mark range 0-3 (max 20)

Organisation (sentence structure and cohesion)

The text is not a recognisable whole. Inadequate or absent intro and conclusion. Little sense of organisation of text or progression of the ideas present

Range (sentence structure, word choice and cohesion)

Inadequate range of sentence structures and word choice. The message is difficult to follow and cohesive devices are inadequate or missing. Inaccuracies in sentence construction predominate and the writer's inadequate syntax mostly obscures meaning. A limited vocabulary is used

Accuracy (grammar, word choice, spelling, punctuation)

Frequent errors of grammar, word choice, word formation, spelling or punctuation cause severe strain for the reader, even within simple sentences. More is wrong than it is right

Register (formal language and academic conventions)

The writer has made little or no attempt to control information from outside sources and no voice is discernible. There are serious deviations from conventions regarding paragraphing, layout, section headings, short and long references. Reading is made difficult by unsubtle and inappropriate language. No use of referencing/clear evidence of plagiarism/no bibliography.

As can be seen, the descriptors specify certain language features which are presented as benchmarks for the award of marks and placement within the relevant bands and again reflect the framework suggested by Shaw and Liu (1998). For example, range of sentence

structure and word choice relates to the development of *complexity*, frequent errors of word choice and word formation relate to *accuracy* and control of information from outside sources relates to the *oral/written* dimension. If a student's work was adjudged to be in this (lower band), then the lowest mark would be zero (4 x 0) and the highest would be 12 (4x3). Twelve marks from 80 represents an inadequate level of written competence with inadequate syntax and limited vocabulary indicative of written production at below satisfactory level for postgraduate study. Having established that there are discrete language features which can be used as an empirical basis for measuring improvement in L2 student writing, it is now necessary to examine studies which have attempted to isolate and measure some of these characteristics in more detail.

2.4 Research on writing development

As Ortega (2003, p.515) points out, it can be difficult to offer generalizations about the findings of empirical studies, because the “elicitation characteristics of particular writing tasks, as well as sample size, timing of the writing, corpus length [if used] and the target languages investigated all [differ] considerably” from one piece of research to another. To this could be added the contexts (for example EFL or ESL) and the genres and variables selected for analysis. Unsurprisingly, given the large number of studies in this area, the literature could be characterised as piecemeal in that specific features have been described and interpretations have been suggested but no coherent picture has yet emerged. Nevertheless, it is possible to impose some degree of coherence on this field by classifying the literature into three main sections: studies with a lexico-grammatical emphasis, studies with a cognitive/developmental focus, and a section on corpus-based approaches to L2

writing development research. Each of these will now be reviewed in turn.

2.4.1 Studies with a lexico-grammatical emphasis

The section begins with a consideration of studies which have focused on vocabulary, and moves from individual word to phrase to multi-word items and then any relevant macro textual issues before reviewing explicitly grammatically-oriented research. Beginning with vocabulary, as Nation and Waring (1997) point out, the most frequent 2000 words in the English language comprise 80% of text coverage, with a falling away of coverage once the intermediate stage is reached. “With a vocabulary size of 2,000 words, a learner knows 80% of the words in a text which means that one word in every 5 (approximately 2 words in every line) are unknown” (Nation and Waring 1997, p.17). This study whose emphasis is primarily upon reading has implications for the production of written text in the sense that second language writers are faced with a finite task. The list of possible words is enormous but a working vocabulary can be readily acquired, with a key issue being that of *time*, and this is encouraging for course designers and teachers. Nation and Waring refer to a study by Milton and Meara (1995) which was concerned with vocabulary growth over a 6 month period amongst university undergraduates from various language groups (German, French, Spanish and Italian) and showed an increase in vocabulary acquisition rates from 500 - 600 in the 6 months prior to departure on a study abroad programme to over 1300 during the study period itself. This increase was attributed to the combined effects of tuition and immersion with the writers concluding that “an extended period abroad does have a real effect on a student’s linguistic competence.” (Milton and Meara 1995, p 24). Whilst offering support for vocabulary development over a 6 month time

frame, averaged at 2500 words per year, the study did suggest that those students with relatively low vocabulary levels (between 2500- 4000 words) showed larger gains than the higher level learners in the study. Milton and Meara's study could be described as concerned with *lexical magnitude*, used a measure of vocabulary size with a 10,000 word upper limit denoting a near native level of competence. A note of caution is sounded by Laufer and Goldstein (2004) who in referring to *vocabulary knowledge*, hypothesise a hierarchy of difficulty (from the easiest, passive and active recognition, to the hardest, passive and active recall) and that any consideration of vocabulary size, especially over a short time period should take this into account. Again, considerations such as vocabulary size and vocabulary knowledge, although of primary application to the reading context, bear direct relevance to the development of a workable vocabulary resource which may be expected to increase over the course of an intensive EAP programme and which can be employed by second language writers to produce higher quality written work.

Yasuda (2011) examined the results of a 15 week writing course for Japanese biology related undergraduate students including email writing tasks using pre and post tests and concluded that students showed improvement in the awareness of genre. The study suggested that writing also improved in terms of task fulfilment, appropriacy, cohesion, grammatical control fluency and language sophistication but *that vocabulary size and lexical diversity* did not increase significantly over the time period. Yasuda offers the view that vocabulary items in student writing may become more contextually meaningful rather than greater in number; this presents an opposition between what Yasuda refers to as lexical sophistication on one hand and lexical diversity on the other. Drawing attention to the time frame in the acquisition of a workable vocabulary resource, Yasuda considers that "attaining substantial development in learners' vocabulary knowledge is a significant challenge in a one semester course. However, ...fifteen weeks may be significant for obtaining significant

improvement in language sophistication within a particular domain” (Yasuda, 2011, pp. 125-126). An important aspect of this study is that students took English classes for 1.5 hours per week making a total of 22.5 hours over a 15 week period. This stands in sharp contrast to an intensive university preessional course of 20 hours per week and invites consideration of the intensity and number of teaching hours of a language programme, in addition to the position of such learners in target language culture, for example the overall academic environment, which as Churchill and Dufon (2006) observe can have a strongly beneficial effect on Japanese learners. In a study, which focused on the definite article and simple past tense, Bitchener, Young and Cameron (2005) in a study of 53 mainly Chinese adult migrant learners of English, investigated the effectiveness of written corrective feedback and found that over a 12 week period, these items were produced with greater accuracy in writing produced at the end of the programme which included substantial amounts of individual written and conference feedback. Other categories such as prepositions showed no increase in accuracy, however, and the writers hypothesise that this may be because the use of the past simple and the definite article is more rule-governed whereas the use of prepositions is more idiosyncratic and therefore less amenable to correction.

In terms of the classification and measurement of vocabulary development, there are various measures which have been employed, for example Laufer and Nation’s (1995) Lexical Frequency Profile. Laufer and Nation also provide a critique of measures and indices designed to assess lexical richness, a broad and sometimes undefined classification which is useful when conceptualising the wide field of vocabulary acquisition and the measures include a variety of related terms such as lexical originality, lexical density, lexical sophistication, lexical variation and lexical quality. *Lexical originality* refers to the uniqueness of the piece of writing by measuring the number of words which are specific to

that particular writer. It compares the writer to the rest of the group and as the reviewers point out may have only context specific reliability. *Lexical density* (LD) measures the percentage of lexical words (nouns, verbs, adjectives, adverbs) in the text, a clear and frequently used measure which does attract criticism that it may not reflect relative syntactic complexity. In other words, it is possible to produce a sophisticated piece of writing with a lower lexical density. The pilot study of the current thesis, which is described in full in Chapter Three, examined this feature and suggested an increase in LD over a relatively short time period, but as an indicator of writing development it should probably be used cautiously.

Lexical sophistication refers to the percentage of advanced words in the text as defined by the Academic Word List (Coxhead 2000). *Lexical variation*, often referred to as the type token ratio (TTR) calculates the number of types of words as a proportion of the number of words/tokens in a text. Again, this would appear to give an effective measure of writing development, but is notoriously sensitive to text length; as a general rule, the longer the text, the lower the TTR. Researchers using TTR also need to consider and adopt a clear policy towards the issue of word constitution, for example whether derivatives of a word such as *PUSH* (i.e. *push/pushes/pushing/pushable*) are to be counted as different words or as belonging to a single word family. (In corpus linguistic terms, the question is whether the researcher should count word forms or lemmas.) Laufer and Nation consider that if the former is used then a high lexical variation may not be indicative of lexical richness. The learner would know only the lexical family although it might be indicative of a morphological flexibility or at least a willingness to try other forms of a word and when testing, for example vocabulary size, then LV would need qualification (see also Richards 1987 for a discussion of TTRs specifically in the context of child language development).

Finally, *lexical quality* is identified by Laufer and Nation as a composite of type and

rare words divided by lexical error. It could also be the case that aggregated lexical measures such as those described are difficult to employ *manually* as a measure of writing development but that measures of lexical development could be more effectively used in a battery form, and a composite index might be created which could offer a stronger overall indication of writing improvement, at least in the lexical sense and this could be a goal for future second language writing assessment.

Turning now to studies focusing above individual word level, there are a variety of terms which have been applied, sometimes interchangeably as descriptive frames, for example lexical phrases and formulaic sequences (pre-assembled multi word chunks) lexical bundles (words which are often located together within a specific register) and n-grams (a sequence of items such as letters and words occurring within a common border) are often used, sometimes interchangeably. Li and Schmidt for example (2009) examined the acquisition of lexical phrases over the period of a *one-year MA programme* for a single Chinese student. They conclude that although new phrases were learned (166 in total) together with an increase in confidence, there was a heavy reliance on a limited range of expressions and no increase in diversity or frequency. The writers also stress the repetitive nature of much of their subject's production and call for programmes which focus on building "more diverse phrasal lexicons" (Li and Schmidt, 2009, p.100). Pointing to previous work by Hyland (2008), Li and Schmidt stress the importance of formulaic sequences such as "*as a result*" and "*it should be noted that*" as "building blocks" (2009, p.86) in the formation of academic texts and consider that if second language learners cannot use these sequences effectively, it may be difficult for them to develop into competent academic writers. Jones and Haywood (2004) looked at a university 10 week preessional programme, finding some evidence of development in the acquisition of formulaic sequences but generally little significant improvement in overall composition writing. We

will return to research on student writers' use of formulaic sequences when we review corpus-based studies in Section 2.4.3.

Turning now to research with a more clearly grammatical emphasis, in a study of the use of English articles by 18 Chinese learners of English, Robertson (2000) confirmed the tendency of many Chinese learners to omit the article in English where a native speaker would use one and identifies a 'remapping' principle which arises from a need for the Chinese learner to move from a discourse oriented (Chinese) to a syntax oriented (English) grammar and that the Chinese learner needs to adjust the mapping between the two semantic systems. This study has direct bearing on the current thesis not least because around 60% of the total data is derived from Chinese students (see section 2.4.3 for a further discussion of Chinese learners and corpus issues). Stanat *et al.* (2012) examine a compressed time frame, namely *3 weeks*, in their study of the effects of a summer course in L2 German for schoolchildren. They conclude that explicit and focused German language instruction produced a significant improvement in grammar as measured by pre and post-test assessments in terms of grammar (articles, prepositions, verbs and tenses) and reading (using standardised German school based-tests) after a programme which emphasized, amongst other things, the structural aspects of language. Although this study was aimed at schoolchildren at 15 years of age, it does offer a suggestion that real improvements in L2 performance may in fact be possible with reference to certain specific linguistic features after a very short time period. Spalding, Wang and Lin (2009) looked at writing development over a *three-week* study programme. Fifty seven Chinese teachers of English were studied during daily writing workshops and data was drawn from pre and post writing samples. Improvements were observed in writing quality by reference to a general set of descriptors which included organisation, ideas, voice, word choice, sentence fluency and conventions and noted improvements in all the characteristics apart from conventions with the greatest

increase being in voice, “the distinctive, individual way in which the writer speaks to the reader” (Spalding, Wang and Lin 2009, p. 23) over the three week period. This is a relatively rare example of a study which has suggested general writing development over a short time period and is significant to the extent that the workshop programme specifically targeted general features of writing which were considered to be important for second language writers (in this case teachers) of English.

On a broader level, various studies have examined features of syntactic complexity such as sentence and clause length, subordination, coordination and the range and complexity of structures. Weston, Crossley and McNamara (2010) for example, delineate just two measures of syntactic complexity, the mean number of words (a relatively simple metric) and the noun overlap (how often a noun of the same form is shared between sentences) in a study of High School L1 ‘freewrites’ (the term is used here to refer to a freewriting exercise over a set time period and with as much information transmitted as possible without specific attention to grammatical rules) and the two measures were identified as clear predictors of writing quality. This study, although non developmental in emphasis, does offer the possibility of using two marked indicators to assess writing development at least in terms of syntactic complexity. Collentine (2004) examines morphological and lexical complexity and uses past tense verbs and past and present participles to explore differences between in-class and study abroad L2 Spanish learners. The study suggests that complexity may not develop in a linear fashion and as such, a progression curve of syntactic/lexicogrammatical development may be a desirable abstraction, but not represented by reality. Larsen-Freeman (2006) points to the often non-discrete nature of L2 development which may include a regression of certain features. The study, although of only 5 (Chinese L1) learners over a 10 month period, points to the need for caution when making generalisations about progression and development and suggests

the importance of an awareness of the individual nature of development and that not all learners may follow the same improvement path.

The time factor in grammatical improvement is clearly of great importance to the current thesis. On this question, the overall picture is varied, with several studies suggesting that more than one term/semester is necessary for development of accuracy and complexity. In a research synthesis of 25 studies of L2 writing containing a variety of sample sizes, observation periods and study contexts, Ortega (2003, p.492) notes that the longitudinal evidence available at the time was limited and concludes that “...an observation period of roughly a year of college level instruction is probably needed for substantial changes in the syntactic complexity of L2 writing to be observed” This is a significant observation, not least because many English language programmes, especially of an EAP character take place over a much shorter time frame. The research synthesis attempts to isolate algorithms which could reliably quantify the relationship between L2 proficiency and L2 writing competence, identify critical magnitudes between proficiency levels and to measure the rate of change in syntactic complexity which she considers to be relevant to growth of an “L2 learner’s syntactic repertoire” (Ortega, 2003, p 492). In addressing the issue of a typical or normal rate L2 syntactic development in L2 writing, and in using the mean length of T-Unit (main clause and modifiers) as an indicator of syntactic complexity, the study concludes that there may be evidence of small MLTU developments over a two to three month period. However, as indicated, a one year period is required before substantial change can be observed. Addressing the issue of syntactic complexity directly, Lu (2010) offers a computational system for analysis and provides 14 indices of syntactic complexity which are drawn from a composite of previous studies notably Ortega (2003) and Wolfe–Quintero, Inagaki and Kim (1998). The review firstly presents Hunt’s (1966, p. 737) definition of a T-Unit as:

the shortest unit into which a piece of discourse can be cut without leaving any sentence fragments as residue. Thus, a T-Unit always contains just one independent clause plus many more subordinate clauses that are attached to the independent clause.

Designed initially as a means of distinguishing between proficiency grade levels, the system is proposed as a means of “analysing the syntactic complexity of any number of writing samples using any or all of the 14 complexity measures. Lu’s system inputs written texts, analyses the syntactic structure of each and then outputs 5 types of measure which include length of production unit (clause, sentence and T-Unit), sentence complexity expressed as a ratio, subordination (T-Unit complexity, T-Unit ratio, dependent clause ratio, dependent clause per T-Unit) coordination (coordinate phrases per clause/per T-Unit, sentence coordination ratio) and particular structures (complex nominals per clause/per T-Unit, verb phrases per T-Unit) containing 14 different indicators of syntactic complexity. As a means of measuring syntactic development, the system would appear to offer an effective set of indicators if limited to the syntactic field of enquiry. For a broader picture of learners’ overall linguistic development, another research tool may be required.

In another study regarding the T-Unit as a measure of syntactic and lexical complexity, Jiang (2013) offers a summary of a study by Wolfe-Quintero (1998) which consisted of a survey of 39 studies measuring L2 development. The study, while offering a framework for examining written development as a whole, includes measures of fluency (speed of production), accuracy (error free production) and complexity (a variety of sophisticated structures and vocabulary) and considers that clauses per T-Unit and dependent clauses per clause to be effective measures of grammatical complexity.

Wolfe-Quintero, Inagaki and Kim (1998) in attempting to correlate linguistic measures of writing development focus on fluency, grammatical complexity, lexical

complexity and accuracy and offer the summary of measures (in no rank order) reproduced in Figure 2.3.

Figure 2.3 Measures of L2 development (based on Wolfe-Quintero, Inagaki and Kim 1998)

dimension	measure
fluency	words per T-Unit; words per clause; words per error-free T-Unit
grammatical complexity	clauses per T-Unit, dependent clauses per clause
lexical complexity	word type measure, sophisticated word type measure
accuracy	error free T-Units, errors per T-Unit

A study by Knoch *et al.* (2015) examines the written development of undergraduate second language learners in an English medium university over a *three year period*. The study concluded that at the end of the three years there was *no evidence of writing improvement in terms of accuracy or grammatical and lexical complexity*, and only fluency (number of words per essay) showed any increase. *Accuracy* was measured by the percentage of error-free T-Units and clauses, *grammatical complexity* by the average number of words per clause, clauses per T-Unit and the ratio of dependent clauses to all clauses. *Lexical complexity* was measured by percentage of words from the Academic Word List (Coxhead 2000), lexical sophistication (using words derived from an AWL sublist and divided into the number of content words) and the relatively simple metric of average word length. The authors of this pessimistic study suggest that universities need to review aspects of their provision for second language learners and recommend close attention to feedback guidelines by higher education institutions. Until fairly recently much of the research into the lexico-grammatical aspects of second language writing development could be classified

as a work in progress with some studies emphasising specific features and others using a broader approach.

Over the last few years however, there has been considerable input into the specific issue of syntactic complexity and in a review by Ortega (2015) the expanding nature of this field of inquiry is highlighted and the conventional conception of syntactic development as a means of indexing “the expansion of the capacity to use the additional language in ever more mature and skilful ways, tapping the full range of linguistic resources offered by the given grammar...” (Ortega, 2015, p.82) is stressed. In considering the relationship between written syntactic complexity development and instruction (teaching) Ortega draws attention to a study by Mazgutova and Kormos (2015) who found that over a four week, sixty hour preessional English course for university undergraduates, there was evidence of greater noun modification, more complex nominals, syntactic structure similarity, conditionals and relative clauses. This particular study has relevance for the current thesis especially as it uses *data from such a limited time frame and employs pre and post course essay analysis* concluding that by the completion of the programme, the students had produced writing which exhibited “a more advanced repertoire of lexical and syntactic choices that are characteristic of expository texts in academic contexts” (Mazgutova and Kormos, 2015) It offers some optimism that development in second language written English is indeed possible over a limited time period.

2.4.2 Studies with a cognitive/developmental emphasis

In the previous section, we looked at studies which focused on the writing itself and the main emphasis of this section is upon the developmental and cognitive aspects of writing in a

second language with reference to those studies which have investigated improvement in competence over a defined period of time. In this sense the focus is now more upon the learner than the production and there will be reference to certain theoretical positions regarding language acquisition, a consideration of the issue of error as it impacts upon second language writing development and reflection on the description of L2 proficiency levels. The first question to tackle here concerns whether EAP students' writing development could take place without any formal instructed language instruction at all, particularly if the student is already living and/or studying in an English-medium environment. One direct attempt to answer this question is offered by Storch (2009), who studied 25 mainly postgraduate East Asian students at an Australian university and concluded that the results of pre and post course tests indicated that after one semester of study at an English medium university *without access to any of the formal English language programmes*, there was “no evidence of improvement in linguistic accuracy or complexity” (Storch, 2009, p.103). The writer considers that “mere immersion in the L2 context and incidental learning will not necessarily lead to improved language use particularly after only one semester” (Storch, 2009, p.116). This observation prompts the consideration that *the course itself* may be a key variable in the development of linguistic competence and may point the researcher towards examination of the content and efficacy of the ELT programme being taught and it certainly invites consideration of the specific features which comprise a course syllabus and of the emphasis course designers may place upon informing students about the nature of academic writing.

Looking now at the broader aspects of second language writing acquisition, a useful cognitive perspective is afforded by the dynamic systems approach (for example, Langacker, 2000 and 2009; De Bot, Lowie and Verspoor, 2007) which has underlain certain recent studies in L2 writing development. From this position, learners understand the features of

a second language through exposure and levels of development are crucially affected by previous stages. From a dynamic systems perspective, the acquisition of an L2 depends on an aggregation of factors such as morphology, lexicon, collocations, formulaic phrases and constructions. These factors and the interaction between them are seen as crucial in L2 development with no one factor being paramount. From this perspective, cognition rather than innate systems is key and language learning involves the mastery of an inventory of patterns. Once a pattern is mastered it can then be abstracted, a process which involves “the progressive retrenchment of recurring configurations” (Langacker, 2009, p. 628).

One such aggregation of factors is evident in a study designed to explore assessment and development of the writing of over 430 Dutch learners of English as an L2. Verspoor, Schmid and Xu (2012) offer a 64-variable framework including sentence, clause and verb phrase constructions together with lexical items and chunks over 5 different stages of proficiency. This study attempts the threefold task of describing the text features which occur at the relevant stages of second language development, identifying characteristics which can be used to discriminate between stages, as stage markers, and also to study the L2 development process in the wider sense by looking at changes in (syntactic) systems and feature use. The study, which emphasises the difficulty in developing and establishing a “common yardstick to measure ... (writing) proficiency” (Verspoor, Schmid and Xu, 2012, p. 258), offers some challenging conclusions; firstly, that in their study, nearly *all types* of constructions were used from level 1 to level 5 (beginner to advanced) but that beginners use them less frequently, with less consistency and with more errors and on a common sense basis. This is possibly to be anticipated. The second main finding is that their more advanced writers did use new constructions (including advanced vocabulary); they increasingly use the more complex recently learned constructions and that “*as proficiency increases, they make fewer errors*” (Verspoor, Schmid and Xu, 2012, p. 258). The study also suggests that

there are regular stages though difficult to clearly define, which the learners pass through despite variation and variability in the overall process of language development. Figure 2.4 is a diagrammatic summary:

Figure 2.4 Second language writing proficiency levels (Verspoor, Schmid and Xu, 2009, p.258)

<i>general</i>	levels 1 to 5 simpler constructions gradually disappear and more complex ones replace them
<i>level 1</i>	predominantly simple sentences with present tense, mainly lexical changes between levels
<i>level 2</i>	focus on vocabulary and chunks, mainly syntactic changes between levels
<i>level 3</i>	reorganisation of syntactic system, lexical and syntactic changes between levels
<i>level 4</i>	main syntactic changes in place
<i>level 5</i>	focus on vocabulary and chunks

By offering this framework of writing development stages, Verspoor, Schmidt and Xu also focus on *what it is that may change* during a programme of language instruction; clearly, this makes their framework of considerable interest and relevance to the current thesis. Also of interest are issues relating to the *rate of change of* language features. Spoelman and Verspoor (2010) in a study of the acquisition of writing competence in Finnish, point to the non-linear development of L2 and draw attention to the rapid rate of accuracy improvement combined with increases in complexity, with noun phrase and sentence complexity competing with each other to some extent. The picture presented here is that accuracy and complexity levels “are characterised by peaks and regressions, progress and backsliding and by complex interaction between variables” (Spoelman and Verspoor, 2010, p.551) The writers also highlight the need to research individual learning paths or “developmental trajectories” (p.551) as an opportunity to gain insights into the second language learning

development process.

A similar issue is tackled by Spinner (2011, p. 530) who calls for a description of “a clear path of grammatical development that can be used as part of an assessment tool” and invites the development of a theoretically grounded set of indicators which can empirically justify the rubrics and descriptors relating to particular levels of L2 development. Drawing on the work of Pienemann and Brindley (1988) and Vanikka and Young–Scholten (2006) Spinner presents a match of language levels, “can do” statements and some relevant morphosyntactic elements with a view to developing an assessment tool. This is illustrated in Figure 2.5 by using selected items at the low and high ends of a scale which also includes intermediate mid, intermediate high and advanced low levels.

Figure 2.5 Indicators of language levels (Spinner, 2011)

level	Description	morphosyntactic elements
intermediate low	strong influence of L1 syntax can ask a few questions can accomplish uncomplicated communicative tasks	L1 word order formulaic questions uses subjects, pronouns, copula
superior	can separate main idea from supporting information can construct and develop hypotheses alternative possibilities abstract elaborations	adverbs, relative and noun clauses if/although until etc. passives

Although Spinner’s data consists of a mixture of oral and written samples, there is a clear target, and a coherent description of grammatical improvement which Spinner refers to with possibly deliberate overstatement as a “holy grail” (Spinner, 2011, p.530).

Another and recurrent perspective on the developmental aspects of second language writing acquisition is represented by the concept of error and the associated idea of accuracy. In many of the studies, these two concepts are often conflated so that a discussion of development in student's accuracy positively correlates with a reduction in error production. A study by Chandler (2003) directly assesses the results of a *fourteen week* undergraduate English language programme, using twenty five page homework assignments and focusing on grammatical and lexical error. In outlining a variety of correction strategies, Chandler suggests that concentrated error correction relating to *twenty six different error categories* can produce significant student improvement without sacrificing fluency. By fluency, Chandler here is referring to the relatively simple metric of text length and Chandler records an average improvement profile of 2 errors per 100 words and a slight reduction in the time taken to produce the same number of words. Chandler's study is of considerable interest and relevance to the current thesis because it draws attention to the fact that these error and time reductions which have taken place over a fourteen week period have in fact required only 24 hours of direct teaching (2x two hour sessions per day). Also, while relating to structured rather than spontaneous student production, Chandler's study offers a glimpse of what can be possibly achieved with a more intense course programme.

Another relevant perspective on the issue of error is provided by Bitchener (2005) who refers to treatable and less treatable categories of error and calls for teachers to identify these for students and to provide both direct and indirect feedback on the more treatable type. The notion of accuracy improvement itself is problematised and the issue of consistency is raised, the idea that it is possible to produce more accurate writing on a specific occasion, but then slip back into repeated error on another. Bitchener calls for investigations *over several semesters* and more research in general into the efficacy of written communicative feedback.

Yet another perspective is offered by Wigglesworth and Foster (2008) who, in addressing the issue of error and accuracy, focus on the clause and suggest the use of weighted clause ratios as a measure. Wigglesworth and Foster ascribe weightings which relate to the retrieval of meaning in the text and which can act as assessors of accuracy. The weightings are listed in Figure 2.6.

Figure 2.6 Weighted clause values (Wigglesworth and Foster, 2008, p.36)

error	Definition	weighting
none	accurately constructed clause	1
level 1	minor errors (e.g. morpho-syntactic) which do not obscure meaning	0.8
level 2	more serious errors (e.g. word choice or order) which makes meaning harder to recover	0.5
level 3	errors which make the intended meaning difficult to recover	0.1

This approach, the classification of accuracy by reference to the number of errors per clause appears to offer a quantitative framework for this aspect of text quality and the measure is supported by Evans *et al.* (2014) who in a study of 81 different L2 users (writing for 10 minutes) found that it was effective as a measure of accuracy especially for lower level writers. Thewissen (2015) in an extensive study of accuracy in second language learning offers a considerable degree of elaboration to the concept and identifies a range of error classification methods including obligatory occasion analysis (a percentage measure relating to obligatory suppliance of a morpheme), T-Units (main clause with superordinate clauses, potential occasion analysis (part of speech category errors), error percentages and error frequencies. Error counting methods include number and average length of error-free T-Units, number of words per error-free T-Unit, number of syntactic/morphological/lexical

errors per clause, ratio of relative clause errors to overall number of relative clauses, focus on most frequent error categories, ratio of lexical verb errors to total of lexical words and error to total word ratio. Thewissen emphasises the importance of longitudinal studies to illustrate a range of developmental processes pertinent to the acquisition of specific linguistic features and while offering a range of suggestions especially for the modification of the European CEFR proficiency framework, stresses the significance of "positive errors" as indicative of overall language improvement as a greater range of linguistic feature are attempted, the errors acting as signposts of development (Thewissen, 2015, p.273). This concept of positive error may be applicable to the improvement profiles of second language learners who may, as discussed earlier, exhibit uneven trajectories as more complex structures are attempted, increasing the likelihood of error. In other words, a learner may overreach themselves in attempting to produce a higher level of written sophistication, resulting in structural error, for example incorrect use of the passive.

Another approach to the concept of improvement profiles is offered by Ferris (1994) who examines the language produced by L2 learners at various stages of development and adopts Biber's (1988) concept of dimensions as its classification system. The main emphasis is upon establishing whether identified linguistic features were characteristic of different levels of language proficiency. The result suggested that the more advanced writers did employ the specific items, for example a variety of syntactic patterns and more cohesive items. This study uses a pre-defined set of linguistic categories and applies them to set of written L2 data using 160 ESL texts from 4 language groups (Arabic, Mandarin, Japanese and Spanish) and after statistical refinement, 28 variables were identified which were used as comparators to a high/low assessment carried out by raters. The pre-selected variables were a mixture of traditional (word length, words per sentence, number of words) and sophisticated (deictic reference, reduced structures) types. The main conclusions supported

the divisions between low and high categories of essays and the writers emphasise the pedagogical implications and recommend micro-level attention and instruction in areas of word choice, lexical and syntactic diversity. This perspective, using a range of measures of assessment and classification, leads us to a consideration of a multidimensional approach to writing production and the compilation of a corpus of texts which may illustrate the features of second language writing development.

2.4.3 Corpus-based studies

So far, this chapter has reviewed research that has used traditional SLA methods to study L2 student writing development. Over the last few decades, however, SLA researchers have gradually begun to realise the potential that the tools and methods of corpus linguistics have to offer to this field, and it is to a consideration of corpus-based studies of writing development that we now turn. The principal advantage of corpus-based research is that it allows empirical researchers to conduct analysis of student data on a scale that would simply not be possible using traditional manual methods. This in turn allows the researcher to conduct far more robust and sophisticated quantitative analyses than was possible hitherto. As a simple illustration of the advantages of a corpus-based approach, Biber and Conrad (2001, p.332) draw attention to the fact that while over 400 different verbs occur over 20 times per million, corpus-based studies have revealed that only 63 of these lexical verbs occur more than 500 times per million words and that only 12 verbs occur more than 1,000 times per million.

It should be noted that there are corpus-based studies which have examined many of the areas and issues addressed in the previous two sections (with a frequent focus on

cognitively-oriented research), but a separate section is offered here as the methodologies used in individual studies are often quite different from each other.

Much of the initial impetus for corpus-based approaches to the analysis of learner language stems from the pioneering work of Sylviane Granger and her colleagues at the Université Catholique de Louvain, Belgium. The Louvain team is most strongly associated with the development and exploitation of benchmark learner corpora such as the International Corpus of Learner English (ICLE) (Granger, 1993 and 2003b; Paquot, 2015) and the Louvain Corpus of Native English Essays (LOCNESS) (Paquot, 2015). ICLE is a 3.7 million word corpus of argumentative essays (higher intermediate to advanced level) from a range of L1 users, while LOCNESS is a 300,000-word reference corpus of final high school year and first year undergraduate essays written by British and American students, which enables researchers to compare learner writing with writing produced by native speakers. For example, Granger (1998) used these corpora to compare native English and advanced French-speaking English learners in terms of their use of ‘prefabricated chunks’ of English. Granger found that learners tended to ‘overuse’ certain chunks, which she described as ‘islands of reliability’ or as Hasselgren (1994, p.237) calls them, “lexical teddy bears”.

In another ICLE-based investigation, Granger and Petch-Tyson (1996) hypothesised that logical connectors would be overused by EFL learners in comparison to native speakers, and found that while certain items were indeed overused, many others were actually underused in comparison to the native speaker benchmark. This led them to the recommendation that lists of such connectors be avoided for teaching purposes. Instead, Granger and Petch-Tyson call for a greater awareness of linguistic (stylistic, semantic and syntactic) properties of connectors and an encouragement of students to think more reflectively about these areas rather than simply trying to learn lists of words.

Echoing the words of Zamel (1984), Granger and Petch-Tyson point to the importance of teaching students when *not* to use the connectors as much as when and where to use them.

Unsurprisingly, perhaps, given the large numbers of Chinese learners studying at universities in Anglophone countries, there has been a pronounced emphasis in much of the EAP research conducted so far on Chinese learners of English. In a study, which examined the tendency of Chinese learners of English to experience difficulty when expressing degrees of certainty, Hyland (1997) using a one million word corpus of Cantonese High School students' English examination scripts, found after analysis of a reference corpus of L1 speakers of similar age a propensity to use a more restricted range of features among L2 users. There was also a noted tendency to be more assertive with stronger commitments. Finally, Hyland draws attention to the means through which writers signal a stance on particular issues. Hyland examined several components of this aspect of language use and listed "categories of epistemic commitment" (Hyland, 1997, p.19) such as modality (e.g. will, could) approximation devices (e.g. almost, to a certain extent) personalised and impersonalised forms (I believe, apparently) and epistemic clusters (it might be possible, it would seem that) and concludes that the Chinese speaking learners used simpler constructions, relying on a more restricted range of devices and tending to offer stronger statements, finding difficulty in expressing an exact degree of certainty (Hyland, 1997, p.20). Hyland calls for greater awareness of these features to be reflected in course design in English medium universities. Another study which focused on Chinese learners, in this case using a large sample of IELTS free writing scripts of *20 to 40 minutes duration*, Kennedy and Thorp (2007) identified a strong tendency towards the use of categorical statements and less hedging than typical of L1 users. Chen and Baker (2010) also emphasise this tendency of Chinese learners to make over-categorical statements and also to use a smaller range of lexical bundles (a generic term which refers to a set of co-occurring lexical

items, as discussed earlier in this chapter). In a comparative corpus study of published academic L1 texts, L1 learner texts and L2 learner texts, Chen and Baker identify L2 writers as using fewer lexical bundles and as having a tendency to underuse certain high-frequency expressions such as “in the context of” and also to use other expressions uncharacteristic of academic language such as “all over the world”. Chen and Baker also emphasise the pedagogical implications of their study, calling for integration of frequency driven expressions into ESL/EFL/EAP curricula.

Chuang and Nesi (2006) in a corpus-based study of errors produced by Chinese undergraduates at a British university, identify a series of error types, with the four most frequent being missing definite article, bare noun count for plural, a redundant definite article and misselection of prepositions. These types accord fairly closely to what a marker of L2 essays might commonly anticipate with recurrent errors in articles, prepositions, plurals/singles and word classes. Chuang and Nesi point out that the three top error types are all related to the article system and that work on this language feature would be a good use of time, with a corpus being viewed as both a teacher’s friend and a students’ study companion.

Thus far, learner corpus analysis has been most frequently used to emphasise the differences between first and second language learners in relation to a wide range of features, from general syntactic development to rhetorical styles, as we have seen above. But more recently researchers have begun to look at how L2 student writing develops over time. Corpus-based longitudinal research can be carried out either by collecting and analysing work by the same cohort of students over an extended time period, or ‘quasi-longitudinally’, by looking at work produced by comparable (but not the same) students at different proficiency levels. An example of the former will be reviewed later in this chapter; for the moment, let us conclude this section by reviewing a recent example of the latter. Parkinson

and Musgrave (2014) studied nominalisation and phrasal compression in essays by two groups of graduate writers of English as an L2; one group consisted of students on a preessional EAP programme (the ‘lower’ level group), and the other consisted of students who were already studying on their chosen MA programmes (the ‘higher’ level group). Parkinson and Musgrave found that there was a greater use of sophisticated forms of noun modification among the more proficient student writers, and a prevalence of more attributive adjectives by the less experienced (and thus less proficient) students on the EAP programme. Parkinson and Musgrave argue that these observations support the developmental path for noun phrase construction proposed by Biber *et al.* (2011). Figure 2.7 provides examples of the type of language which a student may produce along this developmental path.

Figure 2.7 Noun phrase modification features (from Parkinson and Musgrave, 2014, p. 50)

Stage (Biber <i>et al.</i> , 2011)	Grammatical structures	Example from (Parkinson and Musgrave’s) dataset
2	Simple phrasal embedding in the noun phrase: common attributive adjectives less common attributive adjectives	big earthquake potential disaster
3	Relative clauses simple phrasal embedding the noun phrase: nouns as premodifiers Possessive nouns as premodifiers Of phrases	(an) earthquake which happened in Japan power stations/ bomb blast people’s views risk of this technology

	Prepositions other than of	war on the Korean peninsula
4	Nonfinite relative clauses: -ed ing clauses Of phrases (abstract meanings)	risk involved with terrorism people living around the place the production of fossil fuels
5	Preposition + non-finite complement clause Complement clauses controlled by nouns	price of keeping the acceptable security standard viewpoint that using nuclear energy is equivalent to suicide

In considering the evidence above, Parkinson and Musgrave advocate an explicit teaching focus on noun modification; specifically, “a focus on nouns as premodifiers and prepositional phrases as postmodifiers (Parkinson and Musgrave, 2014, p.58).

To summarize the discussion so far, we have seen that corpus-based approaches to the study of second language writing are now becoming increasingly popular, and are now being used to study a wide range of linguistic features. We have also seen that this research allows detailed and robust comparisons to be made between L1 and L2 datasets, and by L2 writers at different proficiency levels. However, all the studies reviewed above are similar to each other in one crucial respect: they all focus on a single language feature or a relatively restricted and well defined group of features, such as nominal groups, prefabricated phrases, stance expressions, types of error, and so on. This sharp focus on relatively narrow aspects of L2 writing is of course perfectly legitimate and entirely welcome, but the advantages of computer-assisted analysis mean that it is now also possible for researchers to study many linguistic features *at the same time*, and to submit the results of such multiple analyses to sophisticated statistical analysis. Thus far, two significant approaches that harness the full

power of computational analysis have emerged: one is associated with the Coh-Metrix project (McNamara *et al.*, 2014), and the other is Multidimensional Analysis (Biber, 1988). Each of these approaches will be reviewed in turn.

2.4.4 Coh-Metrix analyses of L2 writing

Coh-Metrix (McNamara *et al.*, 2005; McNamara and Graesser, 2011; McNamara *et al.*, 2014) is a software programme which its creators describe as “a sandbox of automated language and discourse facilities” (McNamara *et al.*, 2014, p.40). It draws together a range of techniques and resources that have been developed within the field of Natural Language Processing; prominent among these are *lexicons/dictionaries*, for example Wordnet (Fellbaum, 1998) The MRC Psycholinguistic Database (Coltheart, 1981) and the CELEX (Baayen, Piepenbrock and Gullikers, 1995); *programmes* which use text inputs and calculate language/code outputs, for example a syntactic parser (Charniak, 2000); *algorithms* which can measure language components, for example as used in latent semantic analysis, “a statistical representation of world knowledge” derived from corpus data (McNamara *et al.*, 2014, p.42). Coh-Metrix also draws on general theories and research into language and discourse that have been conducted in linguistics, psychology and other related fields in recent decades.

The programme itself is accessible via a simple web interface, and employs a series of databases that provide a wide range of statistically referenced linguistic information on up to 108 different categories, ranging from simple word-based information (for example mean word length), through to more complex measures including latent semantic analysis (for example conceptual similarity between sentences). Coh-Metrix was originally developed as a tool for automatically assessing text readability (i.e. how easy or difficult a text is to read), but has in recent years begun to be used as a tool for studying second

language writing. In the review of Coh-Metrix-based studies that follows, we will begin by looking at general uses of the software in order to establish the range of features that it can analyse, before moving on to review in more detail a study that uses Coh-Metrix to address a very similar research question to that which is being posed in this thesis.

In an analysis of (textbook) cohesion (McNamara *et al.*, 2010) three indices of cohesion – argument overlap, latent semantic analysis (average similarity between each sentence) and connectives frequency – were used to assess the levels of difficulty for readers. The main finding of this study was that ‘traditional’ features such as text and word length (McCarthy *et al.*, 2007) were found to be too “shallow” and that Coh-Metrix was able to provide deeper metrics to analyse potential comprehensibility which are not normally included in traditional readability formulae. In another study, Crossley, Allen and McNamara (2011) developed a model of lexical proficiency in learner texts which is intended as a surer measure of *readability* than other previous measures. This study emphasised the work of McNamara *et al.* (2010) who were concerned to analyse psychology texts at multiple levels of text cohesion (high cohesion = easy to read/ low cohesion=difficult to read) The study concluded that some commonly used readability indexes were inaccurate in distinguishing between high and low cohesion texts and that the Coh-Metrix offered a more sensitive method of discrimination by being able to compute a higher range of discriminatory factors.

In another application of Coh-Metrix, Hall *et al.* (2007) looked at differences between British and American English as reflected in a corpus of legal texts and examined the differences between text types generally thought to be highly similar, namely varieties of legal English. The authors provided strong evidence of differences in language between the two corpora, and concluded that “[t]he algorithm generated in this analysis requires no human intervention at a judgement level and establishes that discourse level features are

sufficiently diverse for sophisticated computational systems to distinguish texts with a very high degree of accuracy” (Hall *et al.*, 2007, p.51).

A salient issue to emerge from the Hall *et al.* (2007) study is the authors’ distinction between *traditional and sophisticated linguistic indices*. The traditional measures include average word length and average sentence length in addition to syllable count and number of words per sentence/paragraph/text which the authors consider can be used in tandem with the more elaborate measures to produce an effective and discriminatory research tool. McCarthy *et al.* (2006) looked at authorship and attempted to disambiguate between the writers Kipling, Wodehouse and Dickens with these authors being chosen because their styles were diverse but sufficiently similar that distinguishing between them would not be easy. After selection of an appropriate range of indicators after discriminant analysis including higher level constituents per word, minimum word imageability per paragraph, incidence of wh determiners and incidence of conditionals, the study concludes that the authorship is discernible by the Coh-Metrix method and that authorship styles which can vary over the period of a career are also distinguishable by the same method – for example, Dickens and Kipling used more pronouns as their careers progressed.

In another study, using a corpus of essays produced by 9th grade and 11th grade US students, Coh-Metrix was used to investigate whether essays produced at different grade levels could be distinguished from one another. This study, (Crossley *et al.*, 2011), examined linguistic features such as word frequency and word concreteness relating to lexical sophistication. It also examined the issue of syntactic complexity, by examining, for example, the number of modifiers per noun phrase and used word overlap and the incidence of connectives to investigate cohesion. The study suggested that the students produced more sophisticated lexical items and more complex sentence structure but fewer cohesive features, as grade level increased and that the writing could be characterised as being of a more

elaborate but less cohesive nature.

The previous four examples of Coh-Metrix-based studies demonstrate the applicability of this type of methodology to a variety of contexts, and suggest that such a programme can be effective as a research tool in the field of second language writing development. A study with more direct relevance to the present research is that of Crossley, Kyle and McNamara (2016), who used Coh-Metrix to investigate features of textual cohesion in a corpus of pre- and post-course timed essays written by 57 university students on a one-semester EAP preessional programme at a university in the United States. Crossley, Kyle and McNamara found that there was increase in several cohesive devices, the largest being that of noun overlap, the repetition of the noun in exactly the same form and plurality, between paragraphs. Their study is thus extremely significant for the present thesis, in that it suggests that there may be grounds for greater optimism than hitherto as to the possibility that there are certain text features which may show an improvement over a relatively short period of instruction. The study is also highly relevant to the current research in that Crossley, Kyle and McNamara stress the implications that their observations have for both testing and teaching. In particular, they argue that essay scoring systems could be adapted to include cohesive features indicative of essay writing quality and that teachers could be usefully informed about semester long improvement profiles and “the possible trajectories of ... students and [to allow] them to better pinpoint instruction and to target specific areas of cohesion development” (Crossley, Kyle and McNamara, 2016, p. 14).

2.4.5 Multidimensional analysis of L2 writing

While Coh-Metrix allows the corpus-based L2 writing researcher to carry out an unprecedentedly wide range of linguistic analyses fully automatically, it remains

somewhat traditional in approach inasmuch as each of the variables it analyses is treated separately. That is, it provides a discrete set of results for each feature, and does not perform any multivariate statistical analysis on the data generated as a whole. In this section, we will review a method which does harnesses computer technology in order to identify and analyse the covariance of multiple linguistic features at the same time:

Multidimensional Analysis (MDA). First developed by Douglas Biber (1988), MDA looks at simultaneous relative presence or absence of large numbers of discrete linguistic features (usually over 100) and uses factor analysis, a multivariate statistical technique, to reduce these large numbers of individual co-variance observations to a small set of basic characteristics, which Biber calls 'dimensions'.

Like Coh-Metrix, MDA was not originally intended as a methodology for studying L2 writing development; in fact, Biber (1988) originally developed it as a means of studying variability across spoken and written varieties of English. Nevertheless, it has subsequently gone on to be used to investigate a wide variety of research questions, including those salient to this thesis. One such study is Grant and Ginther, (2000). This study examined 90 L2 TWE (Test of Written English) essays selected from a 1700 essay data base, which were completed under timed conditions (30 minutes) and completed during standardised testing conditions. The essays were in response to a single prompt and were tagged for linguistic features using Biber's (1988) multidimensional categories. The tagged linguistic features included essay length, lexical specificity (type token ration/word length), lexical features (conjuncts, hedges, amplifiers, emphatics, demonstratives, downtoners), grammatical features (nouns, nominalisations, personal pronouns, verbs, modals, adjectives, adverbs, prepositions, articles), and clause level features (overall subordination, complementation, relative clauses, adverbial subordination and passives). After analysis, the L2 writers, who were of three broad levels of proficiency, were shown to produce characteristic features

marking a profile of what could be termed written sophistication. These were identified as an increase in text length and lexical specificity (type/token ratio and word length) and on a lexical level, increases in all the features listed with the exception of hedges. In terms of grammar, the general use of different parts of speech such as nouns, verbs, adjectives, adverbs, prepositions and articles, increased. This is very much in line with what could be called an intuitive description of a profile of increasing writing competence, with these items being those which teachers may “mark” as praiseworthy or in need of attention. Other grammatical features are also instructive; for example, nominalisation, a characteristic feature of academic writing (Shaw and Liu, 1998), is used more frequently by the better writers and that as writing improves, more awareness is shown of register differences and the need for sensitivity to genre. Pronoun use showed an increase *in the first person* an increase attributed to the “personal opinion style of the question” (Chapters three, four and five of this thesis offer strong evidence regarding this feature). Third person pronouns also increased. Regarding verbs, the types or range seem to increase with competence, in other words, more competent writers seem better able to select from a range of private, public and suasive verbs as appropriate. This, in the view of Grant and Ginther provides support for the idea of writing becoming *more like written than spoken language* (see Figure 2.7). Modals also showed a steady increase as did adjectives, adverbs and prepositions with a considerable rise, a tripling, in adverb use at the highest levels. Articles showed an increase in appropriacy of use. At clause level, competent L2 users showed more subordination and use of passives.

Grant and Ginther do caution that although computer tagged features can be illustrative of writing development, there is a need to be aware of some limitations. Firstly, conclusions based on timed, unstructured essays may not be necessarily extended to longer essays written in different conditions (and vice-versa). Secondly, the nature of a task and its influence on the type of language produced, means that caution again needs to be applied. The writers

also call for supplementary analysis, of a qualitative kind especially with low level writing.

Figure 2.8 Tagged linguistic features of TWE essays (Grant and Ginther, 2000, p.130)

features	description/examples
essay length	total number of words
lexical specificity	indication of how precisely the writer used vocabulary
(a)type token ratio	number of different words per first 50 words of text
(b)word length	mean length of words
lexical features	for example:
conjuncts	however
hedges	sort of, kind of
amplifiers	completely, absolutely
emphatics	such, really, so
demonstratives	this, that, these, those
downtoners	almost, barely, hardly
grammatical features	
nouns	
nominalisations	words ending in -ment, -ness, -ity etc.
pronouns	first, second, third personal pronouns
verbs	tense, aspect (e.g. present, past, perfect); verb types (private, public, suasive)
modals	possibility (can, may, might, could); necessity (ought, should, must); predictive (will, would, shall)
adjectives	attributive only
adverbs	e.g time and place adverbs
prepositions	
articles	definite and indefinite
clause level features	

overall subordination	complements, relative clauses; adverbial subordination
complementation	that comp; infinitive comp
relative clauses	subj, obj, prep relatives
adverbial subordination	“because she was late”
passives	by and agentless passives

Grant and Ginther’s work offers a glimpse of the broad feature range available after the application of a multidimensional analysis. The next study to be described here also adopts a multidimensional approach to the analysis of a range of linguistic features, although the data do not derive from English. Ascension-Delaney and Collentine, (2011) in a 202,000 word corpus study of written L2 (undergraduate) Spanish learners (the subjects were L1 English speakers) and following on from the multidimensional study of native speaker Spanish by Biber *et al.* (2006), focused upon the variety of lexical and grammatical phenomena used to communicate in writing. They observed feature clustering which were considered characteristic of the narrative and expository discourse types. This study which was one of the first to apply a multidimensional analysis to a written Spanish learner corpus, focuses upon second and third year university students and although the study is primarily directed to discourse variation, it contextualizes the issue of L2 linguistic complexity in the development of a learner interlanguage. In studying 78 *lexical and grammatical features* including for example adjectives, verb classes and phrases and features such as dependent clauses and noun phrase configurations, the writers were able, after norming the observed frequencies to submit the findings to a range of statistical techniques, including principal factor analysis. The study concluded that the linguistic complexity was encoded in a variety of dimensions and that while syntactic complexity remained fairly static, there was variation in the degree of informational density by frequent use of nominal items.

In another application of multidimensional methodology, Reid (1992) analysed

768 TWE (TOFEL) scripts and obtained results that emphasised the importance of academic rhetoric and the need for learners to focus upon the linguistic conventions embedded in different writing genres. The main focus of Reid's study was to examine the configurations of academic prose as used by three L2 groups, (Spanish, Arabic and Chinese) and one L1 group from North America. Using two topic types for the writing tasks (comparison/contrast and graph description) it was found that the writers from all language groups used significantly more pronouns in comparison type essays and more prepositions in the graph description ones. In addition, all four groups used more coordinate conjunctions and subordinate openers in the contrast essays. The writers advocate the transmission of sociocultural knowledge and suggest continued development of written corpora for use by teachers and course designers. Reid considers that "The use of a multidimensional methodology that included computer text analysis to assess a broad range of linguistic and rhetorical features can assist ... in adequately accounting for variation in the co-occurrence of certain discourse features among ESL writers in English" (Reid, 1992, p101).

In another application of multidimensional analysis to a large (British Academic Written English) corpus of student academic essays and assignments, Nesi and Gardner (2012) use Biber's (1988) dimensions to investigate the development of student writing at undergraduate and postgraduate levels. The authors conclude that the students produced writing as they advanced through their first, second, third year and postgraduate level degree programmes which progressed towards being generally "highly informational, impersonal, non-narrative elaborated and lacking overt features of persuasion" (Nesi and Gardner, 2012, p. 260). This study, which uses a very large database of both L1 and L2 speakers, offers a dimensional perspective upon student writing development which has applicability to the present thesis.

While the studies reviewed above are quite diverse, one characteristic that they share

is that they are all what we might call full implementations of MDA. Before moving on, it is worth noting that there have also been studies that do not implement a full MDA themselves, but use the outputs of MDA research (notably Biber's dimensions) as a basis for a more traditional qualitative form of analysis. An early and notable example of this approach is the seminal paper by Shaw and Liu (1998), which was discussed at length in Section 2.3 above. Shaw and Liu applied Biber's (1988) dimensions (involved/non-involved, narrative/non narrative, explicit/ situation dependent, overt expression of persuasion/ abstract/non-abstract and information elaboration) to their pre and post study of University students over a short time frame, and found increases in impersonality (reduction of personal pronouns/increase in passives), formality (more formal vocabulary), explicitness (more lexical cohesion, more WH relatives), hedging (more metadiscourse) and complexity of syntax (more subordinate clauses and more nominalisation) and modification. On the involved/non-involved dimension, two of the variables, contraction and subordinate clauses, showed movement towards an academic axis, and an increased use of passives and connectors indicated a movement towards more abstract exposition.

Another important research paper that draws on aspects of MDA rather than implementing it in full is Staples *et al.* (2016). Using a cross section of the British Academic Written English corpus (BAWE) and selecting university level L1 writers, Staples *et al.* provide evidence that phrasal complexity increases in line with levels of academic sophistication, while clausal complexity has an inverse relationship with academic level; that is, as academic level increases, clausal complexity decreases. This study principally addresses the issue of writing complexity, and emphasises the importance of the production of long noun phrases. The authors conclude that developmental trends in relation to this lexicogrammatical feature are discernible among L1 writers. In considering the idea of complexity, Staples *et al.* suggest that "phrasal complexity is increasingly important as

writers develop throughout their university education” and that successful “student writers use more compressed phrasal structures” (Staples *et al.*, 2016, p.31).

The Staples *et al.* study is not strictly a MD analysis in that it does not use factor analysis to identify a set of dimensions of variability, but it is similar in spirit in that it synthesizes a number of different linguistic and contextual variables. Linguistically, it cross-classifies phrasal and clausal complexity, and contextually it considers level of students, discipline and genre, in order to generate a highly detailed and sophisticated account of student writing development.

2.4.6 The need for further research

Whilst the Staples *et al.* (2016) study is methodologically highly relevant to the current thesis, its focus on L1 learner writing mean that they do not directly address the central question of this thesis, which concerns whether measurable improvements in L2 EAP student writing can be observed over a relatively short time frame. The Crossley, Kyle and McNamara (2016) study reviewed at the end of Section 2.4.4 does address this question directly, and provides robust evidence in support of the claim that such improvements can indeed be identified, but is seriously limited in that it focuses on only one small cohort of 57 students over just one course of study. What is needed now is a study that compares larger cohorts over multiple course durations and over more than one cohort year. It would also be preferable for the purposes of triangulation to carry out more than one kind of analysis on such a corpus, for example by running and then comparing Coh-Metrix and MD analyses on the same set of data, in order to see whether these analyses provide the same or at least complementary perspectives on student writing development. This is precisely what the current thesis proposes to do.

Before closing the chapter, there is one remaining question one to be considered, namely

the issues of controlling for L1 as a variable and the evidence is mixed. Flowerdew (2010), for example, has pointed out that internal variation within a given corpus may significantly alter an identified trend, and suggests that the influence of first language may have a particular bearing upon the development of syntactic complexity. This point is elaborated by Lu and Ai (2015, p.20) who consider that “learners with different L1 backgrounds even those at the same or comparable proficiency levels may not develop in the same ways in all areas”. Lu and Ai’s study, which used ICLE version 2 (Granger *et al.*, 2009), examined the use of syntactic complexity measures across language groups and found the influence of the first language to be very significant. Such observations have led Ortega (2015) to call for more research in order to emphasise the influence of the L1 on syntactic development and to “refrain from purely developmental or purely proficiency-based explanations of syntactically less or more complex patterns when the L1 influence has not been accounted for in the data” (Ortega, 2015, p. 85).

Other researchers, however, have argued that the issue of controlling for L1 makes no difference. In a study which investigates the shared features of L2 writing-characteristics *which are not contingent on L1 category*, Crossley and McNamara (2011a) measure syntactic complexity using Coh-Metrix (McNamara *et al.*, 2005), a computer programme that carries out a battery of statistical tests on written language data (See 2.4.4, and, for a full discussion of the construction and characteristics of this programme, see Chapter Four). Specifically, Crossley and McNamara examine the average number of words before the main verb, the number of sentence and embedded sentence components per word and the phrasal and syntactic categories which comprise the syntactic constructions. The study also employs measures of lexical sophistication and text cohesion. Significantly also for the current thesis, which draws upon data from a variety of language groups, the authors conclude that L2 learner writing is characterised by features such as hypernymy and

polysemy (conceptual meanings and levels), stem overlap (sentences sharing one or more word stems) and lexical diversity (number of diverse features), irrespective of L1 background. Similar findings were also suggested by Cumming and Riazi (2000) who, in a study of 108 ESOL (diverse language groups but students from Hong Kong, Taiwan and Japan prominent) learners doing *a two term* intensive programme at a Canadian university, pointed to the uniformity of writing improvements across mother tongues and other factors such background and occupation. Although the current thesis does not adopt a particular position with reference to the theoretical debate concerning L1 influence on L2, it is reasonable to regard these findings as allowing the present thesis to disregard first language background as a factor, and thus to exclude this variable from the research to be reported later in this thesis. The current thesis does not take a position on this particular debate as it is not relevant to research questions posed which focus upon whether L2 EAP writers improve in general. L1 differences in improvement trajectories are only important if they skew the overall results so badly as to make the study meaningless, and as a result there is no consideration of influence of L1 in this study. However, the whole subject of L1 influence upon second language writing development may well be a fruitful subject for further research and I discuss the issue in the final chapter.

2.5 Conclusion

This chapter began by suggesting that writing development could be seen as progressing along three strands, inaccurate to accurate, simple to complex and oral to written. The previous literature on the subject was then reviewed from a variety of perspectives and the consensus position was generally one of pessimism, or at best cautious optimism regarding the possibility of explicit instruction bringing about measurable development in writing

competence along any of these three strands within the short time frame characteristic of preessional EAP programmes. However, recent studies which have used corpus-based methods of analysis have offered a more optimistic scenario, suggesting that for certain features at least, a development profile may be discernible after a relatively short period. In the remainder of this thesis, I aim to investigate empirically whether and to what extent these positions are supported by evidence and to examine the implications for teaching and course design.

CHAPTER THREE

THE EMPIRICAL CONTEXTS OF THE RESEARCH

3.1 Introduction

This chapter begins by describing the multidisciplinary preessional EAP programme at the University of Birmingham, from which the student essays and other empirical data used in this thesis were drawn. I then describe a pilot study that was conducted in order to establish whether a large-scale corpus-based project of the kind envisaged by this thesis would be methodologically viable in itself, and whether it might be expected to yield findings that would justify the time and effort involved in collecting much larger amounts of learner data over more than one academic year. This pilot study involved collecting an initial corpus of EAP student essays in order to facilitate investigation of the issue of linguistic development over the period of the Birmingham EAP courses. This preliminary investigation was carried out using Wordsmith Tools 5 (Scott, 2008). The chapter shows the results of the Wordsmith analysis, offering consideration of these findings in relation to other corpus studies. The chapter then revisits the conceptual framework and proposes that the results so far obtained suggest that further, larger scale investigation into the features so far identified would be a fruitful venture. The chapter concludes with an account of the creation of the EAP Corpus (EAPCORP), the larger and more extensive corpus of Birmingham EAP student writing that will be the focus of the main research to be reported in the rest of this thesis.

3.2 The nature of the programmes

Like many other universities in the UK and elsewhere, the University of Birmingham offers a suite of preessional English for Academic Purposes (EAP) programmes for non-native English speaking students who intend to study at postgraduate level. Following successful completion of the EAP preessional programme, these students subsequently move onto a variety of postgraduate degree courses including such disparate subjects as Economics, (and related subjects such as Accounting and Finance, Money, Banking and Finance and Financial Engineering), Engineering (Mechanical, Chemical and Civil), Biosciences (including Toxicology and Immunology) TEFL/Applied Linguistics, International Development, Law and Social Policy. These subjects are sufficiently represented in terms of student numbers to require at least one class (maximum number 14 per group) per discipline, with Economics requiring 4 or 5 classes. There are subject areas with fewer students, for example, Music, English Literature and Mathematics. The EAP programmes thus serve the entire university and much of its subject and curricular range. As indicated, some departments and courses are more fully represented than others, but there is generally a fairly even Science/Arts split. Students are from a variety of first language backgrounds, with Mandarin Chinese the largest group, representing around 60% of the total cohort. The second biggest group consists of Arabic speakers and there are many other languages represented including Japanese, Korean, Thai, Greek, Farsi, Russian and Spanish. There is a considerable age range, with the youngest being around 22 and the oldest up to 70 with a median age of around 27. The students are, as a consequence, mature and thus beyond the age where broader issues relating to developmental psychology come into play. The programmes vary in length with 20, 15, 10 and 6 week courses available, but all programmes have the same end date, which is usually 2 weeks before the autumn term (the first term of

the UK academic year) begins. Admission onto the programmes depends upon two variables: the language admission level of the target department and the level of English as assessed by external qualification- for example IELTS and TOEFL. The departments decide which level of English is suitable as a minimum score for prospective students. If they do not meet departmental criteria, students are advised either to reapply once the English qualifications have been obtained, or participate in the University of Birmingham Pre-sessional multidisciplinary EAP Programme. Decisions regarding the appropriate length of the course required are taken by the programme organisers and the students advised accordingly. For example, a potential applicant for an Engineering MSc would need to obtain a 6 overall on IELTS, and a 20 week programme may be recommended if their current score is only 5 or equivalent. A prospective law student would need a 7 overall on IELTS so would also be recommended a 20 or 15 week course. The score profile listing the marks which the students are required to obtain is as follows:

40% = IELTS 6 (minimum level for Engineering/ Computer Science)

50% = IELTS 6.5 (minimum level for Economics/ Political Science)

60% = IELTS 7 (minimum level for Law/ Medicine)

Teachers are recruited for 20, 15, 10 and 6 week programmes which all terminate on the same date. Teachers are expected to be experienced in TEFL and preferably EAP and ideally to hold an MA in TEFL or related subject together with an RSA diploma in TEFL. The EAP programme receives a regular four yearly accreditation from the British Council who insist that teachers are of TEFL (Q) status. The British Council allows the University to recruit teachers of TEFL (I) status (a lower level TEFL qualification), providing they can be shown to be on a developmental path towards achieving full TEFL (Q) status. Most of the teachers

who work on the University of Birmingham preessional EAP programme are of fully qualified TEFL (Q) status and those who are not, are expected to upgrade their qualifications.

Assessment consists of a 3,000 word academic research paper (ARP) which attracts 80% of the marks. This paper is also presented orally, and the oral presentation receives 20% of the total marks available. Students are evaluated solely upon their English; this is a crucial characteristic of the programme, which, although serving a wide variety of subject disciplines, is not content based. The various disciplines themselves are not taught although students are familiarised with the relevant subject-based vocabulary in the scheduled subject related sessions where postgraduate researchers are employed as a resource and assistant to the teacher. The issue of content based English teaching is a separate one (Davies, 2003; British Council, 2014) and there is no intention to debate it here, but it is worth emphasising at this point that the Birmingham EAP programme adopts a "bottom" up approach (Ellis, 2008; O' Malley, 1990) where one of the key aims of the programme is to teach the students explicitly about the construction of written academic texts. In other words, the focus of instruction is on the language features of written and spoken academic texts rather than on the academic content conveyed by this language.

The EAP preessional course itself has at its core, a 300 page book (Oakey and Treece, 2008) which is structured around the identified characteristics of written academic English considered important for the students to know. These include such grammatical features as nominalisation, the noun group, relative clauses and the passive, plus the wider issues of text structure, information sequencing, distancing, hedging and the importance of overall objectivity in the UK university context. The main aim is to help the students acquire an authentic academic voice without the need to resort to copying whole chunks of text, or cutting, pasting and plagiarising in general, an issue which appears to be a growing problem

in the UK higher education sector. Students receive twenty hours per week of direct instruction with a full time teacher. The morning classes are oriented towards the production of academic writing and the afternoon classes include two classes entitled “English in Your Subject” where teachers receive the assistance of a postgraduate researcher in the specific discipline area.

3.3 Creating the pilot corpus

Creating a valid learner corpus of any size or complexity is a demanding and time-consuming enterprise. As such, it was decided to create a pilot corpus of student essays as a means of establishing (a) whether a larger corpus building project would be viable within the timeframe available to me as a postgraduate researcher, and (b) whether the findings obtained from a preliminary analysis of the pilot corpus would justify the time and effort involved in compiling a larger and more comprehensive corpus for the main research.

Clearly the scripts for the pilot had to be free writing in nature, with a need for a sample of the students' "productive linguistic resource" (Shaw and Liu, 1998, p.248) and this was better performed by students under controlled conditions, namely in a large lecture theatre on the first and last days of their programmes. The students could have been asked to write in-class essays but there was too much possibility of copying, using dictionaries and a likelihood of exceeding or going under the time limits, with this also being true of a possible home essay option where students are asked to complete the task outside of class time. Students were told on day one that there would be a writing sample gathered and that this was for baselining purposes, to enable their teachers to gain a clearer picture of the quality of their writing and to map out their progress. All students were told that the samples

were also to be used for academic research purposes and asked to give their consent for this anonymous data gathering exercise. Students who did not give permission were not included in the study.

Students were asked to write for 30 minutes on the following essay prompt:

What are the differences between your country and the UK?

This title was chosen because of its simplicity, relevance (especially as most students had only disembarked from a plane two days previously) and the fact that it offers considerable opportunities for explanation, comparison and exposition from a range of perspectives. As a relatively simple task, it offered the writers opportunity to express themselves with a low cognitive burden in terms of topic complexity, thereby (in principle at least) allowing students to expend more cognitive effort on the quality of their writing (Bereiter and Scardamalia, 1987). On the last day of the programme, the same writing task was repeated, under exactly the same conditions. In other words, students received the same essay question at the beginning and the end of their studies but were not told in advance, either that they would be asked to produce a writing sample or that the question would be the same. The aim was to standardise the prompt question without creating too much boredom, and to prevent students from preparing an answer in advance, memorizing it and simply reproducing it under timed conditions. In requiring the students to make an explicit comparison between aspects of their and their host's country, the question contains a cultural element. This is deliberate. Firstly, because the student learning experience at the University of Birmingham involves a cultural aspect. Students are living and studying in a new culture and the learning aspect of their programmes whilst clearly of paramount importance takes place in a cultural

context. Students are also clearly aware of the differences between the UK and their home country and expect and receive a cultural input into their programmes. This takes the form of visits to places of interest across the UK, social events, parties and visits to the wider non-university environment. The selection of this question thus serves the purpose of recognising the importance of the student's own cultural experiences over the duration of the EAP programmes. It also serves to open the student learning experience at the very beginning of the programme and to conclude the first section of the student learning experiences at the University of Birmingham.

The issue of task familiarity arises at this point, that is, the question of whether the students might be expected to produce a better writing sample on the second occasion because they have been previously exposed to the question. This 'familiarity effect' was anticipated but was not expected to significantly affect the quality of the students' output for several reasons. Firstly, the reduction in the cognitive load produced by having to write on a topic that is not only simple but also already very familiar and salient to student writers who would have only just arrived in the country as discussed above, could have enabled students to express themselves more or less as fluently in the pre-test as in the post-test. Secondly, whilst a student might be expected to have a greater range of lexico-grammatical options available after a period of study, this does not necessarily lead to a more complex, accurate or coherent writing sample even if the topic has been addressed earlier in the programme. Also, the essays were produced after a complete absence of preparation with the task requiring spontaneous writing on a topic which was unlikely to have been anticipated.

The corpus was then compiled and initially produced with 22 matching pairs for the 15 week programme 2009. This was then augmented by 35 scripts for the 15 week

programme 2010, 86 scripts for the 10 week programme 2010 and 40 scripts for the 6 week programme 2010. The essays (all of them hand written) were transcribed into plain text computer files. Nothing was corrected; all errors were reproduced verbatim, thereby not only maintaining the integrity of the original data but also avoiding the need to make (often arbitrary) decisions about which part of speech or linguistic category a corrected word should be reformulated into. Secondly, it removed the possible inconsistencies which may have resulted from inaccurate categorisation of linguistic features. This approach did entail the disadvantage that the programme was often unable to recognise a word produced in error. For example, if the word *country* is spelt *coutry* then Wordsmith software classified it as a different word and created a separate frequency count for this item. It was also important to record the misspellings accurately and to pay close attention to each script for spelling errors.

3.4 Identifying linguistic features for pilot analysis

The next step was a consideration of how the data could be used to offer points of comparison between work produced at the beginnings and ends of the programmes. As indicated in the introductory chapter, the features of interest at this stage included mean sentence length, error free units and lexical densities. It was also considered to be of interest to investigate those words whose frequencies had increased or decreased after the instruction period. Clearly, it would be neither appropriate nor even feasible to compare every word in the data, so it was decided to focus just on any observable frequency differences in the top 20 words in word lists obtained for both pre and post-course data sets. A detailed rationale for choosing each of these features will now be provided.

From the limited data then available, sentence length was selected because it offers an easily quantifiable measure and could provide a ‘quick and easy’ statistical basis for

making judgements about the broader potential of the project as discussed earlier. As a measure of the development of second language writing capacity, it was felt important to ascertain how much the students were able to produce in the time available, on the assumption that the capacity to write sentences of increasing length could be said to be congruent with writing development. In other words, it was an assumption of the pilot study that longer sentences may be characteristic of maturing, well taught writers. Clearly, this assumption is not always or necessarily the case, as succinct writing (consisting of shorter, less rambling sentences) may often be more highly valued by academic readers. Nevertheless, it did seem reasonable to expect overall, that as students gain knowledge, confidence and information about the nature of the writing process, they may well produce longer sentences. It was also felt relevant at this early stage in the study to discover whether students could write more accurately at the ends of their respective programmes, and how much writing could be produced without error. For this reason, the error free unit count was included in the list of features for analysis. An error free unit is defined here as a segment of writing, which most closely corresponds to the sentence (subject verb object usually between full stops) containing no morpho-syntactic mistakes, this being a fairly simple description of the concept of error. The sentence was selected as the unit largely for practical reasons, chief of which was the fact that it is generally a more easily operational construct, easier to identify in a stream of text and for the non-specialist to understand. If a course designer pronounces that, for example the number of errors per sentence has decreased, it probably has more resonance with students and the general public than the number of errors per clause. Having decided upon an all-inclusive classification, a manual approach was used to log errors, which in practice meant reading each sentence and identifying those with no errors of any kind, including for example spelling, punctuation, grammar or word choice. These error-free units were then recorded as a percentage of the total sentence count.

The complexity and variety of students' writing was another topic of interest and as a result, the lexical density (ratio of content to grammatical words) indicator was considered to require investigation. However, there are issues to consider at this point. For example, on a superficial level one might expect an increase in lexical densities to be indicative of writing improvement, but this may be predicated on a view that more content words mean better writing. Such a perspective may be too impressionistic, as the lexical density (LD) count is in some ways, a simple aggregation. For example, a student may write: *My country is very big. My country is developing country. My country is very old.* (LD 70% counting very as an adverb)

Or this: *My country is of considerable size, it has a developing infrastructure and a long tradition.* (LD 60% not counting ellipsis)

The second example would be classified at a higher level on most language assessment criteria, so it may well be appropriate not to use the LD count without qualification. There is also the question of whether and how to deal with errors within the LD framework. For example: *I like the education in this country as it can help student learn more and cultivate the **confident**.* In addition, modals and auxiliaries are difficult to place. For example: *I eat rice everyday / I may eat rice every day.* If the word *may* is counted as a grammatical word then the LD reduces, it may therefore be appropriate to regard LD as an indicator but to be sensitive to its efficacy in different contexts.

Finally, and as mentioned earlier, individual word frequencies within the top 20 words for each data set were also considered to be worthy of investigation and comparison, and were therefore added to the list of features to be studied in the pilot analysis. It was anticipated that many, if not most, of the words in these lists would be 'grammatical' words such as articles, prepositions and conjunctions rather than 'lexical' words such as nouns, verbs and adjectives. This was considered beneficial to the analysis, as it would allow me to

observe, for example, whether the article *the* was used more frequently after a programme of instruction as a possible indicator of more extensive use of noun phrases. Similarly, a greater use of the preposition *of* could be seen as indicative of greater nominalisation together with a range of feature movements which might suggest a greater approximation to a native English academic writer frequency profile.

As indicated earlier, the tool chosen to investigate these features was the Wordsmith Tools 5 suite of corpus analysis software. Wordsmith Tools allows the researcher to analyse textual data in a number of ways including listing word frequencies, showing collocations and obtaining key word lists. For the pilot study reported here, Wordsmith was used to determine which (if any) of the features described above could be said to have shown significant development. The initial data consisted of the paired 15 week scripts for 2 successive years, plus the 10 and 6 week scripts for 2010. At this stage of the data collection these were the only cohorts for which data had been gathered.

3.5 Results of pilot analysis

The first feature to be analysed is the word frequency count which is illustrated in Tables 3.1 to 3.4.

Table 3.1 Top 20 most frequent words for 15 week 2009 (22 matched scripts)

rank	pre	%	post	%	rank	pre	%	post	%
1	the	5.23	the	6.84	11	my	1.39	people	1.06
2	in	4.10	in	4.46	12	people	1.32	it	1.02
3	is	3.11	is	4.34	13	it	1.17	that	0.97
4	and	2.97	to	2.76	14	a	1.11	country	0.93

5	I	2.67	and	2.72	15	that	1.11	have	0.89
6	to	2.15	UK	2.07	16	but	1.02	I	0.79
7	UK	2.05	of	2.01	17	can	0.85	there	0.73
8	of	1.62	are	1.58	18	you	0.85	between	0.71
9	country	1.56	a	1.49	19	very	0.83	more	0.71
10	are	1.15	for	1.10	20	more	0.73	differences	0.66

Table 3.2 Top 20 most frequent words for 15 week 2010 (35 matched scripts)

rank	pre	%	post	%	rank	pre	%	post	%
1	the	6.34	the	7.57	11	my	1.17	I	1.06
2	in	3.63	in	3.89	12	there	1.10	people	0.97
3	is	3.56	is	3.16	13	it	1.02	it	0.92
4	and	3.40	to	2.97	14	people	1.02	different	1.08
5	to	2.47	and	2.86	15	country	1.01	four	0.82
6	I	2.43	of	2.18	16	have	0.97	between	0.78
7	UK	2.33	UK	2.13	17	but	0.90	have	0.78
8	of	1.65	are	1.41	18	we	0.87	they	0.70
9	are	1.49	China	1.41	19	different	0.80	more	0.69
10	a	1.36	a	1.30	20	China	0.77	that	0.67

Table 3.3 Top 20 most frequent words for 10 week 2010 (86 matched scripts)

rank	pre	%	post	%	rank	pre	%	post	%
1	the	6.51	the	6.43	11	people	1.14	people	1.08
2	in	4.08	in	3.97	12	China	1.06	it	0.94
3	is	3.04	and	3.24	13	it	1.02	China	0.88
4	and	2.94	is	2.53	14	that	0.93	that	0.88
5	to	2.51	to	2.47	15	for	0.82	as	0.83
6	I	2.43	of	2.39	16	my	0.82	have	0.82
7	UK	2.15	UK	1.84	17	there	0.79	different	0.73
8	of	2.08	are	1.58	18	country	0.79	for	0.72
9	a	1.35	a	1.44	19	have	0.75	they	0.72
10	are	1.35	I	1.14	20	but	0.74	with	0.64

Table 3.4 Top 20 most frequent words for 6 week 2010 (40 matched scripts)

rank	pre	%	post	%	rank	pre	%	post	%
1	the	6.6	the	7.54	11	people	1.26	China	1.16
2	in	4.4	in	4.41	12	China	1.19	it	1.04
3	and	3.2	and	3.16	13	my	1.08	people	0.94
4	is	2.92	is	2.70	14	it	0.97	have	0.94
5	I	2.69	to	2.49	15	country	0.85	different	0.91
6	to	2.56	of	2.03	16	for	0.83	that	0.81
7	of	1.89	UK	1.95	17	have	0.81	for	0.80
8	UK	1.78	a	1.68	18	that	0.79	there	0.76
9	are	1.46	I	1.53	19	different	0.73	as	0.72
10	a	1.45	are	1.51	20	there	0.73	between	0.71

The word frequency information conveyed in these tables highlights first of all the striking similarities between the four pre and post samples. In terms of frequency profiles, the four tables are similar, with the first two ranks identical across the two successive years. The profile in fact can be said to be fairly typical of first language writing (Nation and Waring, 1997) with high coverage items ranked at the top together with some of the vocabulary which would be expected to be elicited from the prompt question (for example *China* and *different*). There are a number of differences, however, regarding L2 learners which are worthy of comment. *The*, for example, generally regarded as the commonest word in written English with a typical text coverage of around 8% in native speaker writing, shows an increase from 5.82 to 6.84 for the 15 week programme 2009 and from 6.34 to 7.57 in 2010. Similarly, the percentages increased from 6.6 to 7.54 for the 6 week programme in 2010 (there was no significant difference for the 10 week programme 2010). These increases suggest that the students may be using more nouns, or are at least moving towards a more 'nativelike' use of the English article system, and that further investigation of this feature increase would be useful. Another striking feature is the consistent decrease in the use of the first person pronoun *I*, which from its place in the top 5-10 frequencies, falls considerably in both rank and percentage terms across the entire data set of the pilot study. For example, from rank 5 (2.67%) to rank 16 (0.79%) with the 15 week 2009. This simple and easily identified trend was considered to be a good reason for persistence with the current study and as the basis for wider, more detailed investigation of other linguistic areas. In particular, the reduction in use of *I* may relate to increased familiarity with the need to depersonalise, and to a greater use in the post-test data of the passive in academic writing in general. Further encouragement was provided by the increase in the frequencies of the preposition *of*; although not substantial, the figures for this item do suggest a greater use of nominalisation (*of* being the first element in the postmodifying phrase in very many complex

nominalizations in academic English). Again, this observation was interpreted as inviting and warranting an investigation on a larger and more ambitious scale.

The Wordsmith analysis produced several other findings relating to mean sentence length, lexical density and error-free units at sentence level. The main results can be summarised as follows:

Table 3.5 Mean sentence length, feature movement (*increase or decrease of mean sentence length*)

	pre	post	movement
15 week 2009	15.56	19.23	+3.97
15 week 2010	14.50	16.50	+2.0

Table 3.6 Lexical densities (percentage of lexical words nouns, verbs adjectives, adverbs) feature movement

	pre	post	movement
15 week 2009	41.68	46.63	+ 4.95
15 week 2010	49.13	55.17	+6.04

Table 3.7 Error-free units, feature movement (increase or decrease of error free units)

	pre	post	movement
15 week 2009	24%	33%	+9%
15 week 2010	23%	36%	+13%
6 week 2010	33%	43%	+10%

Regarding mean sentence length, the evidence at this initial stage seems to suggest that students do write slightly more on average and as discussed in Section 3.3 it could be said to be desirable to make a claim that, after completion of one of our courses, students can write longer sentences, but much more solid evidence is required before this claim can be verified. Regarding lexical densities, the figures show an increase for the 15 week 2009 and 2010 and this is despite the fact the mean sentence lengths have also increased. Lexical densities are sensitive to text length so these figures are quite impressive. Table 3.7 also offers what appears to be quite strong evidence, at least at this initial stage, of an increase in error-free units over the two successive years. This is emphasised by consideration of the 6 week figures for the 2010 programme which identify a 10% upward movement.

3.6 Revisiting the initial conceptual framework

At this point it was considered worth reviewing the initial typology suggested by Shaw and Liu (1998). In assessing the findings of my pilot study in relation to Shaw and Liu's *incorrect to correct* dimension, I looked at word classes, tenses, and prepositions. Along the *simple to complex* axis I looked for a greater range of lexis and a wider range of complex grammatical features such as passives, relative clauses, long noun combinations, increased lexical densities and longer sentences. Regarding movement from *spoken to written forms*, I looked for (more) nominalisations, fewer phrasal verbs and for the output to be altogether less personal. As an addition to the results shown in Tables 1-7 these rudimentary, manually counted figures for 2009 (22 matched scripts) offer a suggestion that there may be some development of students' writing even if over a relatively narrow front.

Table 3.8 Additional feature movement for 2009

15 week 2009	pre	post	movement
passives	4	16	+400%
relative pronouns	22	44	+100%
word class errors	17	17	0

The pilot study, then, suggested not only that it is possible to use corpus-based methods to identify language features suggesting development in second language writing on EAP programmes, but also that such features are actually attested in data collected from students on the EAP programme being studied in this thesis. On the basis of these observations, it was decided that it would be worth conducting a larger-scale corpus-based project, studying a much wider range of features, over longer and more varied periods of time. The two main research instruments, Coh-Metrix and MAT which are described in Chapters four and five were not used in the pilot study. Coh-Metrix was at this time, still in the development stage and not yet being used for second language acquisition research. MAT had not yet been developed when the study was initiated and is quite complex for pilot study purposes so the Wordsmith programme was the preferred option. The raw frequencies presented in Tables 3.1 to 3.4 were considered the most relevant information to extract from the corpus using the Wordsmith programme and for this reason it was not considered necessary to use the Key Word facility.

It was felt that if the study was going to yield any practical effects, there would need to be work towards the identification of a set of features which had shown development, which had stayed the same and which had regressed, and it was possible that students might not have made any measurable progress at all. It was also felt important to move towards

some tentative description of the language characteristics that would provide a much broader picture of the students' overall development. These characteristics were considered at this stage of the study to include the following: vocabulary distribution, common/ rare nouns: part of speech classes (nouns, verbs, first and second person pronouns, prepositions: grammatical characteristics (nominalisations, past tense verbs, passive voice): syntactic structures (e.g. that relative clauses, to complement clauses) and lexico-grammatical combinations, among many others. With this desire to paint a broader multidimensional linguistic canvas, and taking cognisance of the work of Biber (Biber, 1988; Biber, 2002; Biber and Reppen, 1996), the next phase of the study began. The aim was to obtain a clearer idea of the development of the students' second language writing and also to find a way of making statistical confirmations of the trends which had been tentatively indicated by the manual analysis and Wordsmith programmes and the next step was to compile a larger, more extensive corpus, this took the form of the EAPCORP.

3.7 EAPCORP

EAPCORP stands for *English for Academic Purposes Corpus* and this consists of 526 individual scripts and 263 matched pairs (pre and post course). In order to produce this corpus, the task was to compile a complete set of essays for two separate EAP summer programmes thus offering a profile for analysis of data which is more robust statistically than for a single year. All the scripts were written under exam conditions by hand, so the first task was to read and copy them into plain text format so they could be processed and analysed by the software. This took several months and they were copied in their entirety with no corrections.

The corpus itself presents an opportunity to examine data for two successive EAP

programmes (2010 and 2012 separated by a year), offering an opportunity to examine developmental features. The EAPCORP itself may have some advantages over some other learner corpora in that it is (as discussed in Chapter Two) designed to measure linguistic feature movement for a specific programme and can be considered in this sense to be purpose-built. It also covers two separate years and has a relatively large number (263) of matched pair samples, offering a degree of statistical robustness and this is illustrated in the following table.

Table 3.9 The EAPCORP. Number of matching pairs.

20 week 2010	15 week 2010	10 week 2010	6 week 2010	total
9	27	57	35	128
20 week 2012	15 week 2012	10 week 2012	6 week 2012	total
8	26	65	36	135
total	total	total	total	overall total
17	53	122	71	263

3.8 Conclusion

This chapter firstly outlined the nature of the EAP programmes which were undertaken by the students whose written work forms the basis of this study. It then described how a corpus of essays was created and proceeded to show the results of an initial analysis with an emphasis on identification of significant increases and decreases in the incidence of linguistic characteristics over a limited number of the programmes. The pilot study suggested that there was evidence of such feature change and that it was appropriate to embark on a detailed investigation after the compilation of a larger, more extensive corpus (the EAPCORP) and the use of more sophisticated analytical instruments.

CHAPTER FOUR

THE COH-METRIX ANALYSIS OF EAPCORP

4.1 Introduction

This chapter firstly offers a detailed description of the constitution and characteristics of the Coh-Metrix programme (McNamara *et al.*, 2005; McNamara and Graesser, 2011; McNamara *et al.*, 2014), including a review of both its principal components and individual indices. The results of the Coh-Metrix analysis are then presented. The final part of the chapter provides a discussion of the observed linguistic feature movements from pre-test to post-test, interpreting these in terms of whether and to what extent they may be seen as indicative of improvements in student EAP writing performance as represented by EAPCORP.

4.2 Coh-Metrix: components and indices

One of the two main research tools chosen for this study is Coh-Metrix. As mentioned in Chapter Two, Coh Metrix is a web-based programme that contains many separate indices, grouped under different principal components. Because it provides such a detailed feature analysis, it is necessary here to elaborate its inner workings in some detail, itemising them and describing the relevant scoring systems. Essentially, Coh-Metrix can be described in terms of a small set of ‘components’ each of which can be further subdivided into discrete ‘indices’. In the discussion that follows, I will look at each of these layers in turn.

4.2.1 The principal components

These principal components are 11 in number. The first group (1-11) are called *descriptive* indices and are largely concerned with describing what could be termed the basic arithmetic of the data including mean syllable, word and sentence length. The next group (12-27) are described as *text easability principal component scores* (Graesser, McNamara and Kulikowich, 2011) The purpose is to provide measures of how easy or difficult it is to read a text. They provide "metrics of text characteristics on multiple levels of language and discourse" (McNamara *et al.*, 2014, p.84). The dimension offers 8 components with z scores (how many SDs there are above or below the mean) and percentile scores (0-100% reflecting difficulty relative to each other) for easability with the two scores presenting a “monotonic but not linear relationship to each other” (McNamara *et al.*, 2014, p.86)

The third group of components (28-39) is termed *referential cohesion* and relates to overlap in content words between sentences with coreference being described as “a linguistic cue that can aid readers in making connections between propositions, clauses and sentences in their text base understanding” (McNamara *et al.*, 2014, p.63) There are two dimensions- local (as measured by consecutive sentences), global (as measured by all the sentences in a text) and in terms of the degree of explicitness of the overlap and there are four types listed, noun, argument, stem and content. Component four is termed *latent semantic analysis* and is a measure of semantic overlap between sentences or between paragraphs. LSA “considers semantic overlap between explicit words and words that are implicitly related. (McNamara *et al.*, 2014, p.66) The measures vary from 0 (low cohesion) to 1 (high cohesion) and are offered across 8 components. The calculation of these 8 indices is premised upon a statistical

technique called singular value decomposition whereby a large number of corpus texts is condensed into a narrow range of statistical dimensions which are then used as a basis to calculate conceptual similarity (McNamara *et al.*, 2014; Landauer *et al.*, 2007).

Lexical diversity (indices 48-51) constitutes component five and refers to the variety of unique words (types) in relation to the total number of words (tokens) and this may prompt a consideration of a range of related issues (see Laufer and Nation, 1995.) For example, generally speaking, a low lexical diversity score indicates higher cohesion as fewer words need to be integrated into the discourse field. Complete lexical diversity may indicate a fragmentation of cohesion so it is necessary to exercise caution when drawing any conclusions about writing development with reference to this dimension. There are three indices of lexical diversity, type-token ratio (TTR), measure of textual lexical diversity (MTLD) and measure of vocabulary diversity VOCD. Principal component six (52-60) *connectives* gives an incidence score per 1000 for five general classes of connectives, causal, logical, adversative/ contrastive, temporal and additive (Halliday and Hasan, 1976; Louwse, 2001). Principal component seven (61-68) refers to the *situation model*, an abstraction into mental representation of the text which goes beyond specific words, for example the plot of a story or with informational text relating to for example the circulatory system, the situation model might be the *flow of blood* (McNamara *et al.*, 2014, p.69). The indices for this component relate to the reader's understanding of the situation model.

Component eight, *syntactic complexity* (69-75) is concerned directly with parts of speech categories (nouns/verbs, adjectives etc.), phrases (noun, verb, prepositional, and clauses together with syntactic 'tree' structures. Within this component, Coh-Metrix also measures uniformity and consistency of sentence structure by reference to minimal edit distance (McCarthy, Guess and McNamara, 2009). The minimal edit distance "calculates ... the distance that parts of speech, words or lemmas are from one another between consecutive

sentences in the text”. (McNamara *et al.*, 2014, p.70). As this principal component measures similarities across sentences (Crossley, Greenfield and McNamara, 2008), the minimal edit distance (parts of speech) index calculates “the extent to which one sentence needs to be modified (edited) to make it have the same syntactic composition as a second sentence” (McNamara *et al.*, 2014, p.70.). This index (71) looks at parts of speech but not the words themselves, whereas minimal edit distance (words) and minimal edit distance (lemmas) consider the words. Component nine *syntactic pattern density* is an extension of the previous component by reference to the density of syntactic patterns, word types and phrase types. A high noun and verb phrase incidence for example, is likely to produce more informationally dense text. Component ten is concerned with *word information* and computes incidences of part of speech categories (per 1000 words for indices 84-93) and word frequencies which are calculated using the CELEX database (for indices 94-96). It also employs psychological ratings which relate to psychological and semantic dimensions- 97-105 (the two databases are MRC (Coltheart, 1981) and Wordnet (Fellbaum, 1998). Finally, there is component eleven, **readability** (106-108) a broad dimension discussed subsequently which uses three readability indices.

4.2.2 The indices

There are 108 indices (reduced to 106 in 2013 with omission of indices 38 and 39) grouped under various headings or dimensions. Figure 4.1 presents these in full; individual items will be described and elaborated where appropriate.

Figure 4.1 Coh Metrix indices

index	description
1	the number of paragraphs in the text
2	the total number of sentences
3	the total number of words . A word is defined as " ... anything that is tagged as a part of speech by the Charniak parser" (McNamara <i>et al.</i> , 2005)
4	the mean length of paragraphs
5	the standard deviation of the mean length of the paragraphs
6	the mean number of words per sentence .
7	the standard deviation of the mean length of sentences . A large standard deviation reflects large variations in the text in terms of sentence length
8	the mean number of syllables in words
9	the standard deviation of the mean number of syllables in words - large SDs indicate large variations in syllable lengths.
10	the mean number of letters for all the words in the text
11	the standard deviation of the mean number of letters in the words in the text
12	narrativity z score "a robust component" (ibid) and affiliated to word familiarity and world knowledge. "High narrativity reflects the use of more familiar words combined with a tendency to focus on events and characters rather than objects and ideas" (McNamara <i>et al.</i> , 2014, p.89)
13	narrativity percentile
14	syntactic simplicity z score the degree to which the sentences contain fewer/ more words and simpler/ more complex syntactic structures.
15	syntactic simplicity percentile
16	word concreteness z score concrete words versus abstract words- texts containing more abstract words would be considered more difficult/challenging to understand
17	word concreteness percentile
18	referential cohesion z score words and ideas that overlap across sentences. Texts with high referential

	cohesion are easier to process as they have more connections- more “explicit threads” (McNamara <i>et al.</i> , 2014, p.85)
19	referential cohesion percentile
20	deep cohesion z score - this relates to the match between causal and intentional connectives and causal and logical relationships within the text. High scores on this component reflect more explicit relations between ideas.
21	deep cohesion percentile
22	verb cohesion z score - the degree of verb overlap leading to "a more coherent event structure" (McNamara <i>et al.</i> , 2014, p.85) and identified as more relevant for narrative texts
23	verb cohesion percentile
24	connectivity z score relating to the explicit conveyance of logical relations in the text with explicit use of adversative, comparative and additive connectives.
25	connectivity percentile
26	temporality z score number and consistency of temporal cues
27	temporality percentile
28	noun overlap (local) mean number of sentences with noun overlap from one sentence back to the previous sentence. The noun must match exactly in form and plurality.
29	Index 29 argument overlap - (local) where there is overlap between nouns and pronouns. The term ‘argument’ is used to emphasise contrast between noun pronoun arguments and verb adjective predicates
30	stem overlap (local) a relaxation of the noun constraints by using lemmas which are in common-
31	noun overlap (global) overlap with every other sentence
32	argument overlap (global)
33	stem overlap (global)
34	content word overlap (local) Measures “...proportion of explicit content words that overlap between pairs of sentences (McNamara <i>et al.</i> , 2014, p. 65)
35	content word standard deviation (local)
36	content word overlap (global)
37	content word standard deviation (global)

38	anaphor overlap (local) – deleted in Coh Metrix 3 version
39	anaphor overlap (global) – deleted in Coh Metrix 3 version
40	LSA overlap (local) conceptual/semantic similarity (adjacent)
41	LSA overlap standard deviation (local)
42	LSA overlap (global)
43	LSA overlap standard deviation (global)
44	LSA overlap paragraphs (local)
45	LSA overlap paragraphs (standard deviation) local
46	LSA given/new mean This measure examines how much <i>given</i> versus <i>new</i> information is to be found in each sentence
47	LSA given/new standard deviation
48	type/token ratio content word lemmas- correlated with text length
49	TTR all words- correlated with text length
50	Measure of textual LD all words “the mean length of sequential word strings that maintain a given TTR value (McNamara <i>et al.</i> , 2014, p. 67)
51	Measure of vocabulary diversity all words- “(computational) procedure which fits TTR random samples with ideal TTR curves” (McNamara <i>et al.</i> , 2014, p. 67)
52	all connectives
53	causal connectives (e.g. because, so)
54	logical connectives (e.g. or)
55	adversative and contrastive connectives (e.g. although, whereas)
56	temporal connectives (e.g. first)
57	expanded temporal connectives (e.g. until)
58	additive connectives incidence (e.g. and, moreover)
59	positive connectives incidence (e.g. also moreover)
60	negative connectives incidence (e.g. however, but)
61	causal content- causal verbs (e.g. hit and move) incidence reflecting a change of state
62	causal content causal verbs and causal particles incidence

63	intentional content intentional verbs (e.g. contact, talk) incidence (intentional actions, events and particles)
64	causal cohesion (ratio) of causal particles to causal verbs
65	intentional cohesion (ratio) of intentional particles to intentional actions/events. Texts are judged to be more causally cohesive if there are proportionally more connectives that relate actions and events in the text (McNamara <i>et al.</i> , 2014, p.69)
66	LSA verb overlap the extent to which verbs linked to actions events and states are repeated
67	wordnet verb overlap
68	temporal cohesion tense and aspect repetition mean- the consistency of tense and aspect across a passage of text.
69	mean number of words before the main verb/ left embeddedness
70	mean number of modifiers per noun phrase
71	minimal edit distance parts of speech
72	minimal edit distance words
73	minimal edit distance lemmas
74	syntactic structure similarity adjacent
75	syntactic structure similarity all sentences
76	noun phrases incidence
77	verb phrases incidence
78	adverbial phrases incidence
79	preposition phrases incidence
80	agentless passives incidence
81	negation expressions incidence
82	gerunds incidence
83	infinitives incidence
84	nouns incidence
85	verbs incidence
86	adjectives incidence
87	adverbs incidence

88	personal pronouns incidence
89	first person singular pronouns incidence
90	first person plural pronouns incidence
91	second person pronouns incidence
92	third person single pronouns incidence
93	third person plural pronouns incidence
94	content word frequency mean
95	word frequency all words mean
96	minimum word frequency in sentences mean
97	age of acquisition norms content words (related to children's language acquisition and the premise that certain content words are likely to occur at an earlier developmental stage- of only tangential relevance to this study)
98	familiarity content words related to adult language processing- how familiar the word seems to an adult – average ratings multiplied by 100 (example of high familiarity- <i>milk</i> , example of low familiarity <i>cornet</i>)
99	concreteness content words (concrete/abstract continuum average ratings multiplied by 100 (example of low concreteness- <i>protocol</i> , example of high concreteness – <i>ball</i>)
100	imagability content words – how easy it is to construct a mental image of word (example of low imagery word- <i>reason</i> , example of high imagery word <i>hammer</i>)
101	meaningfulness - derived from a corpus developed by Toglia and Battig (1978) high meaningfulness words are highly associated with other words (for example <i>people</i>) low meaningfulness words are less strongly associated with other words for example <i>abbess</i>)
102	polysemy - the range of senses or meanings of a word (example <i>account</i> , <i>table</i> the mean wordnet polysemy values for all content words. Indicative of text ambiguity but may also reflect presence of higher frequency words.
103	hypernymy - Coh-Metrix here relates to nouns derived from wordnet and reflects the use of more or less specific words, in a hierarchy. For example, <i>bolt</i> , <i>run</i> , <i>travel</i> , <i>go</i>) Low values reflect less specific words. The term is defined as “the number of levels in a conceptual taxonomic hierarchy that is above the word” (McNamara <i>et al.</i> , 2014, 44)

104	hypernymy verbs
105	hypernymy nouns and verbs
106	Flesch reading ease - scored 0-100, the higher the score the easier it is to read
107	107 Flesch- Kincaid Grade level - reading ease score converted to a (US high school) grade score. Higher numbers indicate more difficulty in reading
108	the second language readability score . This uses content word overlap, sentence syntax similarity and word frequency and relates to “... challenges at the sentence and word level... and considers the cohesion between sentences” (McNamara <i>et al.</i> ,2014, p.80) The authors consider this index an advance on indices 107 and 106 because it adds text cohesion to sentence and word level difficulties. A higher score means the text is more readable (for a discussion of the issues related to LSA and wordnet algorithms, see Sabatini, Albro and O’Reilly, 2012).

4.3 The procedure

The scripts were then entered individually, pair by pair into the Coh-Metrix, using a simple web interface (user name and password protected). The 15,000 character maximum capacity afforded sufficient space in the programme to cover the number of words that most students could produce in 30 minutes. The Coh-Metrix was then applied to the corpus with approximately three minutes taken for the Coh-Metrix to analyse each script. The results were produced plus a hard copy for every matching pair. In its printed version, it has two formats, the landscape and the portrait and it is usually possible to select a preferred option. The scripts for the 20 week cohort 2010 and 2012 were entered first (for both 2010 and 2012) and this was repeated for the 15, 10 and 6 week programmes with the resulting data examined using eight large spread sheets (4x2) one for each cohort. A threefold categorisation was used:

1 indicates that there has been a **score increase**

-1 indicates that there has been a **score decrease**

0 indicates that there is **no difference** between the totals (to the nearest two thousandths either side)

Table 4.1. Score range for programmes

programme	maximum score	minimum score
20 week 2010	+9	-9
15 week 2010	+27	-27
10 week 2010	+57	-57
6 week 2010	+35	-35
20 week 2012	+8	-8
15 week 2012	+26	-26
10 week 2012	+65	-65
6 week 2012	+36	-36

After the information was entered on the spreadsheets, the data was examined and scores awarded for each individual index. Each positive (1) and each negative (-1) and any 0s (these being very few in number) were counted and listed under the various course length headings. As a consequence, a series of results was produced.

4.4 The results of the Coh-Metrix analysis

As Table 4.2 will indicate, there are 11 columns per index, each showing the aggregated scores for the respective programmes. An aggregate score of 0 would indicate no overall increase or decrease so the higher or lower the score, the more significant the movement. For example, with the 10 week programme in 2010 a score of 57 would be a maximum, as there are 57 pairs indicating that every pair has shown an increase. Conversely, a score of -57 would indicate that every pair has shown a decrease. For 2012 the maximum score would be 65 and the minimum -65 as there are 65 pairs for this cohort. Columns 9 and 10 show the totals for each year and column 12 shows the total for both 2010 and 2012. This final, two year aggregate column is referred to extensively in the subsequent chapter and frequently used as a comparator with the other main research technique employed in the study, the MAT.

Reading Table 4.2, for *index 1* the 2010 20 week total is minus 1, meaning that of the 9 scripts, 4 were positive, 5 were negative giving a total of -1. For *index 1* 20 week 2012, of the 8 scripts 5 were positive and three negative, giving a total of 3. For the 15 week 2010, of the 27 scripts 15 were positive and 12 negative giving a total of plus 3 and in the same way for the rest of the cohort.

Table 4.2 Full results of the Coh-Metrix analysis

index	20 10	20 12	15 10	15 12	10 10	10 12	6 10	6 12	2010	2012	total
1	-1	3	3	-9	-6	3	4	1	0	-2	-2
2	2	2	4	-6	-6	-7	8	-1	+8	-12	-4
3	5	4	12	-2	6	-6	14	-11	+37	-15	+22
4	5	-2	4	10	3	-7	0	1	+12	+2	+14

5	-4	3	2	-8	-5	1	4	-1	+3	-5	-2
6	3	0	4	2	17	8	10	1	+34	+11	+45
7	1	2	7	-8	-9	8	6	-7	+5	-5	0
8	7	-7	19	10	28	23	15	7	+69	+33	+102
9	9	-6	17	5	16	32	14	-2	+56	+29	+85
10	7	-8	19	12	30	32	16	9	+72	+45	+117
11	6	-6	19	6	25	30	10	-2	+60	+28	+88
12	-1	2	-19	-6	-31	-24	-10	-1	-61	-29	-90
13	-1	5	-15	-5	-23	-24	-14	-5	-53	-29	-92
14	-1	0	1	0	-8	-15	-6	-7	-14	-22	-36
15	-1	0	-1	4	-4	-9	-3	-2	-9	-7	-16
16	-3	2	-1	10	-6	-4	0	-3	-10	+5	-5
17	-3	2	1	-6	4	4	2	-9	4	-9	-5
18	3	2	1	-8	8	-6	6	-3	18	-15	+3
19	3	2	5	-4	1	-7	6	-5	+15	-14	+1
20	-3	-4	-3	2	-13	-18	4	-9	-12	-29	-41
21	-3	-4	-3	0	-5	-16	-2	-7	-13	-27	-40
22	-1	0	3	-5	6	-5	-2	-1	+6	-11	-5
23	-2	1	-1	-1	1	-5	2	-1	0	-6	-6
24	5	4	3	-2	-10	-4	-8	0	-10	-2	-12
25	0	3	5	2	5	13	0	3	+10	+21	+31
26	-2	-4	1	-6	5	-4	6	3	+10	-11	-1
27	-2	-6	-5	-2	13	4	6	1	+12	-3	+9
28	5	-2	5	4	13	2	10	0	+33	+4	+37
29	3	2	5	-4	0	-4	10	-6	+18	-12	+6
30	5	0	13	2	8	11	12	1	+34	+14	+48
31	3	-2	9	-1	17	10	7	7	+36	+14	+50
32	2	2	-9	-10	6	0	-3	3	-4	-5	-9
33	7	2	3	5	14	18	4	-11	+28	+14	+42

34	5	0	3	-1	4	-15	-1	-1	11	-17	-6
35	-1	0	-2	-6	-9	-15	-1	-11	-13	-32	-45
36	0	1	-3	-3	4	-9	-4	-5	-3	-16	-19
37	-1	1	3	-3	-5	7	-8	-10	-11	-5	-16
38*	X	x	X	X	-18	-13	X	1	x	x	X
39*	X	x	X	X	-25	-21	X	-4	x	x	X
40	5	2	16	-2	-4	-3	9	-3	+26	-6	+20
41	1	0	8	-7	-12	-6	4	0	+1	-13	-12
42	5	1	8	8	11	0	8	-1	+32	+8	+40
43	5	4	-3	-9	-1	2	-5	-10	-4	-13	-17
44	-6	3	1	-11	-4	3	4	-1	-5	-6	-11
45	-6	0	-1	-2	1	1	0	0	+6	-1	+5
46	6	0	9	-10	8	-2	7	-3	+30	-15	+15
47	0	0	-3	-8	-18	-5	-4	-4	-25	-17	-42
48	3	-2	-6	0	-4	4	-15	-3	-22	-1	-23
49	-1	-2	-15	5	3	1	-9	-10	-22	-6	-28
50	3	-6	-5	6	-15	0	-4	-7	-21	-7	-28
51	3	-2	-7	10	-21	0	-14	-3	-39	+5	-34
52	-5	-4	-7	-2	-5	-10	4	1	+13	-15	-2
53	-1	-4	7	-4	-3	-9	-6	-5	-3	-14	-17
54	-5	-2	-4	0	-9	-20	8	-5	-9	-27	-36
55	-5	-6	-5	-2	0	-8	8	3	+2	-13	-11
56	3	2	-1	1	-5	0	-6	5	-13	+8	-5
57	2	0	-7	6	9	-16	1	-3	+5	-13	-8
58	-1	-2	-11	0	0	-8	8	3	+4	+7	+11
59	0	0	0	0	1	0	0	0	+1	0	+1
60	0	0	0	0	1	0	0	0	+1	0	+1
61	1	-2	-9	2	-8	-4	-20	-5	-36	-9	-45
62	-3	2	-3	6	-15	-6	-16	-5	-37	-7	-44

63	-1	4	3	-1	-5	-4	2	1	-1	+2	+1
64	-1	2	-8	-3	-5	-12	8	1	+6	-12	-6
65	1	-2	4	4	2	x	1	-9	+8	-7	+1
66	-1	-2	2	-3	5	3	5	3	+9	+1	+10
67	-1	0	-3	2	0	2	0	2	-4	+6	+2
68	-4	-6	0	-3	10	0	8	-1	+14	-10	+4
69	7	4	17	4	5	6	10	-1	+39	+13	+52
70	5	4	7	10	21	26	3	-4	+39	+36	+75
71	-1	1	-1	-10	-15	-1	-13	-11	-30	-21	-51
72	2	-2	-9	1	-15	0	-9	-5	-29	-6	-35
73	-1	0	-6	-1	-23	-8	-13	1	-43	-8	-51
74	1	0	-2	8	-9	-5	-4	-2	-14	+1	-13
75	1	-4	2	-4	-7	7	-3	4	-7	+3	-4
76	-1	-6	1	6	-10	-10	-6	-1	-16	-11	-27
77	-1	-2	7	0	-23	-22	10	-3	-7	-27	-34
78	1	0	-5	4	-10	-6	-6	-3	-20	-5	-25
79	-1	-4	11	-2	6	8	12	5	+28	+7	+35
80	7	-4	6	4	6	27	6	12	+25	+39	+64
81	1	-2	-7	-7	-8	-5	-11	8	-25	-6	-31
82	-5	2	3	4	0	22	5	16	+3	+44	+47
83	1	-4	15	9	13	-9	-4	-3	+25	-7	+18
84	5	-4	5	6	9	22	4	-1	+23	+23	+46
85	-3	0	1	4	-15	0	0	1	-17	+5	-12
86	7	-4	7	16	-7	3	-12	1	-5	+16	+11
87	-1	6	-9	10	-21	-6	-6	-1	-37	+9	+28
88	-5	4	-17	-6	-31	-23	20	7	-33	-18	-51
89	-8	6	-17	-8	-46	-44	-15	-9	-86	-55	-141
90	-4	-1	-8	-10	-27	-18	-15	-9	-54	-38	-82
91	-5	-5	-2	4	-17	-6	3	-4	-21	-11	-32

92	3	-3	0	0	0	1	-3	4	0	+6	+6
93	3	-3	3	3	0	11	0	8	0	+19	+19
94	-5	2	-17	-10	-28	-24	-3	-3	-53	-35	-88
95	-7	4	-5	-16	-10	0	6	-5	-16	-17	-33
96	-5	0	-3	-1	-4	-8	-10	-3	-22	-12	-34
97	-9	-4	15	10	11	22	16	1	+33	+29	+62
98	-9	4	-15	4	-28	-16	-12	-11	-64	-27	-93
99	-7	2	5	-6	-7	10	-12	3	-21	+9	-12
100	-7	4	5	-10	-5	8	-10	-1	-17	+1	-16
101	-1	2	9	2	-3	2	-14	-13	-9	-7	-16
102	1	4	-12	2	-15	-14	-4	-3	-30	-11	-42
103	1	2	11	6	9	24	0	12	+21	+54	+75
104	3	-4	9	0	-1	8	10	11	+21	+15	+36
105	5	4	11	6	14	19	5	11	+35	+40	+70
106	-7	6	-21	-6	19	-3	-20	-7	+29	-10	+19
107	7	-2	17	2	17	2	18	1	+59	+3	+62
108	-3	2	-10	-3	-15	-20	-2	-1	-30	-22	-52

A note on the indices: the number of indices was reduced from 108 to 106 in October 2013. For the sake of clarity it was decided to map the new categories onto the old and show the omissions. The two indices which were omitted were index 38* (anaphor overlap adjacent sentences) and 39* (anaphor overlap all sentences).

The following two tables show a graphic representation of the feature movements identified by the COH-Matrix analysis.

Table 4.3 (a) Coh-Metrix feature analysis- movements from the beginnings to the ends of the programmes, twenty features which have shown the most increase.

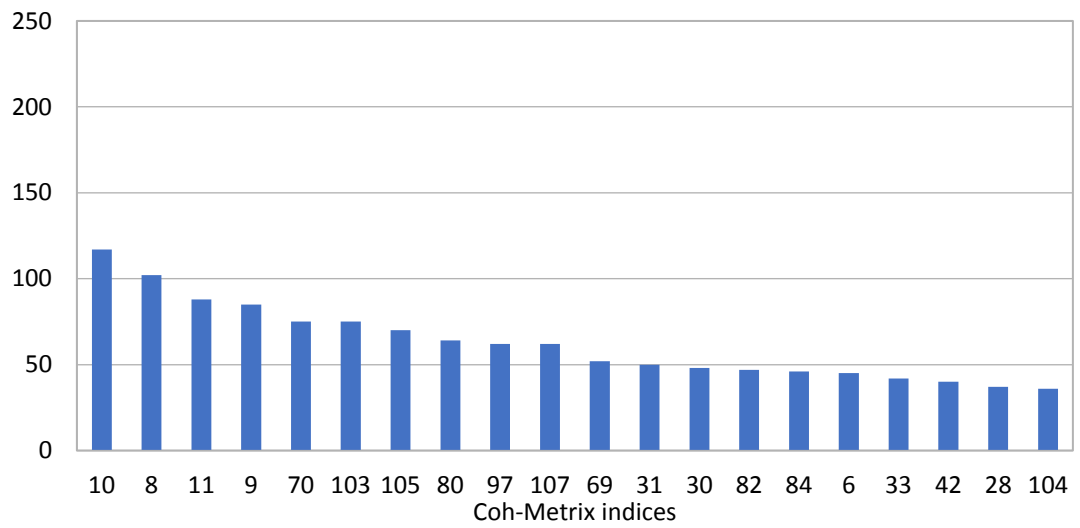
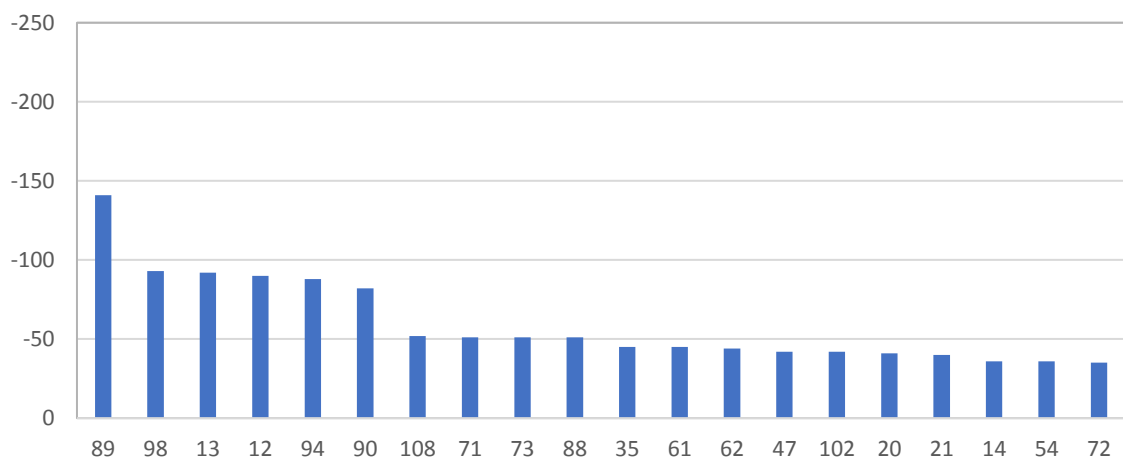


Table 4.3 (b) Coh-Matrix feature analysis- movements from the beginnings to the ends of the programmes, twenty features which have shown the most decrease.



Coh-Matrix indices

4.4.1 Discussion of increases in individual indices

The indices which have shown movement are now discussed and comment is offered on the possible explanations and implications where relevant. The main results are found in Table 4.2 and further examples are offered to illustrate feature movements over specific course ranges. (Those indices which appear in Table 4.3 (a) have been emboldened in the text.)

Firstly, **index 6, mean numbers of words per sentence** shows a moderate, if not spectacular increase for 6 out of 8 courses with students generally writing more in the time available which could be construed tentatively as a positive finding. One explanation for this could be that students are more comfortable and experience less stress when writing in English at the end of the programme rather than the beginning. Table 4.4. illustrates this:

Table 4.4 Mean number of words per sentence for all courses

course	+	-	0	% of cohort increase
20 week 2010	6	3	-	66%
20 week 2012	4	4	-	50%
15 week 2010	16	9	1	62%
15 week 2012	12	15	-	44%
10 week 2010	32	24	-	57%
10 week 2012	40	25	-	62%
6 week 2010	22	12	-	65%
6 week 2012	18	17	-	51%

% increase = % of the cohort which has increased (50% means no overall increase- 100% means every student has increased score level- 0% means that every student has decreased their score for this particular feature)

Regarding **index 8, mean number of syllables in words**, which shows a clear increase for 7 out of 8 courses, it is possible to suggest that students are using more latinate, formal words which have in general, a longer syllable length. This may imply that academic vocabulary, a feature of the syllabus, is being increasingly used and possibly internalised to a greater extent. This simple finding is positive, and might lead to the suggestion that the course has achieved at least one of its stated objectives, namely to help students become more competent users of academic English. **Index 9** which indicates the standard deviations of **index 8** would suggest that there is considerable variation in syllable length across the sample. **Index 10, word mean number of letters**, shows a fairly strong increase with the exception of the 20 week programme 2012 and it appears that students are tending to use longer words.

The following two examples of first sentences are taken from the 10 week programme 2010 and the 15 week programme 2012 respectively.

Example 1

I come from China. (pre) Average word length of 3.5

First of all the weather of my motherland and the UK are different (post) Average word length of 4.0

Example 2

There are many differences between my country and the UK. (pre) Average word length of 4.7

Different cultures between China and the UK have led to many differences. (post) average

word length 5.08

These may well be from a more academic or at least more formal discoursal range and this could be the result of exposure to multisyllabic, academic vocabulary. The figures in Table 4.5 illustrate the increases across the cohort.

Table 4.5. Average word length all courses

course	+	-	0	% of cohort increase
20 week 2010	8	1	-	78%
20 week 2012	0	8	-	0%
15 week 2010	23	4	-	79%
15 week 2012	19	7	-	73%
10 week 2010	47	17	-	72%
10 week 2012	44	12	-	66%
6 week 2010	25	9	-	74%
6 week 2012	22	13	-	66%

Regarding **index 11, the average word length standard deviations**, there is considerable variation in mean numbers. There is an increase in **noun overlap (index 31)**, the overlap between all the nouns in the text which is designed as a measure of referential cohesion by the Coh-Metrix. There is also increase in **stem overlap, index 33** and **index 30** which is the **measure of stem overlap in adjacent sentences** where a noun in one sentence is matched with a content word in the previous sentence that share a common lemma e.g. cost/costly (McNamara *et al.*, 2005). This may indicate that the students' writing is becoming more cohesive in terms of argument and that they are becoming more able to present an

argumentative thread. This issue of development of argument is a complex one and for a discussion of the issue, see Crossley *et al.* (2011) where in a study of grade and college level essays, the learners developed in terms of syntactical complexity and word diversity but exhibited lower cohesion in the form of explicit cohesive cues.

The results for **left embeddedness (index 69)** and **modifiers per noun phrase (index 70)** appear to be a firm indicator of progress, and this is a feature of the EAP programme to explicitly focus upon the noun phrase. It was made clear to the students during their courses of study, that one of the characteristics of academic texts is that they tend to possess long noun phrases along with attendant modification and extensive exercises were provided with the explicit aim of expanding the noun phrase. This result then, was positive and may suggest increase in syntactic flexibility and evidence of increased awareness of the syntactic structure of the target language- English.

The figures in Table 4.6 illustrate the increases for this feature across the cohort

Table 4.6 Average number of modifiers per noun phrase

course	+	-	0	% of cohort increase
20 week 2010	7	2	-	78%
20 week 2012	6	2	-	75%
15 week 2010	17	10	-	66%
15 week 2012	16	8	2	62%
10 week 2010	40	14	2	71%
10 week 2012	43	18	-	66%
6 week 2010	18	15	1	53%
6 week 2012	15	19	1	43%

Within the syntactic pattern density principal component, two indices suggest positive

movement: *preposition phrase density* (index 79) and **agentless passive incidence (index 80)**. It may be the case that students are loading their texts with more information. It is often observed that prepositions do cause problems for learners, especially from East Asia, (Cho, K. 2002) so there is a focus on these during the EAP programmes and it may be that as students become more comfortable with prepositions, then one might expect their use to increase. Regarding **agentless passives (index 80)**, this increase may be influenced by students' exposure to the variety of academic texts, particularly scientific and economic ones which do contain a high density of agentless passives (Yamamoto, 2006; Perez, 2004; Martin-Martin, 2008).

The following two examples are taken from the same (post) text produced by a student on the 10 week programme 2010. The student had produced no passives in the pre text.

Example 1 (post)

In China, the weather is stable and predictable. Umbrellas *are not always requested*. In the UK, the weather is changable and unpredictable.

Example 2 (post)

The courses of China are unitary. In other word, the main lectures are the main body our course. In the UK, compositions **were included** in the preessional courses.

In fact, one of the features of writing in academic discourse is this very use of agentless passives (Biber, 1998, p. 938) and this finding would probably not be expected in the pre samples where the writing may well be *more talk influenced*. This thread will be discussed further in Chapter Six.

There are three indices within the word information principal component which show increases. The first is **index 84, noun incidence**, which may give a slight indication of greater awareness of the centrality of nouns in academic texts. **Hypernymy for nouns**

(**index 103**) and for **nouns and verbs (index 105)** also show increases. This may reflect both an increased use of more specific words as the students' vocabularies develop and the fact that subject-specific English is taught on the EAP programme with a strong vocabulary focus included within our teaching materials. Students are invited to consider words in a variety of contexts. Interestingly this hypernymy for nouns and verbs shows a small if unspectacular increase across *all 8 cohorts*.

4.4.2 Discussion of decreases in individual indices

Those indices which show a decrease are now considered. (Those indices which appear in Table 4.3 (b) have been emboldened in the text). This does not necessarily mean that the *feature in question indicates a regression* on the part of the student. In other words, a reduction in use of some the features may be desirable and the best example of this would be **index 89**, use of the **first person pronouns**, which will be discussed shortly. Firstly, **indices 12 and 13** which relate to **narrativity (z score and percentile)** show large decreases. According to the Coh-Metrix overview (McNamara *et al.*, 2005), low scores on these two indices would indicate that the students are composing texts in a less narrative, less familiar fashion. The students may be trying to use a more expansive, less personally oriented style and it may be that the nature of the question which the students were asked - *What are the differences between your country and the UK* - offers the opportunity to produce a more open, less narrative essay. Also within the text easability component/dimension is **index 20 deep cohesion z score**. This refers to the level of explicitness between causal relationships in the text as expressed by connectives. Low deep cohesion may reflect a lack of fit between the students' ideas and their use of connectives. This reduction is also found in **index 21** which measures the **deep cohesion percentile**, a feature which relates to the degree of inference

needed by a reader in terms of ideas in the text (McNamara *et al.*, 2014, p.85). This may possibly be a characteristic of a developing writing style where a writer allows the text to run ahead of the cohesive elements so the developing writer forsakes precision in order to convey ideas. This is supported to some extent by the next index to show a marked decrease, ***logical connectives incidence, index 54*** and it might be suggested that students are gaining confidence and beginning to express their own ideas but that their output has a tendency to be disconnected in terms of cohesion.

Within the principal component of syntactic complexity, ***index 73, minimal edit distance lemmas***, shows a marked decrease and this finding may be related to the increased subject-specific vocabulary that the students might be expected to have at their disposal after an intensive EAP programme. If there is greater degree of hypernymy which the results would suggest, then this may reduce the edit distance for lemmas. The next finding is somewhat contrary with a slight reduction in *noun phrase density* (index 76) with a higher noun phrase density indicative of more informationally dense text. *Verb phrase incidence* (index 77), also an indicator of information density is markedly lower for the 10 week programme in comparison with the other 20, 15 and 6 week cohorts (-23 and -22 for 2010 and 2012 respectively). The reasons for these reductions are unclear.

The incidence (per 1000 words) of ***pronouns (index 88)*** which is augmented by the considerable and consistent reduction in the use of ***first person singular pronouns***, as shown by ***index 89***, is notable. This is the most striking of the Coh-Metrix findings to date and one that overtly supports the findings of the pilot study using Wordsmith 5 software (see also Chapter Three.)

The following two long examples of pre and post scripts are taken from the 15 week programme 2012,

Pre:

There are a lot of differences between China and the UK, such as culture, custom and food. **I would** like to make a comparison in something about traffic between China and the UK. First, roads and streets in China are different from those in the UK. Generally speaking, roads and streets are extremely straight that lead the north, to the south and the West to the east, except the ring roads. However, **I found roads** in the kingdom seem totally different, most of roads here are curvy. Second, the public transport is different between two countries, In China, bus is the main public transport and a few cities have underground systems. compared with that, the public transport system is more advanced in the UK. **I find it is** very easy to take bus, underground and railway, simply because stops covered the whole city. Third, people have different custom when driving cars

Post

First of all the car run much quicker than Chinese styles, while, when you ride a bicycle, be careful the road and each sides. The shop and supermarket usually closed early in the afternoon, so there were nothing to do then, and lots of the inhabitants choose to join in the party with familiar friends. On the converse, especially in the Southern part of china, the shop will open till midnight. We can enjoy ourselves shopping, eating and playing anytime we want. Moreover, gambling in Britain is not banned or restriction for people adventure with their fortune, while it is quite different in china with only macao is allow to do so. The inhabitants are usually politely and socially, every morning when you go out for walk or running, the inhabitant you meet would always kind and friendly to say hello to you, which would make you to keep energetic with your daily life. There are also some discomfortable manners for us to follow in Britain, especially when you ride down a bicycle during the night. For the laws in Birmingham, you have to wear colourful jacket and with light on both sides of the bicycle although it's much safer for the riders to take, the strict require also caused boring item to be careful. Lastly, the teachers in Birmingham are usually humrous and patient when they were giving a lecture or have a class. However, moving to the final examination, no teacher would give you a help in order to assist you to pass it. However, it is quite different in China, relating to the boring education system, Large numbers of teachers only pay attention to what they have demonstrate to their students while careless about what the students have learn

The pre sample contains three examples of the first person pronoun **I** and the post sample

contains none.

It can be assumed that a reduction in the personal nature of writing is a positive characteristic and if we consider the acquisition of the features of academic writing to be a target for our students, then such a large reduction is welcome. There are many perspectives to consider here, especially as there are occasions when a personal approach may be appropriate in academic writing (there are many online guides related to this topic, for example, Vance, 2005). However, it is possible to suggest that there is evidence here of a more expansive, complex and less personal writing style which is *less like talk and more like writing* (Shaw and Liu, 1998). The figures in Table 4.7 illustrate the reductions across the cohort and for this feature the percentage *reductions* are shown.

Table 4.7 Incidence of first person singular pronouns

course	+	-	0	% of cohort decrease
20 10	1	7	0	88%
20 12	-	8	1	89%
15 10	5	22	-	81%
15 12	9	17	-	65%
10 10	6	50	-	83%
10 12	17	47	1	72%
6 10	9	24	1	70%
6 12	12	21	2	60%

Also within the word information principal component is **mean word frequency for content words (index 94)**. In this component, nouns, verbs, adjectives and adverbs are classified as content words whilst prepositions, determiners and pronouns are assigned to

the function word category so the reasons for this fall in content word frequency are difficult to ascertain. If confirmed, this would represent a reduction in lexical density which in a sense goes contrary to the findings of the pilot study which suggested small but significant increases. Higher lexical densities do tend to be characteristic of academic writing (Laufer and Nation, 1995; Laufer and Goldstein, 2004). However, as discussed in Chapter Two, measures of lexical density may not reflect relative syntactic complexity, so it is possible to produce a sophisticated piece of writing with a lower lexical density and it may be the case that students are producing work with a higher level of syntactic sophistication. **Index 98, familiarity for content words**, also shows a large reduction. The score on this index (which was constructed by using raters who assigned familiarity scores to words) may suggest that the students are tending to use more academic, specialised and less familiar vocabulary.

The following pre and post example from the 20 week 2012 course shows the reduction of familiarity for content words. The post sample contains more specialised vocabulary as emphasised by the words in italics.

20 week 2012

Pre

It is my first time to come UK. I have been found a lots differences among CHINA and UK. As you now, I am a Chinese boy. So what remind me at first is the language. Because of my poor English, It is hard to discuss problems with others. Sometime, I can not translate my opinions from Chinese to English.. Secondly, the weather there is terrible. When I get up on the morning, I thought that It is gone be a nice day. But a few hours pass, the sky turn to dark. heavy rains in a sudden. When the days which I was in China, Tt can not be a day with several weathers. I think that the weather forcast in England must be very difficult. Thirdly, the difference between China and UK is the traffic rules. The cars should be driven on the right side of the road in China. But in the uK, it is totally in the other side. And when people wanne across the road, they should look right first. Except the rules, the car's speed is very fast, even if the driver is an old lady. After **I arrived** on the uK, I found that the second- hand cars price in Uk is much cheaper than it in China.

That's all.

Post

After studying at University of Birmingham for more than four months, I found that there are many differences between China and the UK. For me, the most important difference is that the learning style of these two countries. In general, Chinese students prefer passive teaching methods such as lecture and demonstrations. The *rote learning style* is widely used by Chinese students. This kind of *learning and teaching style* could be traced by the *Confucian teaching method*. According to the *experince of pre-sessional course* in the EISU, I found that *Problem solving ability* is playing a *significant role* in Britain teaching systems. However, in China, the *student's achivement is largely judged through the written examination*, which are not designed to test their ability to work with others and solve practical problems. It means the *Problem- based system* is completely different from my *past learning method* when **I study** in the university of Brimingham at first time. In another side of *study progress is the teacher*. In China, due to the *power distance*, teachers are treated in a high respect place, if the students ask questions or debate with teachers in the class, it would be regarded as a *disrespectful behavior*. But, in the UK, it is widely different with China. In the class, teachers are gladed to have a conection with students and usually ask students to find the answers of *questions by team-work* under teacher's lead. It is true that Chinese students will face many problems when they begin to study abroad, because of the differences between China and the UK. Although, the different education systems could be a tough problem, we would able to get well used to it.

Also within the word information component **index 102, polysemy for content words**, a clear reduction is shown. This may be as the Coh-Metrix designers suggest, a result of the reduced number of higher frequency words. These tend to have multiple meanings and this may account for the reduction in polysemy which the index suggests and may again point to the development of a more specialised vocabulary. This may also complement the findings for **index 103, noun hypernymy** (4.4.3)

Readability, component 11, yields one noticeable reduction, that of the **Coh Metrix L2 readability** score and the construction of this **readability index (108)** is worth comment. Crossley, Allen and McNamara (2011) offer a detailed explanation of the index, comparing

its capacity to measure text comprehensibility. In stressing the cognitive nature of the index, the authors emphasise the limitations of traditional readability formulae when applied to L2 texts, because of a lack of syntactic and rhetorical reference. The Coh Metrix formulae were designed to include variables which "better reflected the psycholinguistic and cognitive processes of reading" (Crossley, Allen and McNamara, 2011, p.88). These variables are a word frequency index - the CELEX frequency (log mean for content words), a word overlap index- content word overlap (how often they overlap in adjacent sentences) and an index of syntactic similarity - measuring the "uniformity and consistency of parallel syntactic constructions" (Crossley, Allen and McNamara, 2011, p.91). The idea is that if the structures are similar then the cognitive demands on the reader are lower. These three variables are combined using statistical techniques to produce a readability formula which reflects more closely, the readability of L2 texts because it incorporates recognition of psycholinguistic and cognitive reading processes, which are often lacking in traditional readability formulae (Crossley, Allen and McNamara, 2011, p.88)

On a broad level, this **reduction in readability scores** would suggest that the students' writing is becoming more difficult to read, based on the readability formula used in this index which is constituted by reference to content word overlap, sentence syntactic similarity and word frequency (see McNamara *et al.*, 2014). It is possible therefore to regard this method of assessing readability as indicative of levels of sentence complexity, in which case the lower scores recorded would suggest that students' work is becoming harder to read because it is more complex. This is in many ways contrary to the general notion of readability indicating better writing performance and this is an open debate with these L2 readability scores possibly indicative of the development of a more complex, expansive writing style. The figures in Table 4.8 illustrate the reductions across the cohort and percentage *decrease* scores are shown.

Table 4.8 L2 readability scores

course	+	-	0	% of cohort decrease
20 10	3	6	-	63%
20 12	3	5	-	63%
15 10	9	18	-	60%
15 12	12	14	-	54%
10 10	18	38	-	68%
10 12	25	40	-	62%
6 10	16	18	-	53%
6 12	17	18	-	51%

4.4.3 Feature movements across programmes

The following tables show the feature movements for 8 out of 8 and 7 out of 8 courses.

These are presented to illustrate the breadth of significant feature movement which has taken place over a high proportion of sample pairs. As Table 4.9 illustrates, there are only two indices which have shown movement across the whole range of courses (263 from 263 pairs) However, this range is considerably extended for seven out of eight (255/6 from 263 pairs) of the programmes.

4.4.4 Feature movement for all courses

The following features show movement for *all* cohorts. Raw figures are listed together with percentages as appropriate to the programme

Table 4.9 Indices showing movement (+) or (-) for all 8 cohorts. 263 pairs

index	description	principal component	2010	2012	total
90	1st person plural pronoun	word information	-27 (41%)	-18 (32%)	-36.5%
105	hypernymy nouns and verbs	word information	+25 (19%)	+40 (31%)	+25%

Table 4.9 describes only 2 indices which show movement across *all* 8 cohorts and this may appear, superficially at least to be a negative result. With 108 categories of variation it would certainly be possible to expect movement across a greater range of features, however the 20 week results may be untypical of the overall student body and this is discussed subsequently.

4.4.5 Feature movement for 7 out of 8 courses.

Regarding movement observable for 7 out of 8 courses, Table 4.10 indicates that there is a broader range of features which show movement.

Table 4.10 Indices showing movement for 7 cohorts (15, 10, 6 week programmes plus 20 week either 2010 or 2012 and excluding indices 90 and 105) **254/5 pairs**

index	description	principal component	2010	2012	total
8	mean word length no of syllables	descriptive	+69 + 54%	+40 +29%	+31.5%
10	mean number of letters in words	descriptive	+72 +57%	+53 +39%	+48%
12	narrativity z score	text easability	-66 -52%	-31 -23%	-27.5%
13	narrativity percentile	text easability	- 53 -42%	-34 -25%	-37.5%
21	deep cohesion percentile	text easability	- 13 -10%	-27 -20%	-15%
33	stem overlap all sentences	referential cohesion	+28 +17%	+14 +10%	+13.5%
70	av modifiers per noun phrase	syntactic complexity	+36 28%	+28 +21%	+24.5%
71	minimal edit distance parts of speech	syntactic complexity	-30 -24%	-21 -15%	-19.5%
80	agentless passive density	syntactic pattern density	+18 14%	+43 +32%	+23%
89	1st person singular pronoun incidence	word information	-86 -68%	-61 -45%	-57.5%
94	frequency content words	word information	-53 -42%	-37 -27%	-34.5%
96	log frequency for content words	word information	-22 -17%	-12 -9%	-13%

107	Flesch-Kincaid grade level	readability	+59 46%	+3 2%	-24%
108	L2 readability	readability	-30 -24%	-24 -18%	-21%

*Index 21 15/12= 0 *Index 33 6/12=minus11 *6week12=minus 4

Table 4.10, which presents an expanded range of feature movement, shows that the two headline characteristics, those with the largest levels of movement are for **mean number of letters in word (index 10)** with an increase of +48% and for **first person singular pronoun incidence (index 89)** with a decrease of 57%.

4.5 Summary and conclusion

This chapter began with a detailed description of the Coh-Metrix programme, offering a review of its individual indices and principal components. It then presented the main findings of a Coh-Metrix analysis of EAPCORP, offering possible explanations for the observed feature movements and considering their significance. Any tentative summary of these results clearly needs to be cognisant of the probabilities involved and the fact that the observed movements are not conclusive verification of change along any of the indices. However, there are certain features which have shown increases for a significant number of paired samples and these include mean sentence, syllable and word lengths, noun modification, agentless passives and hypernymy for nouns and verbs. Similarly, there is evidence of significant reduction in the use of first person pronouns, narrativity, polysemy for content words and overall readability. This leads to the tentative suggestion that there has been an increase in certain aspects of writing *complexity* and a decrease in personalisation offering evidence of a *more written and less spoken* style of writing. It also

appears that our (University of Birmingham EAP preessional) teaching programme is having a positive effect, especially as the Coh-Metrix index can quantify movement in certain features which are explicitly targeted by the EAP programme. The overall picture so far is not one of unqualified success as a close perusal of the results in Table 4.2 shows, but it is possible to discern real change and to attribute this at least partly to the efficacy of our teaching programme. The stage is now set for the next part of the study where a newly developed Multidimensional Analysis Tagger (Nini, 2014) is applied to the aggregated data sets. It is hoped that this triangulation of Wordsmith, Coh-Metrix and MAT will set the Coh-Metrix results in a wider context and confirm the initial premise of this study, that we, as teachers and course designers, can and do make a difference to students' progress in second language writing.

CHAPTER FIVE

MAT ANALYSIS OF EAPCORP

5.1 Introduction

This chapter firstly outlines the multidimensional approach to language analysis listing the large range of linguistic categories presented by Biber (1988) and then describes the adaptation and application of these features in the Multidimensional Analysis Tagger or MAT (Nini, 2014). It then presents the findings of a MAT analysis of EAPCORP together with (uncorrected) examples from the corpus in terms of both individual and dimensional changes together with discussion of possible implications.

5.2 A multidimensional approach

Multidimensional analysis (MDA) forms a central component of the study, and is an approach developed by Biber (1988) which has the basic premise that it is possible to identify register differences between different types of language by the assemblage and analysis of a large number of textual features which then can then be ascribed to certain dimensions. As discussed in Chapter Two, MDA uses computers to count the frequencies of linguistic features in tagged corpora, and then to carry out a factor analysis on these feature counts in order to identify ‘dimensions’, that is, sets of meaningful associations among individual variables.

The linguistic features typically studied in MDA are listed in Figure 5.1.

Figure 5.1: Multidimensional features (Biber, 1998, pp. 77-78)

<p>tense and aspect markers</p> <p>1 past tense</p> <p>2 perfect aspect</p> <p>3 present tense</p>
<p>place and time adverbials</p> <p>4 place adverbials (e.g. above, beside, outdoors)</p> <p>5 time adverbials (e.g. early, instantly, soon)</p>
<p>pronouns and pro verbs</p> <p>6 first person pronouns</p> <p>7 second person pronouns</p> <p>8 third person pronouns</p> <p>9 pronoun it</p> <p>10 demonstrative pronouns (that, this, these, those, those as pronouns)</p> <p>11 indefinite pronouns (e.g., anybody, nothing, someone)</p> <p>12 pro-verb do</p>
<p>questions</p> <p>13 direct wh questions</p>
<p>nominal forms</p> <p>14 nominalisations (ending in -tion, -ment, -ness, -ity)</p> <p>15 gerunds (participial forms functioning as nouns)</p> <p>16 total other nouns</p>
<p>passives</p> <p>17 agentless passives</p> <p>18 by- passives</p>
<p>stative forms</p> <p>19 be as main verb</p>

20 existential there

subordination features

21 that verb complements (e.g., I said that he went)

22 that adjective complements (e.g. I'm glad that you like it)

23 WH clauses (e.g., I believed what he told me)

24 infinitives

25 present participial clauses (e.g. Stuffing his mouth with cookies, Joe ran out the door)

26 past participial clauses (e.g., Built in a single week, the house would stand for fifty years)

27 past participial WHIZ deletion relatives (e.g., the solution produced by this process)

28 present participial WHIZ deletion relatives (e.g. the event causing this decline is)

29 that relative clauses on subject position (e.g., the dog that bit me)

30 that relative clauses on object position (e.g. the dog that I saw)

31 WH relatives on subject position (e.g. the man who likes popcorn)

32 WH relatives on object position (e.g. the man who Sally likes)

33 pied-piping relative clauses (e.g. the manner in which he was told)

34 sentence relatives (e.g. Bob likes fried mangoes, which is the most disgusting thing I've ever heard of)

35 causative adverbial subordinators (because)

36 concessive adverbial subordinators (although, though)

37 conditional adverbial subordinators (e.g. if, unless)

38 other adverbial subordinators (e.g. since, while, whereas)

prepositional phrases

39 total prepositional phrases

40 attributive adjectives (e.g. the big horse)

41 predictive adjectives (e.g. the horse is big)

42 total adverbs

lexical specificity

43 type/token ratio

44 mean word length

<p>lexical classes</p> <p>45 conjuncts (e.g., consequently, furthermore, however)</p> <p>46 downtoners (e.g., barely, nearly, slightly)</p> <p>47 hedges (e.g. at, about, something like, almost)</p> <p>48 amplifiers (e.g. absolutely, extremely, perfectly)</p> <p>49 emphatics (e.g., a lot, for sure, really)</p> <p>50 discourse particles (e.g., sentence initial well, now, anyway)</p> <p>51 demonstratives</p>
<p>modals</p> <p>52 possibility modals (can, may, might, could)</p> <p>53 necessity modals (ought, should, must)</p> <p>54 predictive modals (will, would, shall)</p>
<p>specialised verb classes</p> <p>55 public verbs (e.g. assert, declare, mention, say)</p> <p>56 private verbs (e.g. assume, believe, doubt, know)</p> <p>57 suasive verbs (e.g. command, insist, propose)</p> <p>58 seem and appear</p>
<p>reduced forms and dispreferred structures</p> <p>59 contractions</p> <p>60 subordinators that deletion (e.g. I think... he went)</p> <p>61 stranded prepositions (e.g. the candidate that I was thinking of)</p> <p>62 split infinitives (e.g. he wants to convincingly prove that)</p> <p>63 split auxiliaries (e.g. they are objectively shown to</p>
<p>coordination</p> <p>64 phrasal coordination (noun and noun; adj and adj, Verb and verb; adv and adv)</p> <p>65 independent clause coordination (clause initial and)</p>
<p>negation</p> <p>66 synthetic negations e.g. no answer is good enough for Jones</p> <p>67 analytic negations (e.g., that's not likely)</p>

These categories of linguistic description are then subjected to factor analysis which produces groups of statistically referenced co-occurrences for the categories. These are then narrowed to a smaller group of factors, producing *dimensions* which are qualitative descriptions of the factors. A basic premise here is that the multiplicity of linguistic features does not adequately describe the differences between speech and writing and that analysis of the clustering of features offers a better way of characterising the two modes.

The current study applies multidimensional analysis to a learner corpus, namely EAPCORP (see Chapter Three, section 3.7) which offers a specific body of data, compiled for the purposes of assessing second language writing development over a limited time frame. This corpus can then be analysed in terms of both individual features as described in Figure 5.1 and the dimensions which I describe in section 5.6. There are many advantages of using a multidimensional approach to assess learner development. Firstly, results are probably more generalisable than other more narrowly focussed approaches as they pertain to “patterns of register variation” (Biber *et al.*, 2003, p.152). In addition, they tend to be based on real, often large quantities of corpus data, a position which could possibly be summarised as “the more the better” i.e. the more evidence available, the more likely it is to be applicable to other contexts and the easier it is to apply statistical techniques to measure significance. Another advantage is that the sheer range of linguistic categories employed by MD analysis makes it possible to examine a large range of linguistic features without necessary ascription to a particular theory or a priori position. In other words, MD makes it easier to allow the data to guide the research and to help us to discover more about language. It also offers the possibility of adaptation to specific research contexts with not all the categories needing to be employed on any one occasion. The multidimensional approach can also be used to identify individual features which can then be quantified and compared

so that it is possible to examine a particular text, identify a feature, for example the incidence of agentless passives or average word length and compare it with another text. This enables an analysis of pre and post-writing samples which the current study presents as a key methodological technique.

5.3 The Multidimensional analysis tagger: MAT

The multidimensional analysis tagger (MAT) as the tagger manual indicates, "replicates Biber's (1998) tagger for the multidimensional functional analysis of English texts ... and generates a grammatically- annotated version of the corpus or text selected [and] the statistics needed to perform a text-type or genre analysis" (Nini, 2014, p.1). The MAT programme operates by employing the algorithms used by Biber (1988) and employs an adaption of the Stanford tagger (Toutanova, K. *et al.*, 2003) to analyse the text. The programme offers a range of analytical categories of which two are used in the present study, the individual variables and the dimensions.

5.3.1 The individual variables

The wide spread of the categories identified and harnessed into the MAT software programme are used to investigate which features of the writing samples are notably different pre and post-course. The main emphasis here is upon the individual features themselves and how they illustrate the second language developments shown in the paired samples. The variables are listed in Figure 5.2.

Figure 5.2. Multidimensional analysis variables

abbreviation	variable
AMP	amplifiers (absolutely, completely etc.)
ANDC	independent clause coordination
AWL	average word length
BEMA	be as main verb
BYPA	by passives
CAUS	causative adverbial subordinators (because)
CC	coordinating conjunctions
CD	cardinal numbers
CONC	concessive adverbial subordinators
COND	conditional adverbial subordinators (if/unless)
CONJ	conjuncts (instead, namely, moreover etc.)
CONT	contractions
DEMO	demonstratives
DEMP	demonstrative pronouns
DPAR	discourse particles (e.g. well, now)
DT	determiners
DWNT	downtoners (e.g. almost, partially, somewhat)
EMPH	emphatics (e.g. really, most)
EX	existential there
FPP1	Includes first person pronouns (I, me, my, myself, we, our, ourselves)
GER	gerunds
HDG	hedges (e.g. maybe, sort of etc.)
INPR	indefinite pronouns
JJ	attributive adjectives
LRB	left round bracket

NEMD	necessity modals (ought, should, must)
NOMZ	nominalisations (nouns ending in -tion, -ment, -ness, -ity)
NN	total other nouns (not gerunds or nominalisations)
OSUB	other adverbial subordinators
PASS	agentless passives
PASTP	past participle clauses
PEAS	perfect aspect
PHC	phrasal coordination
PIN	total prepositional phrases
PIRE	pied-piping relative clauses (e.g. the way in which she was informed)
PIT	pronoun <i>it</i>
PLACE	place adverbials
POMD	possibility modals (can, may might, could)
PRED	predicative adjectives
PRESP	present participial clauses
PRMD	predictive modals (will, would, shall)
PROD	pro-verb do (do used as a main verb)
PRIV	private verbs
PUBV	public verbs
QUAN	quantifiers
QUPR	quantifier pronouns
RB	total adverbs
RRB	right round bracket
SERE	sentence relatives
SMP	seem/appear
SNYE	synthetic negation
SPAU	split auxiliaries

SPIN	split infinitives
SPP2	second person pronouns (you, your, yourself etc.)
STPR	stranded prepositions
SUAV	suasive verbs
SYNE	synthetic negation
THAC	<i>that</i> adjective complements
THATD	subordinator <i>that</i> deletion
THVC	<i>that</i> verb complements
TIME	time adverbials
TO	infinitives
TOBJ	that relative clauses on object position
TPP3	third person pronouns (she, he, they, their etc.)
TSUB	that relative clauses on subject position
TTR	type-token ratio
VB	verb base forms
VBD	past tense
VBG	<i>ing</i> form of the verb
VBN	verb participle
VPRT	present tense
WHCL	WH clauses
WHOBJ	WH relative clauses on object position
WHQU	direct WH questions
WHSUB	relative clauses on subject position
WZPAST	past participial WHIZ deletion relatives (e.g. the solids created by this method)
WZPRES	present participial WHIZ deletion relatives (e.g. the rainfall causing this flood)
XXO	analytic negation
.	full stops

,	commas
”	speech marks
:	colons

These variables are scored by the MAT as incidence per 100 tokens (this contrasts with Biber’s 1988 frequency lists per genre which are measured out of 1000) and the only exceptions to this are AWL (average word length) and TTR (type token ratio). The reason for this 100 based scoring method is that the MAT was created using short texts, and percentages were considered to be more effective. There appear to be no real advantages or disadvantages with either scoring system.

5.4 The procedure

The first step was to apply the MAT to the data set and then record the results. Each of the data sets received a MAT results file which contained statistics and dimension information. The statistics file contained standardised incidence scores for each of the language features for each pre and post-cohort plus some modification (for example raw scores for AWL and TTR); the dimensions file contained the aggregated dimension scores again for each of the programmes pre and post for 20, 15, 10 and 6 week for both years 2010 and 2012. The next step was to record the individual feature scores on separate sheets. This was done manually. Each pair was matched and the features recorded for upward (+) or downward (-) movement. Table 5.1 illustrates the range of possible scores for each programme. For example, the 20 week 2010 had 9 students enrolled so a score of +9 describes a feature increase for all the 9 students on the programme and a score of -9 would describe a feature decrease for all students. An overall score of 0 would indicate that the plus and minus scores balanced out

and that there was no overall observable trend. Similarly, the 10 week programme 2012 for example, had 65 course participants so a score of +65 would indicate that all students had made an observable increase for a particular language feature and score of – 65 would indicate that all students had made a decrease for this feature. Again, a score of zero means that the scores balanced out and that there was no observable trend. The actual results as discussed subsequently were rarely as conclusive as either example offered, but the closer to the possible maximum or minimum score that each feature recorded, the more confident the claim concerning movement for each pair. This is illustrated in Table 5.2 where the movements across the score ranges are expressed as percentages.

Table 5.1 Score range for the programmes

programme	maximum score	minimum score
20 week 2010	+9	-9
15 week 2010	+27	-27
10 week 2010	+57	-57
6 week 2010	+35	-35
total	+128	-128
20 week 2012	+8	-8
15 week 2012	+26	-26
10 week 2012	+65	-65
6 week 2012	+36	-36
total	+135	-135
overall total	+263	-263

The next step was a qualitative interpretation of each feature with cognisance of the significance of upward, downward and absence of movement, in other words, increases, decreases and stasis. It is clear that some of the increases recorded are indicative of writing development and a good example of this would be average word length. This basic indicator expressed, as a simple mean may be obvious, but an increase especially one across the board would suggest that students are using words of a more academic character. Similarly, some *decreases* may also be indicative of writing development and examples of this are reduction in the use of first person pronouns or in the use of contractions. While some feature movements were relatively simple to interpret, others were more opaque, for example a reduction in the use of private verbs or a reduction in the use of independent clause coordination. When discussing the results in section 5.5, upward and downward movements are described without a priori classification and qualitative assessments, whether the movements indicate progression or otherwise, are based on evidence from grammar reference, especially Biber *et al.* (1999) and to some limited extent, experience as an EAP teacher. It was important when recording the data to avoid the temptation to look for confirmation and some of the results seemed to be pointing strongly in one direction only to be confounded by an apparent “rogue” finding. A good example of this is the finding for infinitives (TO) on the 10 week 2010 programme (Table 5.2). The cohort shows positive and upward movements for 7 out of 8 cohorts (+ 1 +3 +15 +12 +18 +3 and +3) apart from the 10 week 2010 with a score of -12 and this prompted a recheck of the data which in turn produced the same result. There was throughout the procedure an emphasis on elimination of confirmation bias, although small clerical errors are always a possibility. The individual dimension scores were retrieved and logged and this enabled a more detailed overview of the individual features and utilised the inherent capacity of MAT to illustrate correlations, clustering and to amplify language feature co-occurrences.

5.5 Results of MAT analysis

This section is divided into two main parts. Section 5.5.1 presents the findings of the quantitative analysis of individual features, and section 5.2 provides a qualitative discussion of the main groupings identified by the quantitative analysis.

5.5.1 Results of quantitative analysis

Table 5.2 shows the results for the MAT feature analysis for the whole data set. The first column lists each individual linguistic variable in the analysis. The next eight show the aggregate scores for each programme (20, 15, 10 and 6), and the last column shows totals for each year and for both years together. Taking **AWL** (average word length) as an example (highlighted in bold as it shows considerable feature movement): *every* course has shown an increase for this feature for *both* years. The totals are 109/128 for 2010 and 97/135 for 2012 and the overall total score spread is from +263 to – 263 (526 pairs). As can be seen, the overall **AWL** score is +**206** so the score percentage is: $263 + 206$ divided by 526 multiplied by 100 = + **89%**. An aggregate overall score of zero indicates no movement for the feature. An aggregate percentage of plus or minus 60% of the overall total of 263 pairs is selected as a threshold to highlight clear movement for a specific feature. This represents a minimum 10% increase or decrease and offers a baseline for consideration of movement and possible tractability of the specific feature. 10% itself can be regarded as a minimum and the discussion in 5.4.2 uses this figure although there is also discussion of slightly lower baseline percentages.

Table 5.2 MAT feature analysis- movements from the beginnings to the ends of the programmes.

+ increase - decrease 0 no change X no entries

Variables in **bold** show features which have shown movement of **60% and above** for **all (8/8) or 7/8 courses**

MAT variable	20/10	20/12	15/10	15/12	10/10	10/12	6/10	6/12	2010	2012
									Overall total	
AMP	+1	-2	-1	+4	-2	-8	+3	+12	+1	+6
									+7	
ANDC	+3	0	-13	-11	-9	-7	-5	-3	-24	-21
									-45	
AWL	+9	+8	+27	+19	+47	+50	+24	+20	+109	+97
									+206 (89%)	
BEMA	-5	-2	-2	-3	-12	-4	0	-1	-19	-10
									-29	
BYPA	+3	+2	0=	-3	+7	0	+1	+2-	+11	+1
									+12	
CAUS	0	-4	-5	-2	+5	-8	-4	12	-4	-2
									-6	
CC	-5	+4	-7	-1	-18	-5	-4	+1	-34	-1
									-35	
CD	0	-1	+4	-13	+7	+12	+5	-2	+16	0
									+16	
CONC	+3	+2	-3	+4	-1	-3	+2	-8	+1	-5
									-4	
COND	-1	-3	0	0	-5	-2	+4	+2	-2	-3
									-5	

CONJ	+5	+4	+7	+9	+4	+23	+19	+10	+35	+46	+81 (+65%)
CONT	-2	0	-7	-5	-16	-16	-16	-1	-41	-22	-63 (-62%)
DEMO	+3	+8	+9	+2	+15	+16	+8	+9	+35	+41	+76(+64%)
DEMP	-2	-2	+2	-8	-19	-17	-12	-8	-31	-35	-66 (-63%)
DPAR	x	x	0	+1	0	-1	+1	+2	+1	+2	+3
DT	-1	-4	+6	-1	+5	+24	+7	+1	+17	+20	+37
DWNT	-1	-1	-3	-10	+9	-10	+3	+6	+8	-15	-7
EMPH	+5	-3	-1	-7	-16	-16	-5	-19	-17	-45	-62 (-62%)
EX	-3	+2	-10	-9	-17	-16	+5	+3	-25	-20	-45
FPP1	-9	-7	-19	-23	-35	-46	-27	-7	-90	-84	-174 (-83%)
GER	-1	+2	+9	+2	+2	-11	+10	+4	+20	-3	+17
HDG	-1	x	-4	-5	+2	+9	-5	+3	-8	+7	-1
INPR	+1	x	-1	+4	x	+3	-1	+2	-1	+9	+8
JJ	+3	0	+6	+7	+18	+13	+6	-4	+43	+16	

									+59 (+61%)
LRB	0	+1	-3	-1	-1	-2	+2	-3	-2 -5 -7
NEMD	+2	0	-2	0	+2	-8	-5	+3	-3 -5 -8
NOMZ	+8	+7	+9	+11	+7	+12	+10	-9	+34 +21 +55 (+60%)
NN	-3	-2	+3	+1	+17	+14	-6	-3	+11 +10 +21
OSUB	+1	+2	+1	+2	+8	+6	-1	+2	+9 +12 +21
PASS	+8	+3	+10	+4	+21	+9	+7	+8	+46 +24 +70 (+63%)
PASTP	-1	x	+1	-2	+1	+1	+3	-3	+4 -4 0
PEAS	+1	-4	+6	-3	-1	+10	+7	+2	+13 +5 +18
PHC	+3	-2	+5	-1	+4	+16	+2	+8	+14 +21 +35
PIN	X	-1	+15	+6	+10	+17	+3	-1	+28 +21 +49
PIRE	X	x	+1	+1	+5	+4	-1	+3	+5 +8 +13
PIT	-1	x	-6	-2	+6	-11	-9	+15	-10 +2 -8
PLACE	+3	+2	+2	-2	+4	-5	-4	-7	+5 -12 -7

POMD	-2	0	-6	X	+5	+7	0	+7	-3	+14
									+11	
PRED	-3	+4	-9	+9	-16	+7	+4	+16	-24	+36
									+12	
PRESP	x	+1	+1	+3	+2	+4	+1	+1	+4	+9
									+13	
PROD	-1	x	+3	+3	-2	-6	-5	+1	-9	-2
									-11	
PRIV	-6	0	-20	-4	-25	-21	-2	-5	-53	-30
									-83 (-66%)	
PUBV	+2	-1	+1	+1	-1	-7	-3	+16	-1	+9
									+8	
QUAN	-1	+2	-5	-5	-5	-17	0	0	-11	-20
									-31	
QUPR	+1	+1	+1	-1	-7	0	+3	+4	-2	+6
									-4	
RB	+1	+6	-17	+7	-13	-17	+4	+5	-25	+1
									-24	
RRB	x	+1	-3	-1	-3	-2	+1	-3	-5	-5
									-10	
SERE	0	+1	+8	+2	+9	+12	+14	-2	+31	+13
									+44	
SMP	+1	+1	+9	-1	+5	+5	x	+2	+15	+7
									+22	
SNYE	x	-4	-4	+1	+2	-8	-7	+8	-9	-3
									-12	
SPAU	+2	+3	+5	+2	+7	+6	0	+3	+14	+14

									+28
SPIN	x	x	+1	x	x	x	x	-2	+1 -2 -1
SPP2	-5	-5	-1	+5	-4	-15	+3	-4	-7 -19 -26
STPR	+1	+1	+2	X	-4	-11	-6	-3	-9 -13 - 22
SUAV	x	+6	+10	+2	+2	+4	x	-5	+12 +7 +19
THAC	x	+1	+6	+1	+2	-3	+5	+9	+13 +8 +21
THATD	-3	0	-8	0	-13	-13	-11	0	-37 -13 -50
THVC	+2	+1	+6	+1	+2	-3	+5	+9	+15 +8 +23
TIME	x	0	-1	+2	+4	-8	-11	-5	-14 -11 -25
TO	+1	+3	15	+12	-12	+18	+3	+3	+7 +36 +43
TOBJ	+1	x	+4	+1	x	-2	-2	+3	+3 +2 +5
TPP3	+3	+1	+4	-4	+7	+8	0	+8	+14 +13 +27
TSUB	+1	x	X	X	-2	x	+1	1	0 -1 -1
TTR	+7	-2	+6	-2	+9	+11	+14	-10	+35 -2 +33

VB	+3	0	+5	+9	-13	-10	+7	+1	+2	0
									+2	
VBD	x	-2	-13	-7	-16	-30	-7	0	-36	-39
									-75 (-65%)	
VBG	-5	+1	-3	+9	+20	0	0	+5	+12	+15
									+27	
VBN	+5	-2	+11	+3	+25	+16	+6	+3	+47	+20
									+67 (+63%)	
VPRT	-1	-2	-7	-9	-20	-28	0	-19	-28	-58
									-86(-66%)	
WHCL	0	0	0=	-1	-3	+1	+4	-3	+1	-3
									-2	
WHOBJ	-1	-1	X	-2	+3	-3	+1	+1	+3	-5
									-2	
WHQU	0	-1	-1	+3	-1	0	0	+1	-2	+3
									+1	
WHSUB	+4	+2	+1	+4	-2	+7	-3	+1	-1	+14
									+13	
WZPAST	x	-1	-1	-3	+1	+7	0	-5	0	-2
									-2	
WZPRES	0	-1	+1	+2	-1	+7	+2	-2	+2	+6
									+8	
XXO	0	+6	-11	-2	-6	-16	-10	+5	-27	-15
									-42	
.	+1	0	-6	-1	-7	-23	-15	+1	-27	-23
									-50	
'	+3	0	+13	+3	+1	-1	+8	-7	+25	-5

									+20
"	0	-2	1	-2	-4	1	+2	0	-3 -3 -6
:	0	-1	+2	+6	-1	+4	-8	+4	-7 +13 +6

percentages to nearest whole number

Table 5.3 shows the movements recorded in Table 5.2 in graphic form with a twenty feature plus or minus range. There are two graphs, Table 5.3 (a) which illustrates the movements of features showing the highest increases and Table 5.3 (b) which shows the movements of features showing the highest decreases.

Table 5.3 (a) MAT feature analysis- movements from the beginnings to the ends of the programmes, twenty features which have shown the most increase.

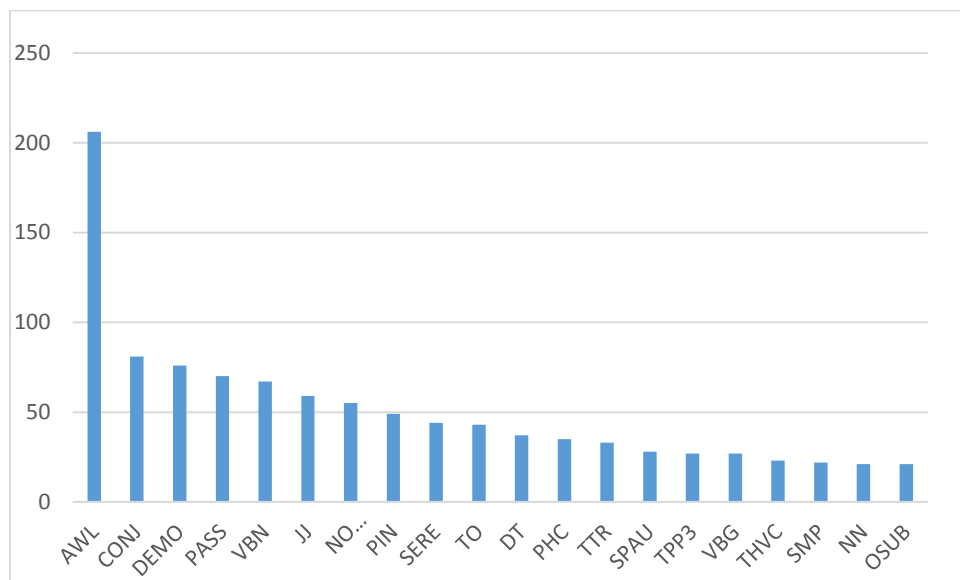
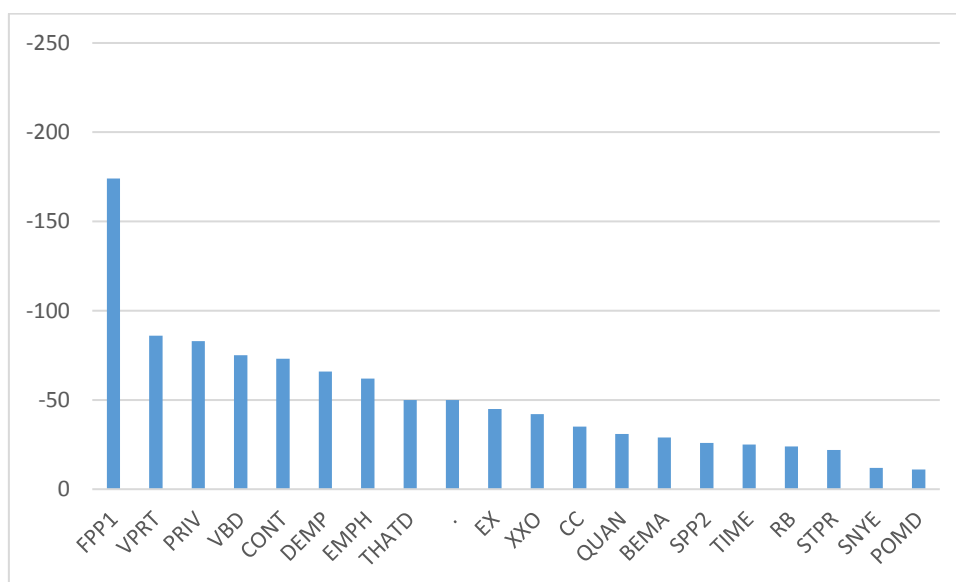


Table 5.3(b) MAT feature analysis- movements from the beginnings to the ends of the programmes, twenty features showing the most decrease.



5.5.2 Qualitative analysis of the MAT findings

This section presents and discusses the results of the MAT analysis. The main focus of the discussion will be on those variables which show movement for **all** or **7 out of 8** programmes (20, 15, 10 and 6 week courses). There is some additional comment on variables which have shown possible but less conclusive evidence of movement such as for 6 out of 8 programmes or for strong 10 week figures (65 pairs for 2010 and 2012 respectively). Language features showing contradictory finding are discussed in section 5.4.8. In addition, there are some features such as discourse particles (DPAR) and WH relative clauses on object position (WHOBJ) which were very infrequently recorded, sometimes appearing with only a few pairs per cohort and these are listed in section 5.5.2.8. The dimensional analysis of the MAT programme is not included in this section although there is tangential reference to it when

considering the move from narrative to expository writing style, which appears to be characteristic of many of the matched pairs. A fuller description and discussion of the dimensional analysis offered by the MAT will be provided in section 5.6.

5.5.2.1 MAT variables with significant increase (60% or above) for **all** programmes

Three variables were found to have increased significantly for all programmes: average word length, conjuncts and agentless passives. Each of these will be discussed in turn.

Average word length (AWL) which is calculated as a simple mean, shows a significant increase in all four cohorts for both years 2010 and 2012 and this is one of the clearest features of overall written language development so far suggested by the current research. This characteristic, which is also identified in both the pilot study and the Coh–Metrix analysis, suggests that students at the end of their preessional studies are producing texts which appear to be more written than oral in character than they were at the beginning of their preessional studies. Academic prose is likely to contain a higher proportion of polysyllabic nouns (the polysyllabic nature of many of the key academic terms is exemplified in the academic word list, Coxhead, 2000), and a reduced number of pronouns (Biber *et al.*, 1999, p.235) which tend to be shorter in length. One would hope and possibly anticipate that the exposure of students to a range of academic texts and related word development exercises, which are carried out as part of the general EAP programme and in subject specific classes might result in a greater use of academic terms and generally more academically related vocabulary.

Conjuncts (CONJ) for example instead, moreover, in comparison, in contrast, in any case)

show an increase across the board with scores of +5, +7, +4 and +19 and +4, +9, +23 and +10 for the 20, 15, 10 and 6 week 2010 and 2012 programmes respectively with the 6 week figures being quite noticeable. The lexical items which fall under this CONJ category are coordinating conjunctions and sentence connectors which are characteristic of sentence complexity whereas the coordinating conjunctions for example *and*, *but* which are coded separately as CC, tend to be viewed as linking clauses in a relatively simple, speech-like fashion. The feature scores for coordinating conjunctions (CC) show a decrease for six out of eight programmes and an overall reduction of – 35 (-57%)

Here are three examples of conjuncts being used.

In comparison to the Eastern country, the UK is a island country. (10 week post 2010)

In other words the main lecture are the main body [of] our course (10 week post 2010)

Another important thing, the UK and Jordan have totaly different culture ... in addition the population and size are widely different (20 week post 2010)

These examples are typical of much of the post course writing samples and this suggests that students are beginning to acquire a more characteristically academic style of writing. Biber (1999, p. 880) considers that such linking expressions are characteristic of academic prose and their use in supporting arguments “as a communicative need” is noticeable.

Agentless passives (PASS) are widely recognized as a feature of academic and technical writing (Biber *et al.*, 1999; Grant and Ginther; 2000; Tribble, 2002). In the current research, the frequency of agentless passives was found to increase in all eight cohorts, with

movements of +8, +10, +21 and +7 and +3, +4, +9 and +8 for the 20, 15, 10 and 6 week 2010 and 2012 programmes respectively. It is reasonable to suggest these movements may be linked to the effects of the courses of study undertaken by the students and the EAP programme explicitly teaches the many passive forms and aims to sensitise students to their use in essays, dissertations and theses. It appears that students are beginning to use the passive voice more widely in their free writing as illustrated in the following example, taken from the 6 week programme 2012.

Pre

In China you will go to the secondary school when you 12 years old. It's the beginning of the nightmare as well in somebody mind

Post

To begin with, as a student, the differences in the education system is really obvious. In China, competitive pressure must be faced from the primary school, even the kindergarden

The agentless passive is in evidence in the extract and is possibly an indicator of writing quality which can serve to quantify the impressionistic evaluation of the post sample as being *better* than the pre sample. The evidence for the prevalence of the various forms of the passive in academic prose and in contrast to conversation is considerable and as Biber *et al.* (1999, pp. 937-938) show, with the exception of short passives as a verb complement there is a higher incidence of all passive forms in academic writing and this may be as Biber *et al.* (1999) suggest because of the concern with generalisations characteristic of the register.

5.5.2.2 MAT variables with significant increase for **7 out of 8 programmes**

A further 3 variables were found to have increased significantly for all but one of the 8 programmes included in the present study: verb participles, nominalizations, and demonstratives.

Verb participle (VBN) figures show increases of +5, +11, +25 and +3 (2010) and -2, +3, +25 and +6 (2012). This may reflect increased use of the passive plus a greater awareness of more complex grammatical forms which are explicated and targeted by the preessional programme, such as using transitivity to express an opinion of an event (an example exercise being *the Chancellor increased taxes on North Sea oil/ taxes were increased in line with inflation/taxes increased by over 50%*).

Nominalisations (NOMZ) are classified by MAT as a vocabulary item, with nouns ending in -tion, -ment, -ness and -ity. The occurrence of these suffixes are the means by which the MAT tagger identifies the nominalisation process, and with noun phrases containing pre-modifiers being “three to four times more common in expository written registers than in conversation” (Biber *et al.* ,1999, p.589) this may represent a move from a verbal to a nominal style of expression. This increased nominalised characteristic is a feature of academic prose and Biber and Gray (2013, p. 100) give examples of this in scientific texts, as the following shows.

Now that programmed instruction has emerged from the laboratories of experimental psychology, **consideration** is being given to the **expansion** and **utilization** of the media

This feature, the nominal character of academic prose is taught explicitly on the EAP preessional programme and is a key component of our teaching curriculum. Here is an example from a post 6 week student in 2012:

*It is heatedly debated about the difference between China and the UK. Although the **globalization** of world culture leads to the differences between China and the UK becoming smaller, there are still huge differences existent*

Demonstratives (DEMO) Nini describes these features as found when the words *that, this, these, those* have not been tagged as either demonstrative pronouns DEMP (-66), *that* on object position TOBJ (+5), *that* on subject position, TSUB (-1), *that* adjective complements THAC (+ 21) or *that* verb complements THVC (+23) (Nini, 2014, p.19). They show an increase across all eight cohorts. Biber *et al.* (1999, p.350) find that when *this* and *these* are used as determiners, they are of relatively high frequency in academic prose “due to their use in marking immediate textual reference, for example ... on this account” and this offers a suggestion that students may be becoming more aware of the need for explicitness and possibly context independence in formal academic writing.

5.5.2.3 MAT variables with less clearly identifiable upward movement

This category covers those variables, of which there is only one, which have shown over 60% upward movement for six out of eight programmes

Attributive adjectives (JJ) is a classification whereby an adjective is followed by a noun or another adjective, modifying nominal expressions thus excluding predicative adjectives. The scores show increases for 6 out of 8 cohorts with the 10 week programmes showing increases of 18 and 13 (31% and 20% for 2010 and 2012 respectively). This increase in a basic linguistic resource for “expand[ing] and elaborate[ing] the information in a text” (Biber, 1988, p.237) may again be indicative of an increase in range of vocabulary and possibly related to a greater use of nominalisations and nouns in general. An example is offered here by Biber *et al.* (1999, p.510):

One of the most **important** [ways] of achieving this is by the **regular** and **thorough** [implementation] of **planned** disinfection [programmes] of all livestock units.

A (single) example is also presented from the EAPCORP (15 week 2010):

Hong Kong is a city where mix with **western** and **traditional Chinese** culture

5.5.2.4 MAT variables with significant decrease (60% or above) for all programmes

Two variables were found to have decreased significantly for all programmes: first person pronouns and private verbs. Both will be discussed in turn.

First person pronouns (FPP1) show a clear and significant decrease across all cohorts for both 2010 and 2012. This reduction, which in similar fashion to the increases in AWL is a characteristic of developed, written English (Biber, 1988; Biber, 2002), may well be

evidence of a less personalised writing style. Biber *et al.* (1999, p. 235) observe that “nouns are many times more common than pronouns in news and academic prose” and that “the informational content of news and academic prose results in a much more frequent use of nouns and proportionally many fewer pronouns”. The FPP1 count includes first person plural pronouns such as *me/my* and *we/our*, and the observed reductions need to include this wider classification; however, as the Wordsmith feature frequencies show (see Chapter Three, section 3.4) **I** is easily the most significant reduction.

From a wider perspective, however, there is some concern relating to the idea that it is always incorrect to use first person pronouns. For example, Basal and Bada (2012) concluded that pronouns could be used effectively in scientific texts. However, students often report that they have been told that use of first person pronouns represents bad practice. To a certain extent this complements the idea of objectivity in the academy where one of the characteristics of academic writing and thinking is to take away the subjective aspects. Nonetheless, there may well be a place for first person pronouns in academic writing. However, reductions in FPP1 use such as evidenced by these findings are indicative of a change in writing style. The fact that students are using fewer first person pronouns, does indicate that this style has changed over the course of their programme and it is again evidence that the talk to writing strand of development identified by Shaw and Liu (1998) is relevant. This finding complements the trends suggested by the Coh-Metrix (McNamara *et al.*, 2014) and the Wordsmith (Scott, 2008) analyses and it appears that students are becoming aware of the need to depersonalise their writing. Nothing has been proven, but this finding may well be a strong indicator of one area of second language writing development (see Chapter Six, section 6.3.1)

Here is a simple example of reduced use of this feature in the first two sentences of the free

writing of a 15 week student.

Pre

Since I've been here for just 3 days, I don't really figure out the UK look like. But I'll point out the perversely the different of the environment.

Post

There are several differences between the UK and Taiwan. Since the two different countries have their particular culture, ways of think and living style, it is obviously some important differences between these two country.

Private verbs (PRIV) are verbs which “express an intellectual state, often taking a that clause” (Chalker and Weiner, 1998) for example (forget, hear, notice, think, feel) and are sometimes referred to as verbs of cognition. As can be seen in Table 5.2, they show across the board reductions of -6, -20, -25 and -2 (2010) and 0, -4, -21 and -5 (2012). Private verbs such as think, feel are often “used for overt expressions of private attitudes, thoughts and emotions” (Biber, 1988, p.105) and their use, together with first person pronouns, tend to produce an interactive, verbal and less nominal style. Conversation tends to feature a higher proportion of these lexical verbs and they “tend to frame the personal stance of the speaker” (Biber *et al.*, 1999, p. 360). Thus, we might conclude that the observed reductions in these private verbs may suggest that students’ writing is beginning to bear characteristics of a more written, less spoken nature.

5.5.2.5 MAT variables with significant decrease (60% or above) for **7 out of 8 programmes**

A further 4 variables were found to have decreased significantly for all but one of the 8 programmes included in the present study: present tense, emphatics, contractions and demonstrative pronouns and each will be discussed in turn.

Present tense (VPRT) shows a 7 from 8 reduction with the 10 week figures looking quite robust (-20 and -28 for 2010 and 2012 respectively). This reduction in the use of the present tense may relate to the fact that the present tense refers to the immediate action (Biber, 1986) and the immediate rather than removed situations (Biber, 1988). The present tense itself is characteristic of both conversation and academic prose (Biber *et al.*, 1999, p.458) but for different reasons. In conversation, the present tense is used to concentrate on the immediate contexts (*I come from China*). In academic prose, the present tense is often used to express the truth or validity of a premise or proposition (*inflation leads to unemployment*) and it may be the case that students are reducing their uses of the present tense disproportionately, creating an overall reduction. It may also be the case that students are consciously incorporating other verb forms after initial use of a basic present tense, in other words students may be becoming more ambitious in their use of tenses and are experimenting with their writing. Another possible reason is that students are becoming more aware of the nominal character of academic writing and that this is reflected in writing of a more nominalised nature, thus reducing the use of present tense verbs.

Here is an example of two scripts produced by the same student at the beginning and end of a 15-week programme (2012)

Pre

*There **are** many differences between my country and the UK. First of all the weather of my country **is** more warmer than the UK. I **come** from China which **is** an Asia country. I **live** in east of China and there **is** not as much rain as that of the UK*

Post

*Different cultures between China and the UK have led to many differences. First of all, the way of communication **is** an obvious different part between these countries. Chinese people tends to express their feelings. Especially in China if the communication [is] with different social status, the lower one should speak more gentle and respect the one who **has** a higher status.*

The pre example is fairly typical of an initial 15 week EAPCORP essay with the first four sentences showing six occasions of use of the simple present tense in 51 total words. The post example appears more linguistically sophisticated with the first four sentences yielding 60 words and direct use of the present tense 2 times with one modified (tends to express) and one implied [is] use.

Emphatics (EMPH) including *just, really, most, more, for sure, a lot* show a reduction for 7 out of 8 cohorts and this may possibly be attributable to the acquisition of a more conventionally oriented writing style. The emphatics listed appear more characteristic of speech with conversation having relatively high frequencies of really, sort of and kind of (Biber *et al.*, 1999, p.867), so a reduction may be attributable to this lessening process, the students' writing being more informed by written conventions and of a less common, more specialised nature (Biber, 1988, p. 11).

Contractions (CONT) are reduced for 7 out of 8 cohorts and this may be interpreted as one of the least surprising results. It may be anticipated that the students are able to internalise the convention that contractions are used less commonly in academic writing and are more a feature of speech. Biber *et al.* (1999, p.1128) identify two broad categories of contraction, verb (*I'm reading*) and not (*She couldn't hear*) in addition there are structural contractions such as *Do y'know?* or *I'm gonna go*. These are all much more typical of conversation than academic prose and the University of Birmingham EAP programme explicitly teaches this characteristic with the convention being relatively easy to learn as a simple rule and probably requiring a minimum of cognitive processing.

Demonstrative pronouns (DEMP) for example those, these and this followed by a verb, auxiliary or modal show decreases for 7 out of 8 cohorts (-2, +2, -19, -12 and -2, -8, -17 and -8 for 2010 and 2012 respectively). Biber *et al.* (1999) indicate that demonstrative pronouns are far more common in conversation than in academic prose or other written registers so that a reduction in their use may suggest a greater facility with characteristically written form of register. As discussed subsequently in section 5.5.2 from a dimensional perspective, the feature bears a positive loading on factor 1 (Biber, 1988) as a characteristic of a more spoken style of writing, and reduction in the use of this feature may be an indicator of this more characteristically written style of production.

5.5.2.6 MAT variables with less identifiable downward movement

This category covers one variable which has shown over 60% downward movement for six out of eight programmes

Past tense (VBD) records a 6 out of 8 reduction x, -13, -16 and -2 for 2010 and -2, -7, -30 and 0 for 2012.

This reduction may be tentatively interpreted as a correlate to the reduction in present tense forms, if the texts produced by the students are of a more *nominal and less verbal character* then it may follow that both forms of tense may be reduced as a consequence. The register distribution of past and present tense for conversation and academic prose show similarities (Biber *et al.*, 1999, p.456) with both having a preference for present tense forms.

5.5.2.7 MAT variables with contradictory findings

This section considers those variables which show what could be termed a marked or conspicuously uneven profile, where a particular cohort displays a marked or disproportionate difference to the overall total. There are four of these variables.

CD cardinal numbers, show a reduction of -13 for the 15 week 2012 programme in comparison to an overall increase of +16

GER gerunds show a reduction of - 11 for the 10 week 2012 programme in comparison to an overall increase of +17.

PRED predicative adjectives show a reduction of - 16 for the 10 week 2010 programme in comparison to an overall increase of +12

TO infinitives show a reduction of -12 for the 10 week 2010 programme in comparison to an overall increase of +43.

It is not possible to offer an explanation for the above figures for which the disproportionate reductions in TO infinitives (see section 5.3) appear to be the most prominent as use of

infinitives is generally characteristic of academic prose (Biber *et al.*, 1999, pp.698-699) and the overall score of + 43 represents a 58% increase for the feature.

5.5.2.8 MAT variables which are infrequently recorded

Using the 10 week cohorts, these being the most quantitatively robust (57 and 65 pairs) as a benchmark, the features are listed which are recorded for fewer than 10 percent of the total, in other words where 9 out of 10 pairs record a blank space.

Discourse particles (DPAR) for example- well, now, anyhow

Hedges (HDG) for example- maybe, at about, something like

Indefinite pronouns (INPR) for example- anybody, everyone, nowhere

WH clauses (WHCL)

WH relatives on object position (WHOBJ)

Direct WH questions (WHQU)

*Foreign words (FW)

*Letters and numerals used to identify items in a list (LS)

*Exclamations (UH)

Most of the above features are relatively easy to explain and may be interpreted positively. For example, the absence of discourse particles, exclamations, and indefinite pronouns are characteristic of conversation and more informal writing such as personal letters, emails and Facebook entries. There is little surprise, therefore, that they are not apparent in either the pre or post essays. None of the students are beginners, so they all have an idea that academic writing requires a relatively formal style and the initial and continued absence of these features is welcome. The same applies to direct WH questions, the avoidance of which is also a curricular item in the EAP syllabus with foreign words and letters and numerals showing a sustained absence that would certainly be expected after an intensive period of EAP study. WH clauses (any public, private or suasive verb followed by a WH word followed by a word that is not an auxiliary- *he believed what I told him*) are more common in conversation in other registers (Biber *et al.*, 1999, p.688) so again their relative rarity both pre and post could be viewed as a positive result.

The low incidence of hedges appears to be of concern, especially as they could be considered characteristic of academic prose. However, there may be an issue in regards to the formation of the expression *maybe* (an informal expression) or *may be* (more formal) which the MAT tagger does not discriminate between. An AntConc analysis over the 2010 data shows 48 pre examples of *maybe* and 12 post examples, this is a large reduction and suggests a fall in informality and may also suggest that the MAT programme is a little blunt regarding this hedging characteristic.

5.6 Factors and dimensions

A consideration of the individual features described in sections 5.5.2 does offer a series of linguistic characteristics which may contribute to an understanding of second language writing development. Specific movements, whether of increase or decrease do offer the basis for much comment; however, there is so far no consideration of a pattern of *correlated features* to describe student writing at the beginnings and ends of their programmes. It is this correlation which may enable us to sketch in more broadly, the type of texts which students may typically produce. For this reason, the dimensions facility on the MAT programme has been used and observations made concerning movements along the various continua which Biber (1988) has isolated. As I hope becomes clear in discussion of the individual dimensions, it may be possible to make general observations about the evolution of student writing in terms of register similarities and differences.

5.6.1 EAPCORP MAT dimensions

The following Figures (5.3 (a) to 5.3(f) describe the six dimensions and list factor loadings, the related genres typical of each and the characteristics of the dimensions on a continuum.

Dimension one is interpreted to distinguish between texts bearing an informational focus and texts having an involved focus (Biber, 1988). High scores on this dimension are characterised by high frequencies of “private verbs, that deletions, present tense, contraction and second person pronouns... together with markedly infrequent occurrences of nouns,

prepositions, long words, more varied vocabulary and attributive adjectives”. (Biber, 1988, p. 129). As Biber (1988, p. 133) indicates, this is not necessarily a spoken to written distinction and is better understood as “the interpretation of involved real time production versus informational, edited production”

Figure 5.3 (a) Dimension one, *involved production versus informational production* (Biber, 1988, p.128)

mean	genre	dimension continuum
		involved production
35	face to face conversations	
20	personal letters	
	spontaneous speeches	
	interviews	
5	romantic fiction	
	prepared speeches	
0	mystery adventure fiction	
	general fiction	
	professional letters	
-5	science fiction	
	humour	
-10	editorials	
	biographies	
	press reviews	
-15	academic prose	
	official documents	
		informational production

Dimension two “distinguishes between narrative and non- narrative discourse” (Biber, 1988, p.142) with high scores “characterised by frequent occurrences of past tense and perfect aspect verbs, public verbs, present participial clauses and synthetic negation ... and markedly infrequent occurrences of present tense verbs and attributive adjectives.” (Biber, 1988, p.136)

Figure 5.3(b) Dimension two, *narrative versus non- narrative discourse* (Biber, 1988, p.136)

mean	genre	dimension continuum
		narrative
7	romantic fiction	
6	general fiction	
2	biographies	
	spontaneous speeches	
1	prepared speeches	
	personal letters	
0	face to face conversations	
-1	interviews	
-2	telephone conversations	
	academic prose	
-3	official documents	
		non-narrative

Dimension three “distinguishes between informational texts that mark referents in an elaborated and explicit manner and situated texts that depend on direct reference to or

extensive knowledge of, the physical and temporal situation of discourse production for understanding” (Biber, 1988, p.142) High scores on this dimension “are characterised by frequent occurrences of WH relative clauses, pied piping constructions, phrasal coordination and nominalisation” (Biber, 1988, p.142)

Figure 5.3(c) Dimension three, *explicit versus situation-dependent reference* (Biber ,1988, p.143)

mean	genre	dimension continuum
		explicit reference
7	official documents	
6	professional letters	
5	academic prose	
2	editorials/biographies	
0	prepared speeches	
-1	interviews	
-2	science fiction	
-4	personal letters	
-6	telephone conversations	
		situation dependent reference

Dimension four “distinguishes between persuasive and non- persuasive discourse” (Biber, 1988, p. 151) but the various genres are mainly undistinguished regarding this dimension. High scores may reflect use of prediction and possibility modals often used with a first person agent (Biber, 1988, p.148) with other features including necessity modals (e.g. you should go) and suasive verbs” with these features “simply marking the speakers’ persuasion”

(Biber, 1988, p.148)

Figure 5.3 (d) Dimension four , *overt expression of persuasion* (Biber, 1988, p.149)

mean	genre	dimension continuum
		overt persuasion high
4	professional letters	
3	editorials	
2	romantic fiction	
1	interviews/general fiction	
	telephone conversations	
	spontaneous speeches	
0	face to face conversations	
	academic prose	
	biographies	
-1	adventure fiction	
-2	press reviews	
		overt persuasion low

Dimension five “distinguishes between highly abstract, technical discourse and non- abstract types of discourse” with high scores reflecting “a focus on conceptual subject-matter” (Biber, 1988, p.154) and characterised by frequent use of conjuncts, agentless and by passives, past participial clauses, WHIZ deletions and certain types of adverbial subordination” (Biber, 1988, p.151)

Figure 5.3 (e) Dimension five, *abstract versus non- abstract information* (Biber, 1988, p. 152)

mean	genre	dimension continuum
		abstract information
5	academic prose	
4	official documents	
	press reviews	
0	biographies	
-2	spontaneous speech	
-3	romantic fiction	
	telephone conversations	
		non-abstract information

The co-occurrence patterns in dimension six indicate “a dimension marking informational elaboration under strict real time conditions” with “subordination features co-occurring with colloquial features such as final prepositions and demonstrative pronouns” (Biber, 1988, p.156). As can be seen, face to face conversations and academic prose score in the same range on this dimension.

Figure 5.3 (f) Dimension six(on-line) *informational elaboration* (Biber, 1988, p.155)

mean	genre	dimension continuum
		high information elaboration
3	prepared speeches	
	interviews	
2.5	spontaneous speeches	
1.5	professional letters	
0.5	face to face conversations academic prose	
0	biographies	
-0.5	telephone conversations	
-1	press reviews	
	personal letters	
-1.5	general fiction	
	science fiction	
		low information elaboration

5.6.2 Movement along the dimensions

The cohorts were analysed using the MAT dimension facility and the following table shows pre and post-sample movements along the dimensions. The aggregate scores are recorded for the pre and post-samples and decreases (-) or increases (+) are indicated.

Table 5.4 (a) scores and movements for 6 dimensions 2010

course	dim 1	dim 2	dim 3	dim 4	dim 5	dim 6
20 pre	0.70	-5.71	8.96	-6.07	5.77	-1.29
20 post	-9.79 -	-5.36 +	13.84 +	-4.65 +	7.74 +	0.08 +
15 pre	5.56	-4.65	5.92	-4.02	3.11	-1.21
15 post	-1.85 -	-4.94 -	7.49 +	-1.18 +	6.81+	-1.06 +
10 pre	1.48	-4.54	-2.23	-2.23	5.93	-0.48
10 post	-3.32-	-4.17 +	-2.90 -	-2.90 -	7.38 +	0.62 +
6 pre	4.68	-3.76	-2.11	-2.11	-1.40	-1.40
6 post	-1.26 -	-4.36 -	-2.41 -	-2.41 -	-0.80 +	-0.80 +
total	0/4 +	2/4 +	2/4 +	2/4 +	4/4 +	4/4 +

Table 5.4 (b) scores and movements for 6 dimensions 2012

course	dim 1	dim 2	dim 3	dim 4	dim 5	dim 6
20 pre	4.68	-4.11	5.10	-4.03	6.39	-2.00
20 post	-2.88 -	-5.88 -	5.13 +	-0.16 +	12.40 +	-1.12 +
15 pre	4.57	-4.50	4.99	-3.62	2.48	-1.28
15 post	1.54 -	-3.99 +	6.25+	-2.22 +	4.63 +	-1.45 -
10 pre	6.03	-4.69	4.88	-3.09	4.12	-1.09
10 post	-0.16 -	-4.80 -	6.46+	-2.34 +	7.37 +	-0.59 +
6 pre	0.72	-4.39	4.76	-2.00	5.94	-0.97
6 post	1.15 +	-3.07 +	5.30+	-1.74 +	8.12+	-0.31+
total	1/4 +	2/4 +	4/4 +	4/4 +	4/4 +	3/4 +

Table 5.4. (c) movements for 6 dimensions 2010 and 2012

course	dim 1	dim 2	dim 3	dim 4	dim 5	dim 6
total	-7/8	+4/8	+6/8	+6/8	+8/8	+7/8

5.6.3 Dimensions which have indicated movement for all (8/8) courses

All cohorts show increases along dimension five, *abstract versus non-abstract information* and the genres which are typical of relatively high levels of abstraction include academic prose and official documents while telephone and face to face conversation occupy the opposite end of the continuum. The genres with high scores on this dimension “make frequent use of conjuncts, agentless and by passives, past participial clauses, WHIZ deletions and certain types of adverbial subordination” (Biber, 1988, p.151). The results for the MAT *individual* feature scores in Tables 5.2 and 5.3 illustrate this finding, with conjuncts and passives showing clear increases.

In the following examples from a ten week student in 2010, despite the errors prevalent in both, the second sample appears to present the information with more elaboration and contains three instances of the passive.

2010 10 week pre

Second, Korea has alot different weather from UK. Due to geographay of two countries, Korea rains alot in the summer but in the other seasons, we don't see that much rain on the other hand UK has lots of rain through

whole year, but rain just drizzle all the time. Also Korea has extreme weather in summer it goes up to the 35 or 37 degree Celsius but in the UK weather doesn't go extreme.

2010 10 week post

*Secondly I could find the mountains around Birmingham. as a matter of fact, given that Korean peninsular is **composed** of 79% mountain areas and 30% normal lands, this fact **can be considered** as one of the major differences, for examples I have been to 'Lickey Hills country park', which **is located** around southern part of Birmingham, steep hills couldn't be found around that area*

5.6.4 Dimensions which have indicated movement for 7 out of 8 cohorts

Seven out of eight of the programmes show reductions along the continuum for dimension one, *involved production versus vs informational production*. The dimension itself, “a powerful factor [which represents] a very basic dimension of variation among spoken and written texts in English” (Biber, 1988, p.104), offers a possible criterion for evidence of movement from a spoken to a more written form of writing. This is elaborated by Nini (2014, p. 5): “Low scores on this variable (1) indicate that the text is informationally dense, as for example, academic prose, whereas high scores indicate that the text is affective and interactional, as for example a casual conversation. A high score on this dimension means that the text presents many verbs and pronouns (among other features) whereas a low score on this dimension means that the text presents (among other features) many nouns, long words and adjectives”

The results along this dimension, apart from the 6 week 2012, all show a reduction in overall scores with a range from +1.15 (6 week 2012) to -9.79 (10 week 2010). On this

dimension, the increases in word length, nouns, type token ratios, and agentless passives, all negatively loaded features on the involved/ informational axis, would appear to indicate that the students' writing is becoming less personally involved and more informationally dense. It is worth emphasising the important features which comprise factor one and underlie dimension one not least because in combination they produce the most powerfully weighted dimension in the set. In other words, dimension one has significance for the whole data set, even if the generic closest text types are less clearly defined. "Factor 1 represents a dimension which marks high informational density and exact informational content versus affective, interactional and generalised content...I propose the interpretive label 'informational versus involved production' for the dimension underlying this factor". (Biber, 1988, p.107). In referring to the centrality of this dimension of linguistic variation between texts, Biber (1988) cites several studies which relate to 'discoursal dichotomies', nominal versus verbal (Wells, 1960) and oral versus literate (Tannen, 1982; Tannen, 1985) and the findings so far on this dimension do appear to suggest that students' writing is becoming less involved and more informationally productive. Tables 5.4 (a) (b) and (c) show a small but clear movement in scores towards the informationally productive end of the spectrum.

Here is an example which illustrates to some extent, the differences between involved and informational production. The two examples are the beginning of pre and post essays produced by the same student on the 10 week programme 2012 with the samples being of similar length. The pre sample is 60 words long and contains 7 sentences with four pronouns and one adjective. The post sample is 55 words long, has two sentences with one first person pronoun and four adjectives. The average word length for the pre sample is 4.31 and for the post sample is 4.67.

Pre

My name is Zho Hu Zheng. I come from China. In my opinion, there is quite difference between China and the UK. The main is culture. First is the language. In China we speak Chinese every day, And in UK people speak British English. Second the food is difference, Chinese food is much more complicate to cook than British food.*

*not the student's real name

Post

I'm from Chinese, for my own experience, the main difference between China and the UK is the education system, especially the mark system. China's education system mainly focus on what the students got from the books and how many knowledge the students have and ignore where and how the students get the knowledge.

Seven out of eight programmes also show movement along dimension six, which is termed *on-line informational elaboration*. The high factor loadings on this dimension include the following: that clauses as verb complements, that relative clauses on object position, that clauses as adjective complements (demonstrative pronouns, wh relative clauses on object position). The post score range is from + 0.62 (10 week 2010) to - 1.45 (10 week 2012) and the scores, which as can be seen from Tables 5.4 (a) and 5.4 (b) group around the zero, indicate slight but probably not statistically significant movement towards informational elaboration. As Biber (1988, p.113) points out, there are several features on this dimension which suggest “that they function to mark informational elaboration in relatively *unplanned* types of discourse” (my italics) and that some of the features with high factor loadings may mark “informational elaboration that is produced under strict real-time constraints resulting

in a fragmented presentation of information” (Biber, 1988, p.113). This finding echoes the results of a study by Nesi and Gardner (2012) who found that their native speaker students’ writing became “increasingly informationally elaborated as they progress through their degree programmes” (Nesi and Gardner, 2012, p.14).

5.6.5 Dimensions which have shown movement for 6 out of 8 cohorts

Six out of eight programmes show reductions along dimension four, *overt expression of persuasion*. The post scores on this dimension are low with a range of -1.18 (15 week 2010) to - 4.65 (20 week 2010) and this may reflect the fact that the prompt question itself, “*What are the differences between your country and the UK*” may not necessarily elicit writing typical of a persuasive type essay. The low overt persuasion scores may reflect the combination of low suasive verb incidence and reduced use of first person pronouns. As can be seen from Figure 5.4(d), academic prose typically scores around zero on the continuum and the figures for the cohorts show a 6/8 increase in the post scores from a relatively low base. Here is a pair example from the 6 week programme in 2010, taken from the first two sentences (produced by the same student) which may illustrate a general awareness of the importance of lower commitment and a need to depersonalise.

Pre

In my opinion, there are lots of differences between my country and the UK. I come from China.

Post

Cultural distinction is the main difference between China and the UK. Cultural distinction leads to different eating habits, dressing style and greeting customs.

Six out of eight programmes showed increases along dimension three, *explicit versus situation dependent reference*. High scores on this dimension indicate that the text is independent of context and is characterised amongst other features by frequent occurrences of nominalisations, phrasal coordinations and WH relative clauses (Biber, 1988) with low scores indicating text dependence. There is a range of post scores from + 5.13 (20 week 12) to + 13.84 (20 week 10) and the overall differences are relatively slight. Explicit prose is a positive quality in academic writing and this observed increase can therefore be interpreted as a positive (if not always successfully achieved) outcome of the explicit instruction provided by the preessional programme. Here is an example of a student's free writing from the 6 week programme 2012 and there is a degree of explicitness in evidence in the second sample which also appears to be less situationally dependent with the highlighted nominalisation emphasising this.

Pre

Secondly, the UK have good urban planning that makes everything convenient for citizens to live. There are varies of accommodation around the university. The night life is so amazing.

Post

*Secondly, it cannot be denied that the economy of the UK is more developed than Vietnam. In the UK, **the application** of high technology in daily life, as well as production is widespread and at an advanced level.*

5.6.6 Dimensions which have shown movement for 4 out of 8 cohorts

The remaining dimension to consider is that of *narrative versus non-narrative discourse* (dimension two) with 4 out of 8 cohorts showing increase with a post score range of minus 3.07 to minus 5.8. This suggests that narrativity is generally low overall and that although there is no real decrease, the scores are located in a low position on the dimension scale. High scores on this dimension are typified by high frequencies of past tense and perfect aspect verbs, third person pronouns, public verbs, present participial clauses and synthetic negation (Biber, 1988, p.137). These results may suggest that students enter their programmes already aware to certain extent that story style writing is inappropriate in an academic context and this is to an extent reinforced by the EAP course. There are a few students whose target postgraduate subject is history related and for whom narrative is important but overall the results for this dimension are encouraging.

5.7 Summary and conclusions

This chapter has described the multidimensional approach originally developed by Biber (1988) and adapted into a research instrument, the Multidimensional analysis tagger (MAT) by Nini (2014). It then outlined how the MAT was used to analyse the EAPCORP in terms of its individual features and dimensions and presented these findings together with discussion of their possible explanations and consideration of their significance.

The corpus data examined in this chapter appears to suggest that the students' written production is marked by movement along the *talk to writing* strand suggested by Shaw and Liu (1988). In other words, the data suggests that one characteristic of developing student

second language writing, is text of a more recognisably “written” and to a certain degree more complex character. This is evidenced by the MAT programme analysis which has shown that in terms of individual features there has been an increase in use of longer words, passives, nominalisations, verb participles, sentence relatives, demonstratives, prepositional phrases and attributive adjectives. There has been a decrease in first and second person pronouns, private verbs, use of the present tense, use of the past tense and fewer emphatics. In terms of dimensional characteristics, this movement from talk to writing is further evidenced by the observed shift from non-abstract to abstract information which characterises all eight programmes (2010 and 2012) and the change from involved to informational production, from high overt to low overt persuasion and from low to high information elaboration which characterise seven out of eight of the programmes.

CHAPTER SIX

SUMMARY AND SYNTHESIS OF FINDINGS

6.1 Introduction

This chapter attempts to draw together the main results of the two analyses reported in detail in the previous two chapters, which have addressed the key research question of the current study, namely to examine the possible improvements in the written English of students on short intensive English for Academic Purposes courses at a British university. It will be argued that, when taken together, the results of the foregoing empirical analyses do appear to show significant linguistic changes over the course of an intensive EAP programme and that this can best be described as movement from a talk oriented to a more written oriented style of written production.

6.2 Considering and synchronising the evidence

The general theoretical framework of writing development employed in this study has been the threefold (simple to complex, spoken to written and inaccurate to accurate) typology suggested by Shaw and Liu (1998), as introduced in Chapter Two. In discussion of the findings, the evidence is broadly considered from these three positions of writing progression from simple to complex, from spoken to written and from inaccurate to accurate. These general categories which also form the subheadings for description and comment on the findings have been useful in framing conceptions of writing development throughout the study. They do overlap to certain extent, for example it is possible, to refer to a feature such

as increased word length as an example of increased complexity and of being more characteristic of written than oral production. As will be seen, the talk to writing strand is easier to support by empirical evidence as it can be explicitly referenced to research and is presented first in section 6.3. The simple to complex perspective, which is discussed in section 6.4, is more difficult to directly evidence, not least because the category may bear an assumption that simple is inferior and complex is superior, a consideration which may not be true in all cases. The following example shows the first four sentences from an initial essay with the shortness of the sentences suggesting a degree of simplicity, however there is use of the present perfect.

(15 week student 2012 pre)

I come from China. I have been in my country for twenty four years. And the UK is my favourite country. I learn this two countries well.

In addition, the grammar of academic writing is often said to be ‘simple’ at clause level (Halliday, 1985). Take this example from an article in *Journal of Neuroscience* and where the complexity lies in the noun phrases.

*To reconcile these results, Warren et al hypothesised that **different neuronal ensembles encode operant reward and extinction memories** (Warren et al., 2016).*

Another example could be the use of nominalisation which is certainly more characteristic of writing than speech (Biber et al., 1999, p.589) and nominalised writing is a feature of

academic prose (Biber and Gray, 2013) but it could also be said to mark an increase in complexity. The two strands, therefore of simple to complex and spoken to written can be viewed as offering a degree of complementarity which can accommodate the differing metrics and feature focus used by the Coh-Metrix and the MAT programmes. For this reason, the emphasis in this chapter is upon the empirical evidence for student development along the *talk to writing stand*, which as will be seen covers most of the observed changes in linguistic characteristics identified by the research instruments.

The general observation needs also to be made that there were in fact very few areas of *direct contradiction* between the findings of Coh-Metrix and MAT but because the specific indices are often different, it may be necessary to make comment on the *complementarity* or otherwise of the results. One issue which arises is related to the scope of the two techniques, with Coh-Metrix being a composite programme which is designed primarily to measure textual cohesion and so contains measures which are often applicable beyond word or phrase level. The MAT programme by contrast is concerned with specific parts of speech and the dimensional aspects are a separate component. As a result, it is often not possible to make direct correlations between for example, syntactic complexity at sentence level and to relate the findings to each other. This can cause problems when attempting to establish congruence and areas of agreement or disagreement between the two research tools. For example, the MAT focus on specific language items such as emphatics (e.g. really, most, more, for sure) and amplifiers (extremely, absolutely, very, totally) represents one clear difference from the codification system of Coh-Metrix (discussed in Chapter Four) and as a result, the linguistic emphasis differs somewhat. There are features such as word length and agentless passives incidence which are recorded by both programmes and there are many overlaps which may present a congruence between the two and a characterisation of the two approaches might emphasise the stress on cohesion in Coh-Metrix and the identification of specific linguistic

features by MAT.

A closer look, however at the principal components of MAT can partially address this issue, with descriptive (1) connectives (6) syntactic pattern density (9) and word information (10) components (approximately 40 descriptive elements) relatively easy to match with MAT categories. The two main research instruments, Coh-Metrix and MAT could therefore be described as providing complementary but not exactly matching analyses of the corpus data. In other words, they offer different areas of focus, occasionally identical but often different in several aspects. The establishment of areas where they are both in broad agreement is the main emphasis of this section.

The tables used (6.2 to 6.8) are for illustrative purposes and use aggregated figures for all programmes with detailed course by course results suggested for reference in Chapters four and five. In order to compare the results of the Coh-Metrix and MAT programmes, simple percentage figures are used which are rounded up to the nearest whole number. The pair information in Table 6.1 serves to illustrate this:

Table 6.1 EAPCORP Matching pairs per course

Programme	number of matching pairs
20 week 2010	9
15 week 2010	27
10 week 2010	57
6 week 2010	35
Total	128
20 week 2012	8

15 week 2012	26
10 week 2012	65
6 week 2012	36
Total	135
overall total	263

Score range for each year 2010 maximum +128 minimum -128

2012 maximum +135 minimum -135

A 2010 score of + 72 indicates that scores increased for $72/128/256 = +78\%$ of pairs

A 2010 score of -72 indicates that scores decreased for $72/128/256 = -78\%$ of pairs

6.3 A spoken to written perspective

This section focuses on the production of a less spoken and more recognisably written style and if we look at the findings in summary, this does appear to be the case. A very broad characterisation of “typical” speech could be direct reference to the listener (second person pronouns, questions, imperatives) and typical concern with the expression of thoughts and feelings, (emphatics, amplifiers, cognitive verbs such as think and feel) “In terms of its linguistic characteristics, stereotypical speech is structurally simple, fragmented, concrete and depended on exophoric (situation dependent) reference (Biber, 1988, p.47). Typical writing by contrast can be characterised as “more structurally complex than speech with for example longer sentences and more subordination, it is more explicit, more decontextualised, less personally involved and more abstract, and more deliberately

organised” (Biber, 1988, p.47) and as Biber points out, these characterisations are often in need of elaboration and not uniformly accepted, but may be useful as a general framework for conceptualising writing development. The summary is presented at word, sentence and beyond sentence levels, taking into account inevitable overlaps and focuses upon feature movements suggesting significant agreement for 8/8 or 7/8 courses of study (20,15,10 and 6 week programmes) for *both* years (2010 and 2012).

6.3.1 At word and phrase level

The first feature grouping for both Coh-Metrix and MAT concerns individual word information and a characteristic which suggests consistent advance is in one basic lexical area namely the *lengths of words and syllables* with Coh-Metrix indices 8 and 10, word length number of syllables and word length number of letters respectively showing firm increases. These two features offer evidence of the use of longer, more polysyllabic words and this may well be expected if students do internalise the academic vocabulary which is often of this polysyllabic nature (Biber *et al.*, 2002) to which they have been exposed over the course of an intensive writing programme. The related Coh-Metrix indices which show relatively high increases in standard deviations (see 4.5.1) would suggest that the development profile here is uneven- there is lot of individual variety, but the overall movement certainly appears to be forward. An example from the EAPCORP is offered here as an illustration:

Student c17 2010 10 week pre

The main difference between China and the UK is ...

Student c17 2010 10 week post

The obvious discrepancy between UK and China is that ...

Another simple indicator of positive transition between a speech-informed to a writing-informed style of written production could be a reduction in the use of contractions which the MAT identifies by recording apostrophes preceding a tagged word **or** use of *n't*. The MAT shows a mean reduction in this feature of 62% Emphatics, (for example, just, really, most, more, real/so+ adjective, for sure, a lot) which are more likely to be a feature of speech than writing, (Biber *et al.*, 1999, p.867) also show 62% reductions. The increases in hypernymy indicated by the Coh-Metrix (+64%, 57% and 64.5%) may be related to the development of subject-specific vocabulary (see section 4.5.1) and the decrease in polysemy (-58%) which tends to complement this finding may as suggested in section 4.5.2 be a result of the reduced number of higher frequency words which tend to have multiple meanings, again suggesting the development of a more specialised vocabulary. The increase in prepositional phrases which is included here (59.5%+) may also indicate movement from a spoken to a more written form of production and as Biber *et al.* (1999, p.606) observe, prepositional phrases as post modifiers "... [are] relatively rare in conversation and extremely common in academic prose" The increase in agentless passives may be a possible indicator of a more complex, academically informed writing style widely recognized as a feature of academic and technical writing (Biber *et al.*, 1999; Grant and Ginther, 2000; Tribble, 2002) and for a discussion see section 5.5.2.1. Conjuncts and demonstratives indicate upward movement with Biber *et al.* (1999, p.880) considering that such expressions are characteristic of academic discourse.

(In Tables 6.2-6.11 figures over 59% are shown in **bold** to highlight the extent of movement with feature movements of slightly below 60%, included to show complementarity. Blank spaces indicate no metric)

Table 6.2 Talk to writing feature movement at word and phrase level (1)

feature	2010 CM	2010 MAT	2012 CM	2012 MAT	mean CM	mean MAT
word length no of letters	+ 72 +78%	+107 +92%	+ 45 +67%	+97 +86%	+ 72.5	+89%
word length no of syllables	+69 + 77%		+ 35 + 62%		+69.5	
contractions		-41 -66%		-22 -58%		-62%
emphatics		-17 -57%		-45 -67%		-62%
polysemy	-30 -62%		-11 -54%		-58%	
hypernymy (nouns)	+21 +58%		+54 +70%		+64%	
hypernymy (verbs)	+21 +58%		+15 +56%		+57%	
hypernymy (nouns and verbs)	+35 +64%		+40 +65%		+64.5%	
prepositional phrases	+28 +61%	+28 +61%	+7 +53%	+21 +58%	+58%	+59.5%
agentless passives		+46 +68%		+24 +59%		+63.5%

conjuncts		+35 +64%		+46 +67%		+65.5%
demo*		+35 +64%		+41 +65%		+64.5%

*the words that, these, those, this when not tagged as DEMP, TOBJ, TSUB, THAC or THVC

Another feature grouping which presents amongst the most striking findings of the study relates to the use of personal pronouns indicating that that students' writing has become *less personal*. This is evidenced by the Coh-Metrix and the MAT with both in broad agreement offering *evidence of considerably reduced levels of personal pronouns, especially first person pronouns*.

Table 6.3 Talk to writing feature movement at word and phrase level (2) *personal pronouns*

feature	2010 CM	2010 MAT	2012 CM	2012 MAT	mean CM	mean MAT
first person pronouns		-90 -85%		-83 -81%		-83%
first person singular pronouns	-86 -84%		-55 -70%		-77%	
first person plural pronouns	-54 -71%		-38 -64%		-67.5%	
personal pronouns	-33 -63%		-18 -57%		-60%	

*the MAT tagger counts first person pronouns (I/me/myself/us/we/our/ourselves) as one category

These findings, namely a reduction in personalisation suggest strongly that students are acquiring a range of alternatives to presenting their message from a narrow personal position, even in the context of a relatively simple expository essay. The question, *What are the differences between your country and the UK?* might be expected to invoke a lot of opinion, but the evidence suggests that this is done in a less personal form. The simple metrics listed in Table 6.3 would appear to offer fairly strong evidence of *depersonalisation*. The Coh- Metrix and the MAT as the two main research instruments appear to be providing complementary evidence regarding a reduction in personalisation. Biber *et al.* (1999, p. 333) observe that “With the exception of we/us, forms which refer to the speaker and the addressee (I/me, you) are far more common in conversation than in other registers” and in terms of distribution of personal pronouns, the occurrence of I is recorded at 37,000 times per million in conversation, compared to 2,000 times per million in academic prose (Biber *et al.*, 1999, p.334).

Table 6.4 Talk to writing feature movement at word and phrase level (3) *noun orientation*

feature	2010 CM	2010 MAT	2012 CM	2012 MAT	mean CM	mean MAT
noun incidence	+23 +59%		+23 +59%		+59%	
modifiers per noun phrase	+35 +64%		+36 +64%		+64%	
determiners		+17 +57%		+20 +57%		+57%
nominalisations		+34 +63%		+21 +58%		+60.5%
left embeddedness		+39 +65%		+13 +55%		+60%
attributive adjectives		+43 +67%		+16 +56%		+61.5%

Table 6.4 relates to the production of *noun oriented writing* and there are a number of relevant findings. Noun incidence shows an increase of 59% for both 2010 and 2012 and the rise may reflect an awareness of the nominal nature of academic writing. Biber *et al.* (1999) show that nouns occur approximately twice as frequently in academic writing than in conversation and make the following observation that “In conversation, the shared situation and personal involvement of the participants result in a dense use of pronouns. In contrast, the informational purposes of news and academic prose result in a much more frequent use of nouns and proportionally many fewer pronouns.” (Biber *et al.*, 1999, p.235) The 64% increase in modifiers per noun phrase again may serve to emphasise a possible re-focusing on the part of the students towards the production of noun-oriented writing with noun

phrases containing modifiers being relatively unusual in conversation and common in academic writing with Biber *et al.* (1999, p.578) pointing out that almost 60% of all noun phrases in academic writing have modification, either pre, post or both. The relatively modest rise in determiners (+57%) recorded by the MAT complements this finding and as Biber *et al.* (1999) observe here, the differences between choice of a pronoun or a noun phrase may characterise the type of production produced by students pre and post instruction.

“In choosing between a pronoun and a noun phrase with determiner, the speaker/writer can take into account the degree of precision that is required or desired ... Conversation is embedded in a situation shared by the speaker and addressee, it can therefore be less specific ... As a result, conversation is characterised by a very dense use of pronouns. A writer on the other hand must make sure that sufficient specification is given with the text which leads to a high frequency of full noun phrases”. (Biber *et al.*, 1999, p. 284). The increase in left embeddedness, the mean number of words before the main verb, may well be anticipated if the noun phrases contain more modification and this result appears complementary to the other findings in this grouping.

The rise in attributive adjectives can again be viewed as complementary to the overall increase in nouns (Biber *et al.*, 1999, p.506) recording a frequency of 15,000 per million for conversation and over 60,000 per million for academic prose and attributive adjectives are “... one of the primary mechanisms used to pack information into noun phrases”. Nominalisations also indicate upward movement (see discussion in section 5.5.2.2) and their occurrence is characteristic of academic prose (Biber *et al.*, 1999; Biber and Gray, 2013).

It is worth reiterating at this point, that the EAP programme which the students had just completed, offers an explicitly targeted syllabus content related to noun orientation with separate sections on the production of all the above items especially regarding expansion of

the noun phrase and the general procedure of nominalisation (these issues are further discussed in Chapter Seven). Another feature grouping which focuses on verbs and tenses is shown in Table 6.5.

Table 6.5 Talk to writing feature movement at word and phrase level (5) verbs and tenses

feature	2010 CM	2010 MAT	2012 CM	2012 MAT	mean CM	mean MAT
present tense	-28 -61%		-56 -71%		-66%	
past tense	-36 -64%		-39 -64%		-64%	
private verbs	-53 -71%		-30 -61%		-66%	

The reductions in both present and past tense forms may be tentatively interpreted to indicate that students are producing writing of a more nominal and less verbal character with both forms being reduced as a consequence. The register distribution of past and present tense for conversation and academic prose show similarities (Biber *et al.*, 1999, p.456) with both having a preference for present tense forms. (see section 5.5.2.6)

The decrease in private verbs (for example, think, know, want, learn, remember) may be related to the fact that verbs themselves are reduced overall with conversation having a higher frequency than academic prose (Biber *et al.*, 1999, p.66). Another possible interpretation is that private verbs, referred to as mental verbs (Biber *et al.*, 1999, p.366) are more common in conversation, and the question, *What are the main differences between your country and the UK?* may tend to elicit private opinions with heavy use of *think, know,*

and *want* which are amongst the commonest lexical verbs (Biber *et al.*, 1999, p.373) and less common in conversation than in academic prose (Biber *et al.*, 1999, p.375) so a reduction in use of these private verbs such as *I think, I want* and may correlate to the observed reductions in personal pronouns, marking a shift to a less verbal style of writing.

6.3.2 At sentence level

At sentence level, the first and most obvious indicator is the *average sentence length* with a word being defined as a tagged part of speech (Charniak, 2000) and this largely descriptive measure may have a bearing on sentence complexity (see 4.5.1) but as discussed in 6.1, longer does not necessarily mean more complex. However, the production of a more noun-oriented type of writing with its attendant increases in number of words before the main verb, noun incidence, modifiers per noun phrase, determiners, nominalisations and attributive adjectives, all features illustrated in Table 6.4, might be expected to increase the average length of sentences by the modest degree indicated by the evidence. In other words, many of the features which have contributed to the production of more noun-oriented writing, most of which are taught on the EAP programmes, may have resulted in slightly longer sentences. Table 6.6 illustrates this:

Table 6.6 Talk to writing feature movement (6) sentence length and sentence complexity

feature	2010 CM	2010 MAT	2012 CM	2012 MAT	mean CM	mean MAT
average sentence length	+34 +63%		+11 +54%		+58.5%	
minimal edit distance parts of speech	-30 -62%		-21 -58%		-60%	
minimal edit distance words	-29 -62%		-6 -52%		-57%	
minimal edit distance lemmas	-43 -67%		-8 -53%		-60.5%	
hypernymy*	+21 +58%		+54 +70%		+64%	
sentence relatives**		+31 +62%		+13 +55%		+58.5%

*see also section 6.2.1

** identified as a punctuation mark followed by *which*

As Table 6.6 shows, three related indicators identify a reduction in *minimal edit distance* (parts of speech, words and lemmas) which is the “average minimal edit or the distance that parts of speech, words or lemmas are from one another between consecutive sentences in the text” (McNamara *et al.*, 2014, p.70). These findings are unclear but the observed reductions may point to a greater degree of cohesion between the parts of speech as represented in the texts and might be related, as suggested in 4.5.2 to the increased subject specific vocabulary that the students may have available at the end of their language programmes. Regarding the reduction in lemmas, the greater degree of *hypernymy* may have

the effect of reducing the edit distance. In discussion of relativisers (that, which, who, whom, whose, where, when, why), Biber (1999, p.611) identifies *which* as the commonest form in academic prose, so the increase in *sentence relatives* (punctuation mark followed by which) may indicate a greater awareness of this significant feature in academic writing.

6.3.3 Beyond sentence level

One characteristic which appears at discourse level is the reduction in narrativity, shown in Table 6.7 and these mean reductions of 67.5 and 65.5 percent (z score and percentile respectively) which are quite substantial may be related to the degree of text familiarity (see 4.3.1). “High narrativity reflects the use of more familiar words combined with a tendency to focus on events and characters rather than objects and ideas” (McNamara *et al.*, 2014, p.89). As discussed in 4.5.2, these reductions may indicate that students are becoming to an extent, more aware of the importance of the characteristics of academic rather than story style writing.

Table 6.7 Talk to writing feature movement (7) narrativity Coh-Metrix

feature	2010 CM	2010 MAT	2012 CM	2012 MAT	mean CM	mean MAT
narrativity z score	-61 -74%		-29 -61%		-67.5%	
narrativity percentile	-53 -70%		-29 -61%		-65.5%	

The MAT results offer complementary but not substantive findings with movement along

the narrative to non-narrative dimension for five out of eight programmes (see Figure 5.3 (b)) with high scores typified by high frequencies of past tense and perfect aspect verbs, third person pronouns, public verbs, present participial clauses and synthetic negation (Biber, 1988, p.137) and as discussed in 5.6.7 this suggests that degrees of narrativity begin at low levels and decrease to some extent.

The MAT programme also describes aggregate feature movement along the specified dimensions described by Biber (1988). The evidence offered by a dimension shift from non-abstract to abstract information production which characterises all eight programmes (20, 15, 10 and 6 week 2010 and 2012) and movement along three dimensions, involved to informational production, high overt to low overt persuasion and low to high information elaboration which characterise seven out eight programmes, could be said to mark a shift from a spoken to more characteristically written style of production (see discussion in 5.6.2, 5.6.3 and 5.6.4) The aggregate shifts are illustrated in Table 6.8 which is a reproduction of Table 5.4(c). In dimension one for example, -7/8 means decrease in dimension scores for seven out of eight programmes. Similarly, for dimension 2, +4/8 means increase for four out of eight programmes.

Table 6.8 Talk to writing dimension movement for 2010 and 2012

course	dim 1	dim 2	dim 3	dim 4	dim 5	dim 6
total	-7/8	+4/8	+6/8	+6/8	+8/8	+7/8

6.4 A simple to complex perspective: beyond sentence-level feature movement

The simple to complex perspective may be value-laden to a certain extent, implying that complexity is necessarily superior to simplicity (see discussion in 6.1). However, as a means of conceptualising second language writing development, and for the purposes of this study, a simple to complex perspective may well be useful in framing consideration of the reductions in *readability* presented by the Coh-Metrix. As Table 6.9 shows, there is movement in two indices of the readability principal component, the Coh-Metrix readability and the Flesch-Kincaid grade level indicator.

The Coh Metrix L2 readability index which is constituted by reference to content word overlap, sentence syntactic similarity and word frequency where lower scores are indicative of lower readability and which indicates that students' work is becoming more difficult to read, may possibly also be indicative of increasing levels of complexity, (see discussion in Chapter Four). The related Flesch-Kincaid grade level index which presents higher scores as indicative of reading difficulty supports the position to some extent and Table 6.9 illustrates this.

Table 6.9 Simple to complex feature movement, (1) *readability*

feature	2010 CM	2010 MAT	2012 CM	2012 MAT	mean CM	mean MAT
Flesch-Kincaid grade level	+59 +72%		+3 +51%		+61.5%	
Coh-Metrix readability	-30 -62%		-22 -58%		-60%	

This reduction in readability is also supported by decreases in familiarity for content words and in polysemy for nouns and verbs together with increases in hypernymy for nouns and verbs as discussed in section 6.2.2. This may collectively suggest that the students' work is becoming increasingly complex and possibly that the process of experimentation on the part of the developing writer is taking precedence over accuracy with this being reflected in lower readability scores. As can be seen, there is no metric for readability on the MAT programme but at beyond sentence level, the MAT dimension indices are in support of this position, namely that there is some evidence of an increase in complexity in students' written production.

6.5 An inaccurate to accurate perspective.

The pilot study (Chapter Three) attempted to address the issue of accuracy in students' writing development and the investigation was confined to a fairly rudimentary error count, using the sentence as a unit within which errors were manually logged to produce an error-free unit metric. This in itself, although rudimentary in form, did offer a basic outline for assessing the improvements or otherwise in certain aspects of the students' levels of accuracy, a term which in this context refers very much to spelling and punctuation. One of the reasons for this relatively restricted definition of accuracy is that it is easy to measure, at least with an initially small, 20 or 30 pair corpus. The Coh-Metrix and MAT programmes contain few indices which relate to direct error measurement, however the MAT programme does offer some information regarding *punctuation*, in the form of a reduction in the number of full stops and this could possibly be related to the increase in mean length of sentences.

Table 6.10 All course feature movement full stops

feature	2010 CM	2010 MAT	2012 CM	2012 MAT	mean CM	mean MAT
full stops		-27 -61%		-23 -59%	-60%	

Other measures of error incidence are not directly included in either the Coh-Metrix or MAT programmes and although it is possible, for example, to manually log errors for a 300,000 word corpus, it is in practical terms, a separate venture and as suggested in 7.6 a subject for further research, possibly using more sophisticated error counts and error analysis techniques. Regarding other measures of accuracy such as singular/plural agreement and accurate use of articles and tenses, they are not included in either the Coh-Metrix or the MAT analyses. The issue of student writing progressing from inaccurate to accurate then, is largely unaddressed and it has not been possible to follow up the promising initial findings of the pilot study which suggested that error counts were reduced and that students' written work was becoming more accurate over the course of a very few weeks.

6.6 Inconclusive and contrary findings

As discussed in 2.4.1, there are many related indicators which may be applied to the lexical profiles of second language learners and while there is no measure which offers a simple comparator, the type token ratio may offer a basic indication of writing development. The findings of this study regarding lexical density are inconclusive and as can be seen, the Coh-Metrix offers three separate TTR indicators and the results suggest small reductions in the TTR ratios with a noticeably larger decrease in the 2010 figures. The MAT which uses a standardised 400 token TTR, suggests a possible increase although again, the figures do not present evidence of clear movement. Also, Biber *et al.* (1999, pp. 53-54) in regarding the TTR as a relatively “crude measure”, indicate that TTR is lower in conversation than all written registers and that academic prose has the second lowest TTR possibly because “a great deal of academic prose has a restricted technical vocabulary” (Biber *et al.*, 1999, p.54). As a result, we may interpret these findings as too statistically slight (only the 2010 results as recorded by the MAT suggest any significant movement) to be conclusive.

Table 6.11 Lexical density indicators

feature	2010	2010	2012	2012	Mean	mean
	CM	MAT	CM	MAT	CM%	MAT
type token ratio content word lemmas	-22		-1		-54.5	
correlated with text length	-59%		50%			
TTR all words correlated with text	-22		-6			
length	-59%		-52%		-55.5	
TTR (first 400 tokens)		+35		-2		
		+64%		-51%		+57%

6.7 Summary

This chapter has attempted to synthesise the results of the Coh-Metrix and MAT analyses, the results of which have suggested a movement in observed linguistic characteristics from a talk to writing perspective, observable at the end of a programme of EAP instruction. At word and phrase level, these features include increased word and syllable lengths and increased incidences of agentless passives, conjuncts and demonstratives. There is evidence of decreased personalisation, supported by reductions in personal pronouns and an increase in noun orientation evidenced by rises in modification, nominalisation, left embeddedness and attributive adjectives. At sentence level, there is evidence of an increase in hypernymy and a decrease in minimal edit distance. Narrative is also reduced and four of the dimensions identified by Biber (1988), non-abstract to abstract, involved to informational production, high overt to low overt persuasion and low to high informational elaboration indicate the production of a more academic style of writing. In acquiring more characteristically written styles of prose, students appear to be internalising many of the explicitly taught features of the English for Academic Purposes programme at the University of Birmingham and can be said to have benefited from their course of study. The wider theoretical and practical implications of this are discussed in the subsequent and final chapter of this thesis.

CHAPTER SEVEN

IMPLICATIONS OF THE STUDY

7.1 Introduction

The principal aim of this thesis has been to investigate whether it is possible for students to make measurable improvements in their written academic English during the short time span of a summer preessional EAP programme. The evidence of the research reported in the preceding chapters provides strong support for answering this question in the affirmative. There is clear evidence of movement from a characteristically spoken to a more typically written form of production, characterised by decreased personalisation, increased noun orientation and more abstract, more informational and more elaborated writing. This is in striking contrast to the mainstream view in much of second language acquisition theory, which generally holds that improvement is slow and gradual. In the remainder of this concluding chapter, I will consider the practical pedagogic implications of the main findings of the study from a variety of broad and overlapping perspectives which relate to classroom practice, assessment, syllabus design and the EAP profession as a whole.

7.2 Implications for syllabus and materials design

Firstly, an area where the results of the study may be helpful is in the explicit awareness and use of the research instruments and their related feature classifications. Both the Coh-Metrix and the MAT programmes are readily available and relatively accessible to teachers through a simple web interface. They are also free and require only a password for access (Wordsmith

Tools is also highly accessible although users do need to pay a fee). This ease of access may serve to promote an active interest in linguistic enquiry and add another intrinsically interesting perspective for students and teachers alike. Charles (2014), for example, has shown that it is possible to get students to use corpus software in the classroom and to compile personal corpora using research articles, as a resource to inform their writing practice for the duration of their programmes of study. In a similar spirit, Lee and Swales (2006) have described corpus investigations conducted by their students using their own purpose built corpora.

Regarding classroom practice, and considering what could be termed a *language awareness* perspective, one implication is that the framework suggested by Shaw and Liu (1998) could be usefully taught to students as a means of framing their own perceptions of writing development. Students could be taught, for example, that the acquisition of academic writing capacity from simple language features to complex ones, is not only desirable but also attainable and that they, the learners, are on a progression continuum where they will be able to acquire the ability to write at a higher level of academic quality. It can also be firmly suggested that that their writing will become *more written in character and less like speech* with an increasing range of words and flexibility of production. Their writing can be expected to become, if not more accurate, then more appropriate to the context, especially the academic context within which they will be operating once their main academic subject programmes commence. These broad positions can be elaborated by consideration of the ways in which the study can inform aspects of pedagogic practice, one of which may be the teaching of vocabulary, a possibly understated aspect of language instruction, in the EAP classroom. The related issues here are legion, for example, from the advocacy of teaching word lists through to targeted substitution exercises and many other aspects of text related classroom practice. If as the study suggests, personal pronouns are highly tractable to

instructed acquisition, then their explicit inclusion in a syllabus may be justified as an effective, teachable and workable characteristic of academic writing.

While the current study has no direct bearing on the means of delivery practised in the classroom, the case for teaching an academically focused, word, phrase, sentence and discourse level EAP programme may have been strengthened by the results. With the information at hand as to which specific features appear tractable, it would be entirely possible to construct a syllabus and course book which focuses upon these characteristics. This has to a certain extent already been implemented on the University of Birmingham EAP preessional programme, which has evolved since its original inception and new research can be expected to continue to inform the programme in the future. Here are 10 simple examples (there are many others possible) of exercises based on the results of the current study, moving broadly from word to sentence to whole text level:

- Appropriate academic vocabulary substitution

- Increasing the length of the head nouns

- Verb substitution exercises changing active into passive as appropriate

- Relative pronoun substitutions

- Sentence length extension using relative pronouns

- Sentence length extension using conjunctions

- Nominalisation exercises: transforming verb oriented to noun oriented texts
- A rewrite of a first person pronoun dominated text.
- A rewrite of a narrative text in a more academic style
- A rewrite of a speech oriented to a written oriented text

The area of syllabus design has a vast literature of its own, however, and any curricular developments prompted by this research would clearly need to take into account a range of issues that cannot be considered here, such as learner levels, target courses, age profiles and countries of origin. Nevertheless, an assessment of the pedagogic tractability of EAP writing features as suggested by the current study can only be beneficial as a starting point for curriculum development activities in second language oriented writing development programmes. Having identified areas of written language production which appear to have shown development and can be considered tractable, the next step might be to make decisions about what could be included in an EAP writing syllabus based on this information. In other words, the study has identified language features which might be included in an EAP writing syllabus because they have shown improvement over a limited time frame and are worthy of attention. Also in terms of syllabus review and evaluation there may be features or characteristics of academic writing that *should be* taught as the research results indicating previously unnoticed gaps that may need to be filled.

A central premise of the approach used on the University of Birmingham EAP

programme is that the features of academic writing, i.e. those characteristics which distinguish it from other forms of written and (spoken texts) *need to be taught*. There is a strong view among the designers of the programme that improvements in written production are unlikely to occur by chance or indirectly through exposure to subject-based content, and that it is considered necessary to teach, for example, nominalisation, depersonalisation and syntactic flexibility for the simple reason that these characteristics are otherwise unlikely to improve. The current University of Birmingham preessional programme does incorporate some of these features, often by means of specifically targeted exercises. For example, the programme teaches noun orientation with exercises including *identifying the head noun in a noun combination*, *turning noun combinations into long sentences* and *inventing noun combinations*. There is also a strong emphasis on the importance of nominalisations; for example, *changing clauses with transitive and intransitive verbs into noun combinations*. Academic vocabulary is extensively taught and students practice *learning new and common and rare, words both general and academic* and *identifying word classes* and *practicing word extension*. In terms of recognising spoken to written style, the Birmingham programme emphasises movement from *inappropriate to appropriate register* and *informal to formal style*. The programme also aims to foster in students an awareness of active to passive differences and an ability to produce passive sentences accurately and appropriately. The curriculum does not specifically include any input on adjectives, however, although these are often included in activities related to the development of noun phrases. There has also been no explicit practice in the reduction of first person pronouns and these are both areas which could be emphasised and developed with a relevant series of language activities.

Another implication for syllabus design might be the explicit foregrounding of areas of linguistic focus, as the following possible syllabus outline may illustrate.

- Unit one (exercises 1- to 10) academic vocabulary and *using long words*
- Unit two (exercises 11- 20) *expanding the noun phrase*
- Unit three (exercises 21 – 30) using relative pronouns and conjunctions
- Unit four (exercises 31- 40) *verb flexibility- active and passive*
- Unit five (exercises 41-50) *depersonalisation and objectivity*
- Unit six (exercises 51- 60) *nominalisation and noun orientation*
- Unit seven (exercises 61-70) writing academic texts
- Unit eight (exercises 71-80) writing academic texts (control of outside sources)
- Unit nine (exercises 81-90) sentence and overall text coherence
- Unit 10 (exercises 91-100) creating your own academic voice

The last syllabus feature in this list is strongly emphasised on the University of Birmingham EAP preessional programme and would be augmented by consideration of the specific features identified in the current study.

Another possible area of focus lies in the explicit teaching of the dimensions identified by Biber (1988). Postgraduate students, operating as they are at a higher level of academic study, might be expected to be receptive to theory especially as it impacts directly upon their acquisition of English and their learning profiles in general. Nesi and Gardner (2012) identify 13 generic families of student assignments: case study, critique, design specification, empathy writing, essay, exercise, explanation, literature survey, methodology recount, narrative recount, problem question, proposal, research report proposal, critique or essay (Nesi and Gardner, 2012, p.34) and map the dimensions upon each of them, demonstrating the importance of register for each type of student activity. This approach could be applied to the design of a postgraduate English for Academic Purposes curriculum;

indeed, the University of Birmingham EAP programme already includes a quite overt component of linguistic theory to label sections in the teaching programme, as the following example from the University of Birmingham EAP course preessional course materials book shows:

Text structure, general to specific

Task 19: identifying general to specific sequences

Task 20: re-ordering information into a general to specific sequence

A multidimensional framework could easily be included in a syllabus using a selection of Biber's (1988) dimensions as follows:

Involved production versus informational production

Task x: changing shorter monosyllabic words to longer polysyllabic ones

Task y: packing information into noun phrases

Explicit versus situation-dependent reference

Task x use of agentless passives

Task y using relative clauses

Abstract versus non- abstract information

Task x using abstract nouns

Task y using conjuncts

Informational elaboration

Task x nominalisations with transitive and intransitive clauses

Task y extending the noun phrase using adjectives

The register demands of the various assignments could then be taught as appropriate together with the overall concept of student writing progression towards increasing informational and elaborational production (Nesi and Gardner 2012).

7.3 Implications for assessment

One application of the current study relates to the reporting of students' *progress*, a currently somewhat under-researched topic, but one which carries significance both in terms of student language progression and also awareness of personal linguistic development. As an example, there may be an opportunity to replace nebulous and clichéd comments such as "has shown improvement" with more explicit evaluations such as "is now producing writing of a more academic, less personal style" and this may help to augment the battery of information available to teachers and report writers in general. Another area of application is in teacher-led individual assessment, based on the writing that a student has produced and after analysis using the Coh-Metrix or MAT programmes. In other words, a teacher could take a sample of their students' writing and make individual assessment, diagnosis or simply comment, for example, in a tutorial on progression in an individual or range of features. This could help teachers assess the effectiveness of their teaching of a specific item and aid students in understanding their own linguistic development.

There are also several implications for the formation of marking criteria. As

suggested in Section 2.2, the allocation of marks and grades to students' work may have a tendency to be overly subjective, this being to an extent inevitable given the imprecision inherent in many marking schemes. To this extent, the current study may help to add an element of clarity to impressionistic marking descriptors. In one sense, much English language marking assessment has an in-built precision deficit, often attempting to measure and quantify features which cannot easily be quantified and while there is probably a degree of subjectivity in many marking schemes including those in specific subject disciplines, there may still be a need for greater precision when constructing criteria for English language assessments.

Here are some examples of possible descriptors which could be constructed on the basis of the current study:

- shows evidence of nominalisation and production of noun orientated writing
- uses an increasing range of academic vocabulary and longer, more polysyllabic words
- shows evidence of depersonalisation by reduced pronoun incidence
- writing is characterised by a less narrative style

The marking descriptors highlighted in Figure 2.2 could *incorporate* these features at various band levels (as could other briefer or less explicit descriptors at a variety of language levels). For example, a band one (highest level) descriptor could now read:

Range (sentence structure, word choice and cohesion)

A wide range of sentence structures and word choice. The message can be followed effortlessly, and

*cohesive devices within and/or between sentences are skilfully managed so that no attention is attracted. **Writing shows evidence of acquisition of a noun oriented style***

At the bottom end, occasional slips or minor infelicities are tolerable, but there is nothing revealing serious ignorance.

Similarly, a band 5 (lowest level) descriptor for the University of Birmingham's EAP final assessed essay marking scheme could now read:

Range (sentence structure, word choice and cohesion)

*Inadequate range of sentence structures and word choice. The message is difficult to follow and cohesive devices are inadequate or missing Inaccuracies in sentence construction predominate and the writer's inadequate syntax mostly obscures meaning. A limited vocabulary is used and **is more typical of the written rather than the spoken form of English.***

Another set of descriptors could be applied to the production of free writing as follows:

Marking criteria for pre and post free writing samples

Band 1 - marks 16-20 (for each descriptor)

Range

*A wide range of sentence structures and word choice. There is likely to be good **evidence of extended noun phrases, nominalisations and passives together with a selection of vocabulary characteristic of academic writing. The sample is of a completely impersonal, objective and academic character.***

Accuracy

Standards of grammar, word choice, word formation, spelling and punctuation are consistently of a high level.

Band 2 – marks 12-15

Range

*A fairly wide range of sentence structures and word choice. **There is likely to be evidence of extended***

noun phrases, nominalisations, and passives together with a selection of vocabulary characteristic of academic writing. The sample is of an impersonal, objective and academic character.

Accuracy

Standards of grammar, word choice, word formation, spelling and punctuation are of high level.

Band 3 – marks 8-11

Range

A range of sentence structures and word choice. There is likely to be evidence of extended noun phrases, nominalisations and passives together with a selection of vocabulary characteristic of academic writing. The sample is of a generally impersonal, objective and academic character.

Accuracy

Standards of grammar, word choice, word formation, spelling and punctuation are reasonable.

Band 4- marks 4-7

Range

A limited range of sentence structures and word choice. There is likely to be little evidence of extended noun phrases, nominalisations and passives or of vocabulary characteristic of academic writing. The sample is generally personal in character.

Accuracy

Standards of grammar, word choice, word formation, spelling and punctuation are generally inconsistent.

Band 5- marks 0-3

Range

Inadequate range of sentence structures and word choice. Little or no evidence of extended noun phrases, nominalisations and passives or of vocabulary characteristic of academic writing. The sample is entirely personal in character

Accuracy

Frequent errors of grammar, word formation, word choice spelling and punctuation cause severe strain for the reader. More is wrong than right.

The above marking scheme was piloted for the 2016 University of Birmingham preessional

programme, where it was decided to add a free writing element to the programme assessment which was taken by the students after completion of their course and submission of their final research paper. This was implemented for two main reasons. Firstly, the recurrent issue of plagiarism, while minimal over the whole international student cohort (no more than 10 from around 500 students altogether) was still present in a small number of cases, with the most intractable problems being the production of essays which were clearly written by someone other than the student themselves, either as a favour or even as a result of purchase. A free writing sample, it was felt, offered an independent means of verification in terms of the likely authenticity of the 3,000 word research paper in addition to affording another opportunity for the student to demonstrate writing competence. The other reason for the free writing element applied to the whole student cohort; specifically, to act as a progress marker that would be visible to both student and teacher alike.

7.4 Implications for the EAP profession

The current study may bear wider implications for the EAP profession, for associated stakeholders such as institutions of higher education and their related teacher administrative and ancillary staff, and even for the local economies. Such stakeholders would in all probability be pleased to find that the courses available to international students at their universities and in their locales are offering the opportunity for effective language improvement and sound, well designed programmes of study. This research may therefore, if in a limited way, offer substance to the claim that students are indeed benefiting from their programmes, if only as an effective use of time. It may also suggest that well-constructed, research-informed programmes are of benefit to the community as a whole, with course

providers, whether privately or publicly funded, needing to take current research into language development into account when designing and implementing their courses. Failure to do this may render the programme on offer insufficient and of unproven quality. This *quality assurance* element of the thesis may be considered a bold claim, but in the current economic and political climate, with the centrality of English language provision, both to the national economy of the UK and to the UK's status as a provider of high grade English language programmes, it is a prescient issue. At the very least, it is hoped that the current research will have a bearing on the way that course providers conceive of and implement their EAP provision and that an organisation such as a teaching unit, school or faculty for example, given the responsibility of improving international students' writing competence, can in fact do so.

The bodies tasked with EAP programme quality assurance are, as far as the UK is concerned, the British Council and BALEAP (the British Association of Lecturers in English for Academic Purposes; see Chapter One for a discussion of some of the roles of these two organisations). Both institutions could possibly take cognisance of the current study's identification of salient measurable development characteristics. Certainly, both bodies express interest in and invite consideration of current research findings as an element in their evaluation of institutions seeking accreditation. An example of this is within the BALEAP competency framework (BALEAP, 2008):

Competencies relating to curriculum development 8. Syllabus and Programme Development

An EAP teacher will understand the main types of language syllabus and will be able to transform a syllabus into a programme that addresses students' needs in the academic context within which the EAP course is located.

Another example is from the British Council accreditation programme 2015 self-assessment template inspection criteria (British Council Accreditation Handbook, *course design and implementation* (T12), which requires that:

Course design will be based on stated principles. There will be a coherent and appropriate course structure described in writing for teacher's guidance

The syllabus element of the British Council accreditation scheme is emphasised to a lesser extent than the BALEAP programme, as might be expected given the explicit EAP focus of this organisation; however, the British Council regularly inspects university EAP programmes and the findings of the current study may be useful in addressing the course design issues raised by the two examples.

This quality assurance element is also emphasised by the following criterion taken from inspection criteria teaching and learning *-knowledge* (T 23)

Teachers will demonstrate sound knowledge and awareness of the use of English and the linguistic systems underlying it and will provide appropriate models of both spoken and written English

The current study, which has identified linguistic resources that contribute to the production of a more recognisably academic style of writing, emphasises this difference and also provides a justification for a *learning outcome* which requires teachers to be aware of the purposes of their classes rather than simply providing a list of items to be covered in the class, as this criterion taken from the programme's teaching and learning- *planning* document (T25) shows:

Lessons will be based on a coherent sequence of activities leading to relevant learning outcomes

It is relatively easy to provide inspectors and classroom observers with an inventory of items to be covered in a class but the criterion above asks teachers to be aware of the purposes of teaching them and if the syllabus is predicated upon the development of academic style writing, then the learning outcomes are clear and preparation for inspection and classroom observation may be enhanced. A further implication of the current research relates to perceptions of the EAP profession particularly by international students who (in my own extensive personal experience as an EAP course co-ordinator, at least) seem to be increasingly aware of their importance as educational ‘consumers’. As indicated in Chapter One, there are a growing number of options available for study in the UK and abroad, and if students perceive that they are being used as ‘cash cows’, that is, as a financial resource of considerable proportions, they may be increasingly inclined to select ‘prestige’ institutions which have an active research profile and which are cutting edge in terms of research informed teaching, or at least be more attracted to universities who can empirically demonstrate that their EAP programmes really do have a measurable impact on student writing quality. In addition to this, any organisation planning to implement an EAP language programme may need to factor a research element into course design, and not simply use published materials in an unreflective, ‘off the shelf’ fashion. If the current thesis can contribute towards this, then it may well be of benefit to the profession. An allied and highly topical issue is that of the UK’s impending withdrawal from the European Union, which may well have implications for the level of funding available for research projects. As a recent Newsnight (05/07/16) article and broadcast points out, “more than 60% of the UK's international research partners are from other EU countries”. It may be the case that university-managed rather than privately run English language units will become more important as a revenue source if academic research grants are cut or reduced and as a result,

the value of English language provision at Higher education institutions may increase.

7.5 Further issues

The current thesis has very much concentrated upon measurable movements in specific linguistic characteristics. Based on these findings and in emphasising the importance of features which have shown tractability, the thesis has presented a range of implications for course design, assessment and the professional status of the EAP profession. There are other factors which have not been explicitly acknowledged by the current study but which may be of importance in any consideration of the development of second language writing quality.

One such factor is the nature of the task in framing the kind of language produced by the learner. The current study has attached a lot of significance to the production of free writing created under exam-like conditions. Most academic writing is not of this nature and any observation concerning likely progression in second language writing needs to be tempered by awareness of the reality of written production. For example, the writing of a dissertation requires a considerable amount of input at the levels of organisation and register, including such features as providing structure and references, characteristics which the current study has not measured.

It is also relevant to acknowledge the importance of subject discipline in the production of writing. For example, the incidence of agentless passives, although much more prevalent in academic prose in general (See 4.4.1) has a greater frequency in scientific texts than in others such as English literature. The thesis does not propose uniformity of feature development in this regard, but it would be possible to argue that the science student should, after completion of the EAP programme at the University of Birmingham, have a *greater capacity* to use the passive when the occasion arises, as it is likely to do.

Another area of consideration lies in the input provided by individual teachers. There

is no intention here to ignore or down play the input of teachers and their specific classroom styles, preferences and emphases. Indeed, the teacher's role can be said to be vital in the transmission, not only of the information contained in the course syllabus, but also in the production and maintenance of an industrious and harmonious classroom atmosphere, without which little learning could be expected to take place. One could imagine the substitution of a regular classroom teaching schedule with a series of on-line, teacher free classes using the course materials. It would be far cheaper to implement although it would be impossible to determine whether the same degree of progression could be observed without experimentation. The current thesis, therefore strongly recognises the importance of the role of individual teachers and whilst stressing the importance of well-designed, research-led course materials considers that this can probably best be implemented by well-motivated and professionally trained classroom teachers.

From an individual learner perspective, it may be possible to conceive of developments in accuracy being sacrificed as students become more adventurous in their written production. In other words, there may be a trade-off, conscious or otherwise between producing writing of greater linguistic sophistication and writing containing fewer errors. Teachers could possibly be made aware of this by course designers and sensitised to this characteristic of writing development and it might be suggested that teachers recognise that students are likely to make a lot of error as they produce more complex, less orally-informed writing. It could be that an active intervention strategy is effective, for example at tutorials and in the marking of students' work whereby these very errors are identified and highlighted together with a recognition the production of these errors is acceptable and indeed desirable on the part of the learner. This recognition of the balance between the development of accuracy and complexity might also be recognised in terms of assessment. Marking criteria for example might incorporate into descriptors such comments as "shows

ambition in use of language or “shows evidence of linguistic complexity despite errors of a non –intrusive nature” or “shows evidence of movement from a speech to a writing-informed style of written production despite the continued presence of ambition-related error”. These descriptors would clearly require modification but the idea of emphasising the balance between the development of complexity and accuracy may be an important one, especially as individual learners might become discouraged from experimentation in their writing if they feel that they are likely to make too many errors. This may be also be another aspect of a teacher’s role, the ability to recognise that a student is incorporating the syllabus items into their writing output, for example trying to produce more complex noun phrases or using the passive but in doing so is making errors. A successful teacher, one might imagine would have the ability and professional skill to able to do this and to outline in for example, a classroom session, that errors are likely to be made and that students should not worry unduly about it. This issue of complexity and linguistic sophistication running ahead of accuracy is one area for possible further research and I discuss it in the subsequent section.

7.6 Suggestions for further research

One area of possible interest for further research and which might attract research funding lies in following up the performances of individual students in their subsequent degree programmes. Students could be tracked within their target departments in order to discover whether they achieve academic success in terms of grades on their Masters’ courses and this may provide in a sense, a criterion for the predictive validity of Coh Metrix and MAT. These two programmes could also be used to measure linguistic feature development at different times in their MA/MSc study and such the research conducted in the current thesis could be continued in order to identify further progression or indeed regression. It might also be interesting to discover any other specific feature movement which might occur

at the conclusion of the preessional programme which has not been identified after intensive EAP course of 20,15, 10 or 6 week duration.

Another research possibility could involve comparison with postgraduate students who have not taken preessional courses, in other words students who have taken an entry route into the University by means of external examinations such as IELTS or TOEFL. Such students will have not have undertaken preessional programmes and it may be fruitful to measure their individual Coh Metrix and MAT scores for particular linguistic features at the commencement of their degree courses and compare them to a post-preessional cohort.

The variety of subject disciplines might also be the focus of further research with a possible comparison between writing development by students in for example natural science and arts related subjects such as literature and music. To this end, the research tools of Coh Metrix and MAT could be applied at various stages of students' postgraduate programmes and the results subjected to factor analysis to determine whether any significant feature movement could be identified for specific subject areas. A similar approach might also be used to examine the production of linguistic features in certain written genres such as essays, lab reports, dissertations and even PhD theses. All of these genres could be examined at various stages over the time periods of the students' courses of study and observations made concerning any progression patterns which were in evidence.

Being based largely on the computational analysis of large sets of attested data, the current study is highly replicable for other EAP programmes. Perhaps the most obvious and important replication study would involve using the same research tools to investigate a comparable corpus of student writing collected from a preessional EAP programme that has been outsourced to a third party private sector provider such as INTO or Kaplan, as discussed in Chapter One. It would be very interesting to find out whether the improvements

in student writing identified in the current study are matched or even exceeded by programmes being operated in this way. If they are, this would have dire consequences for those who argue (as this thesis has done) that EAP is a highly skilled professional activity that needs to be done on an academically-informed basis. If the third party programmes fall short of the benchmark figures provided by the current study, however, it will be incumbent on private sector providers and the universities who employ them to justify their continued existence, especially given the increasingly competitive and quality-driven international EAP market as described in Section 7.3 above.

As well as extending the methodology of the current study to new institutional contexts, there remains considerable scope for future research to focus on aspects of second language writing development that this study was not able to investigate in detail. In particular, the subject of second language writing accuracy as discussed in 6.4, was not significantly addressed by the current study and future work using error counts and more sophisticated error-analysis techniques may be of value (for a discussion of the subject of writing error see, for example, Ferris 2003; Liu 2008; Housen and Kuiken 2009; Thewissen 2015). Studies of the type and frequency of errors made by second language learner writers before and after their courses of study would also be of great interest. It would be interesting to ascertain whether there is in fact any improvement in terms of the accuracy of students' writing as they progress through their higher degree programmes and to what extent this matches developments in other linguistic features.

The findings suggested by the current research would also be enhanced by a series of follow up studies which could offer further investigation into the features identified to have shown consistent movement. The current study uses a diachronic corpus, but any investigation over a single time frame may also be useful. For example, the identification of

reduced personalisation may present itself as worthy of validation by subsequent empirical studies, not necessarily with the same corpus size (500+ matching pairs) although if resources are available, with a larger data base. It might also be fruitful to investigate a feature such as the increase in agentless passives, and it may be possible to conduct a more focused, smaller-scale study with a control corpus of scientific texts for example. The wide use of a self-developed corpus may indeed be worthy of consideration by a variety of EAP and educational related professionals with a range of possible objectives.

A further issue of possible interest concerns the learning and development *profiles of different language groupings*. Clearly, there are likely to be many factors which impact upon the ways in which, for example, Chinese and Arabic students acquire EAP writing competency. EAPCORP includes data from several different groups of language users, the commonest being Mandarin Chinese, and there has been no specific investigation of differences between them in terms of emerging writing profiles. This may be an area of interest to researchers and teachers, and it may be important to ascertain whether in fact there are any observable differences between, for example, a “typical” Chinese and a “typical” Greek learner in the use of certain linguistic features such as articles, tenses and prepositions. This kind of study could certainly be carried out over the course of a short intensive period of language instruction and the results could, if sensitively presented, offer fruitful contributions to the issue of second language writing development.

An additional contemporary issue for course providers is that of entry level thresholds. Using the International English Language Testing System or IELTS (IELTS 2015) examination as a benchmark, for example, it would be of interest to ascertain whether a greater degree of writing development occurs when the learner has a higher examination graded level of English. The data assembled for this study has suggested firm improvements over a range of course lengths, with a 10 week time frame being the most statistically robust

cohort. Entry levels as indicated, depend on target course requirements but the typical IELTS score of a 10 week student is 6.0 with that of a 6 week student being 6.5 and for a 20 weeker the score is 5.0.

The study has not addressed the issue as to whether it is possible to construct a typical development profile for the various IELTS entry levels, but it is reasonable to suggest that an aggregation of research findings following the methodology developed in this thesis could provide a very firm empirical foundation for such an enterprise. Needless to say, this would be of great interest to researchers, course designers and teachers alike. IELTS levels are invoked here because this is the most common entry examination for UK universities, but there are others, for example the TOEFL series (ETS TOEFL 2015) and it is relatively easy to frame an entry profile of English language levels based on the various examination equivalences. This, a corpus-based analysis of writing competencies at various examination based grade levels, is a venture which may attract interest (and perhaps material support in the form of research funding) from the examination bodies themselves. It might also help to re-establish links between EAP as a teaching enterprise and the wider academic community of applied linguistics which, with the advent of university EAP privatisation, have come increasingly under threat in recent years.

References

- Anthony, L. (2014) *AntConc (Version 3.4.4)* [Computer Software]. Tokyo, Japan: Waseda University. Available at: <http://www.laurenceanthony.net/software/antconc> {Accessed 15 Jul. 2016}.
- Asencion-Delaney, Y. and Collentine, J. (2011) 'A Multidimensional Analysis of a Written L2 Spanish Corpus.' *Applied Linguistics*, 32 (3), pp. 299-322.
- BALEAP competency framework (2008) *Competency framework for teachers of English for Academic Purposes*. Available at: <https://www.baleap.org/wp-content/uploads/2016/04/teap-competency-framework.pdf> {Accessed 1 Jul.2016}.
- BALEAP (2016) *The Global Forum for EAP Professionals*. Available at: <https://www.baleap.org/home> {Accessed 1 Jul.2016}.
- BALEAP (2016) *Course guide*. Available at: www.baleap.org.uk/course/guide/index.aspx {Accessed 1 Jul.2016}.
- Baayen, R., Piepenbroch, R. and Gullikers, L. (1995) *CELEX2 LDC96L* Web Download. Philadelphia: Linguistic Data Consortium. Available at: <https://catalog.ldc.upenn.edu/LDC96L14> {Accessed 15 Jul.2016}.
- Basal, A. and Bada, G. (2012) 'Use of first person pronouns: A corpus study of journal articles.' *Energy Education Science and Technology Part B. Social and Educational Studies*, 4(3), pp. 1777-1788.
- Bereiter, C. and Scardamalia, M. (1987) *The psychology of written composition*. Hillsdale NJ: Lawrence Erlbaum Associates.
- Biber, D. (1986) 'Spoken and written textual dimensions in English; resolving the contradictory findings.' *Language*, 62 (2), pp. 384-414.
- Biber, D. (1988) *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, D., Conrad, S. and Reppen, R. (1996) 'Corpus-based investigations of language use.' *Annual Review of Applied Linguistics*, 16 (1996), pp. 115-136.
- Biber, D., Conrad, S. and Reppen, R. (1998) *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Biber, D., Finnegan, E., Johansson, S., Conrad, S. and Leech, G (1999) *Longman Grammar of Spoken and Written English*. London: Pearson.

Biber, D. and Conrad, S. (2001) 'Quantitative corpus-based research: Much more than bean counting.' *TESOL Quarterly*, 35(2), pp.331-336.

Biber, D. (2002) *University Language: a corpus based study of spoken and written registers*. Amsterdam: John Benjamins.

Biber, D., Conrad, S., Reppen, R., Byrd, P. and Helt, M. (2002) 'Speaking and writing in the university: a multidimensional comparison.' *TESOL Quarterly*, 36 (1), pp. 9-48.

Biber, D., Conrad, S., Reppen, R. and Byrd, P. (2003) 'The Authors Respond: Strengths and Goals of Multidimensional Analysis.' *TESOL Quarterly*, 37 (1), pp. 151-155.

Biber, D., Davies, M., Jones, J. and Tracy-Ventura, N. (2006) 'Spoken and written register variation in Spanish: A multi-dimensional analysis.' *Corpora*, 1(1), pp. 1-37.

Biber, D., Gray, B. and Poonpon, K (2011) 'Should we use characteristics of conversation to measure grammatical complexity in L2 writing development?' *TESOL Quarterly*, 45 (1), pp. 5-35.

Biber, D. and Gray, B. (2013) 'Nominalizing the verb phrase in academic science writing,' in Aarts, B., Close, J., Leech, G., and Wallis, S. (eds.) *The Verb Phrase in English*. Cambridge: Cambridge University Press, pp. 99-132.

Bitchener, J., Young, S. and Cameron, D. (2005) 'The effect of different types of corrective feedback on ESL student writing.' *Journal of Second Language Writing*, 14 (2005), pp. 191-205.

Bitchener, J. (2012) 'A reflection on the language learning potential of written CF.' *Journal of Second Language Writing*. 21 (4), pp. 348-363.

Brill, E. (1995) 'Transformation-Based error driven learning and natural language processing: A case study in part of speech tagging.' *Computational linguistics*, 21(4), pp. 543-565.

BBC Newsnight *Brexit Britain – one month in*. Newsnight Special. Available at: <http://www.bbc.co.uk/iplayer/episode/b07m8n2j/newsnight-brexit-britain-one-month-in-newsnight-special> {Accessed 28 Jul 2016}

British Council *Content and Language Integrated Learning (2014)* Available at: <https://www.teachingenglish.org.uk/article/content-language-integrated-learning-clil> {Accessed 1 Jul. 2016}.

British Council Accreditation *UK Handbook 2014 and 2015*. Available at: <https://britishcouncil.org/education/accreditation> {Accessed 1 Jul. 2016}.

British National Corpus, (2007) *Version 3 (BNC XML Edition)* Distributed by Oxford University Computing Services on behalf of the BNC Consortium. Available at: <http://www.natcorp.ox.ac.uk> {Accessed 15 Jul. 2016}.

- Brown, D. (1987) *Principles of Language Learning and Teaching*. Englewood Cliff, NJ: Prentice Hall.
- Cervatiuc, A. (2008) 'ESL Vocabulary Acquisition: Target and Approach.' *The Internet TESOL Journal*, 14 (2008) Available at: <http://iteslj.org/Articles/Cervatiuc-VocabularyAcquisition.html> {Accessed 15 Jul. 2016}.
- Chalker, S. and Weiner, E. (1998) *Oxford Dictionary of English Grammar*. Published online 2014.
Available at:
<http://www.oxfordreference.com/view/10.1093/acref/9780192800879.001.0001/acref-9780192800879> {Accessed 15 Jul. 2016}.
- Chandler, J. (2003) 'The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing.' *Journal of Second Language Writing*, 12 (3), pp. 267–296.
- Charles, M. (2014) 'Getting the corpus habit. EAP students long term use of personal corpora.' *English for Specific Purposes*, 35 (1), pp. 30-40.
- Charniak, E. (2000) 'A maximum-entropy-inspired parser.' *Proceedings of the First Conference of the North American Chapter of the Association for Computational Linguistics*. San Francisco: Morgan Kaufmann.
- Chen, Y. and Baker, P. (2010) 'Lexical Bundles in L1 and L2 Academic Writing.' *Language Learning and Technology*, 14 (2), pp. 30-49.
- Cho, K. (2002) 'A cognitive linguistic approach to the acquisition of English prepositions.' *JACET bulletin*, 35 (2002), pp. 63-78.
- Churchill, E. and Dufon, M. (2006) 'Evolving threads in study abroad research', in Dufon, M., and Churchill, E. (eds.) *Language learners in study abroad contexts*. Clevedon, England: Multilingual Matters, pp. 1-27.
- Chuang, F. and Nesi, H. (2006) 'An analysis of formal errors in a corpus of L2 produced by Chinese students.' *Corpora*, 1(2), pp. 251-271.
- Clough, P., Gaizauskas, R., Piao S. and Wilks, Y. (2001) 'METER: Measuring Text Reuse.' *Proceedings of the 40th annual meeting on Association for Computational Linguistics*, pp. 152-159.
- Collentine, J. (2004) 'The effects of learning contexts on morphosyntactic and lexical development.' *Studies in Second Language Acquisition*, 26 (2), pp. 227-248.
- Coltheart, M. (1981) 'The MRC Psycholinguistic Database.' *Quarterly Journal of Experimental Psychology*, 33 (A), pp. 497-505.
- Coxhead, A. (2000) 'A new Academic Word List.' *TESOL Quarterly*, 34 (2), pp. 213-238.

Crossley, S. and McNamara, D. (2011) 'Shared features of L2 writing: Intergroup homogeneity and text classification.' *Journal of Second Language Writing*, 20 (4), pp. 271-285.

Crossley, S., Allen, D. and McNamara, D. (2011) 'Text readability and intuitive simplification: A comparison of readability formulas.' *Reading in a Foreign Language*, 23 (1), pp. 86-101.

Crossley, S., Greenfield, J. and McNamara, D. (2008) 'Assessing text readability using psycholinguistic indices.' *TESOL Quarterly*, 42 (3), pp. 475-493.

Crossley, S., Weston, J., Sullivan, M. and McNamara, D. (2011) 'The Development of Writing Proficiency as a Function of Grade level: A Linguistic Analysis.' *Written Communication*, 28 (3), pp. 282- 311.

Crossley, S., Kyle, K. and McNamara, D. (2016) 'The development and use of cohesive devices in L2 writing and their relations to judgements of essay quality.' *Journal of Second Language Writing*, 32 (2016), pp.1-16.

Cumming, A. and Riazi, A. (2000) 'Building models of adult second-language writing instruction.' *Learning and Instruction*, 10 (1), pp. 55-71.

Davies, S. (2003) 'Content-based instruction in EFL contexts.' *The internet TESL Journal*. 1X, Available at: <http://iteslj.org/Articles/Davies-CBI.html> {Accessed 1 Jul. 2016}.

De Bot, K., Lowie, W. and Verspoor, M. (2007) 'A Dynamic Systems Theory approach to second language acquisition.' *Bilingualism: Language and Cognition*, 10 (1), pp. 7-21.

Elbow, P. (1991) 'Reflections on academic discourse and how it relates to freshmen and colleagues.' *College English*, 53 (2), pp. 135-55.

Ellis, N. (2008) 'The dynamics of second language emergence: Cycles of language use, language change and language acquisition.' *Modern language Journal*, 92 (2), pp. 232-249.

Evans, N., Hartshorn, J., Cox, T. and Martin de Jel, T. (2014) 'Measuring written linguistic accuracy with weighted clause ratios: A question of validity.' *Journal of Second Language Writing*, 24 (2014), pp. 33-50.

Fellbaum, C. (ed.) (1998) *Wordnet: An electronic lexical database*. Cambridge MA: MIT Press.

Ferris, D. (1994) 'Lexical and syntactic features of ESL writing by students at different levels of L2 proficiency.' *TESOL Quarterly*, 28 (2), pp .141- 420.

Ferris, D. (2003) *Response to Student Writing: Implications for second language students*. Mahwah New Jersey, Lawrence Erlbaum.

- Flowerdew, J. (2010) 'Use of signalling nouns across L1 and L2 writer corpora.' *International Journal of Corpus Linguistics*, 15 (1), pp. 36-55.
- Gilquin, G. and Paquot, M. (2007) 'Finding one's voice in losing it: Learner academic writing and medium variation.' *Conference proceedings Third BAAHE conference Louvain*. No URL available.
- Graesser, A., McNamara, D. and Kulikowich, D. (2011) 'Coh –Metrix: Providing multi-level analyses of text characteristics.' *Educational Researcher*, 40 (5), pp. 223-234.
- Granger, S. and Petch-Tyson, S. (1996) 'Connector usage in the English essay writing of native and non-native EFL speakers of English.' *World Englishes: Journal of English as an international and intranational language*, 15 (1), pp. 19-29.
- Granger, S. (1998) 'Prefabricated patterns in advanced EFL writing: Collocations and formulae', in Cowie, A. (ed.) *Phraseology: Theory, Analysis and applications*. Oxford, Oxford University Press, pp. 145-160.
- Granger, S. (2003a) 'Error-tagged learner corpora and CALL: a promising synergy.' *CALICO (Computer Assisted Language Instruction Consortium -special issue on Error Analysis and Error Correction in Computer Assisted Language Learning)* 20 (3), pp. 465-80.
- Granger, S. (2003b) 'The International corpus of learner English: A new resource for language learning and teaching and second language acquisition research.' *TESOL Quarterly*, 37 (3), pp. 538-546.
- Granger, S., Dagneaux, E., Meunier, F. and Paquot, M. (2009) *ICLE Version 2*. Available at: <https://www.uclouvain.be/en-277586.html> { Accessed: 25 Jul.2016 }
- Grant, L. and Ginther, A. (2000) 'Using Computer-Tagged Linguistic Features to Describe L2 Writing Differences.' *Journal of Second Language Writing*, 9 (2), pp. 123-145.
- Hall, C., McCarthy, P., Lewis, G., Lee, D. and McNamara, D. (2007) 'Using the Coh-Metrix to assess differences between English Language varieties.' *Coyote papers* 15, pp. 40-54. Available at: <http://arizona.openrepository.com/arizona/handle/10150/126392> {Accessed 1 Jul 2016}.
- Halliday, M. (1985) *Spoken and written language*. Geelong Victoria: Deakin University.
- Halliday, M. and Hasan, R. (1976) *Cohesion in English*. London: Longman.
- Hasselgren, A. (1994) 'Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary.' *International Journal of Applied Linguistics*, 4 (2), pp. 237-258.
- Henriksen, B. (1996) *Semantisation, Retention and accessibility: Key concepts in vocabulary learning*. Paper presented at the AILA Congress, Jyväskylä, Finland (08/96). No URL available.

- Housen, A. and Kuiken, F. (2009) 'Complexity, Accuracy and Fluency in Second Language Acquisition.' *Applied Linguistics*, 30 (4), pp. 461-473.
- Hunt, K. (1966) 'Recent measures in syntactic development.' *Elementary English*, 43 (7), pp. 732-739.
- Hyland, K. (1997) 'Qualification and Certainty in L1 and L2 Students' Writing.' *Journal of Second Language Writing*, 6 (2), pp. 183-205.
- Hyland, K. (2008) 'As can be seen: Lexical bundles and disciplinary variation.' *English for Specific Purposes*, 27 (1), pp. 4-21.
- IELTS (2015) International English Language Testing System*. Available at www.cambridgeenglish.org/exams/ielts { Accessed 25 Jul. 2016}.
- INTO the global education partner (2015)* Available at: www.into-corporate.com {Accessed 25 Jul.2016}.
- Issitt, S. (2008) 'Does it stick? Some evidence from students' free writing.' *BALEAP Professional Interest Meeting. Nottingham Trent University*. Available at: <http://qmlanguagecentre.on-rev.com/baleap/archive/media/uploads/pim/Steve-Issitt.pdf> {Accessed 25 Jul. 2016}.
- Issitt, S (2010) We do make a difference: assessing the development of writing over an EAP programme. *BALEAP PIM Re-assessing assessment in EAP: challenges and opportunities*. University of Nottingham. URL not available.
- Jiang, W. (2013) 'Measurements of Development in L2 Written Production: The case of L2 Chinese.' *Applied Linguistics*, 34 (1), pp. 1-24.
- Jones, M. and Haywood, S. (2004) 'Facilitating the acquisition of formulaic sequences', in Schmitt, N. (ed.) *Formulaic sequences*. Amsterdam.: John Benjamins, pp. 269-300.
- Kaplan International College (2016) *Our partner universities*. Available at: <https://www.kic.org.uk/pathways/quality/universities/> {Accessed 25 Jul 2016}.
- Kennedy, C and Thorp, D. (2007) 'A corpus investigation of linguistic responses to an IELTS Academic writing task', in Taylor. L. and Falvey. P. (eds.) *Studies in Language Testing 19: IELTS collected papers, 2: research in speaking and writing assessment*. Cambridge: Cambridge University Press, pp. 316-378.
- Knoch, U., Roushad, A., Oon, S. and Storch, N (2015) 'What happens to ESL students' writing after three years of study at an English medium university?' *Journal of Second Language Writing*, 28 (2015), pp. 39-52.
- Landauer, T., McNamara, D., Dennis, S. and Kintsch, W (2007) *Handbook of latent semantic analysis*. Mahwah NJ: Erlbau.

- Langacker, R. (2000) 'A dynamic usage-based model', in Barlow, M and Kemmer, S. (eds.) *Usage Based Models of Language*. Stanford: CSLI Publications, pp. 1-63.
- Langacker, R. (2009) 'A dynamic view of usage and language acquisition.' *Cognitive linguistics*, 20 (3), pp. 627-640.
- Larsen-Freeman, D. (2006) 'The Emergence of Complexity, Fluency, and Accuracy in the Oral and Written Production of Five Chinese Learners of English.' *Applied Linguistics*, 27 (4), pp. 590-619.
- Laufer, B. and Nation, P. (1995) 'Vocabulary size and use: Lexical Richness in L2 Written Production.' *Applied Linguistics*, 16 (3), pp. 307-322.
- Laufer, B. and Goldstein, Z. (2004) 'Testing Vocabulary Knowledge: Size, Strength and Computer Adaptiveness.' *Language Learning*, 54 (3), pp. 399-436
- Li, J. and Schmidt, N. (2009) 'The acquisition of lexical phrases in academic writing: A longitudinal case study.' *Journal of Second Language Writing*, 18 (2009), pp. 85-102.
- Liu, Y. (2008) 'The effects of error feedback in second language writing.' *Arizona Working Papers in SLA & Teaching* 15, pp. 65-79. Available at: <http://slat.arizona.edu/sites/default/files/page/awp15liu.pdf> {Accessed 25 Jul. 2016}.
- Louwerse, M. (2002) 'An analytic and cognitive parameterization of coherence relations.' *Cognitive Linguistics*, 12 (3), pp. 291-315.
- Lu, X. (2010) 'Automatic analysis of syntactic complexity in second language writing.' *International Journal of Corpus Linguistics*, 15 (4), pp. 474-496.
- Lu, X. and Ai, H. (2015) 'Syntactic complexity in college level English writing: Differences among writers with diverse L1 backgrounds.' *Journal of Second Language Writing*, 29 (2015), pp. 16-27.
- Martin-Martin, P. (2008) 'The mitigation of Scientific claims in research papers. A comparative study.' *International Journal of English Studies*, 8 (2), pp. 133-152.
- Mazgutova, D. and Kormos, J. (2015) 'Syntactic and lexical development in an intensive English for Academic Purposes Programme.' *Journal of Second Language Writing*, 29 (2015), pp. 3-15.
- McCarthy, P., Guess, R. and McNamara, D. (2009) 'The components of paraphrase evaluations.' *Behavioural Research Methods*, 41 (3), pp. 682-690.
- McCarthy, P., Lewis, G., Duffy, D. and McNamara, D. (2006) 'Analysing writing styles with Coh-Metrix.' In *Proceedings of the Florida artificial intelligence research society international conference (FLAIRS)* Available at: <http://aaai.org/Papers/FLAIRS/2006/Flairs06-151.pdf> {Accessed 25 Jul. 2016}.

- McCarthy, P., Lightman, E., Duffy, D. and McNamara, D. (2007) 'Using Coh-Metrix to Assess Distributions of Cohesion and Difficulty: An Investigation of the structure of High-School Textbooks.' *Proceedings of the 28th annual conference of the Cognitive Science Society, Vancouver, Canada*. 2006. Available at: <http://csjarchive.cogsci.rpi.edu/proceedings/2006/docs/p2556.pdf> {Accessed 25 Jul. 2016}.
- McNamara, D., Louwrese. M., Cai, Z. and Graesser, A. (2005) *Coh-Metrix version 1.4*. URL not available.
- McNamara, D., Louwrese, D., McCarthy, P. and Graesser. A. (2010) 'Coh-Metrix: capturing linguistic features of cohesion.' *Discourse Processes*, 47 (4), pp. 292-230.
- McNamara, D., Louwrese, M., Cai, Z. and Graesser, A. (2013) *Coh-Metrix version 3*. cohmetrix.com/documentation_indices.html {Accessed 25 Jul. 2016}.
- McNamara, D. and Graesser, A. (2011) 'Coh-Metrix. An Automated Tool for Theoretical and Applied Natural Language Processing', in McCarthy, P., Graesser, A and Boonthum, C. (eds.) *Applied Natural Language Processing: identification, investigation and resolution*, Hershey PA. IGI Global, pp.188-205.
- McNamara, D., Graesser, A., McCarthy, P. and Cai, Z. (2014) *Automated Evaluation of Text and Discourse with Coh-Metrix*. Cambridge: Cambridge University Press.
- Milton, J. and Meara, P. (1995) 'How periods abroad affect vocabulary growth in a foreign language.' *ITL Review of Applied Linguistics*, 107-108 (1995), pp. 17-34.
- Nation, P. and Waring, R. (1997) 'Vocabulary size, text coverage and word lists', in Schmidt, M. and McCarthy, M. (eds.) *Vocabulary: Description, Acquisition, Pedagogy*. Cambridge: Cambridge University Press, pp. 6-19.
- Nesi, H. (2009) 'A multidimensional analysis of student writing across levels and disciplines', in Edwardes, M. (ed.) *Taking the measure of applied linguistics: Proceedings of the BAAL annual conference. University of Swansea, 11-13 September 2008*. London BAAL: Seitjungil Press
- Nesi, H. and Gardner, S. (2012) *Genres across the disciplines: student writing in higher education*. Cambridge: Cambridge University Press
- Nesi, H., Gardner, S., Thompson, P. and Wickens, P. (2008) *The British Academic Written Corpus*. Available at: www.coventry.ac.uk/bawe {Accessed 25 Jul. 2016}.
- Nini, A. (2014) *Multidimensional Analysis Tagger 1.2 - Manual*. Available at: <https://sites.google.com/site/multidimensionaltagger/versions> {Accessed 26 Jul. 2016}.
- Northwest Regional Educational Laboratory. (2006) *About 6 +1 Traits writing*. Available at: <http://educationnorthwest.org/traits> {Accessed 26 Jul. 2016}.

O'Malley, J. and Chamot, A. (1990) *Learning strategies in second language acquisition*. Cambridge: Cambridge University Press.

Oakey, D. and Treece, P. (2008) *English for Academic Purposes Programme: Course materials*. University of Birmingham.

Ortega, L. (2003) 'Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing.' *Applied Linguistics*, 24 (4), pp. 492-518.

Ortega, L. (2015) 'Syntactic complexity of L2 writing: progress and expansion.' *Journal of Second Language Writing*, 29 (2015), pp. 82-94.

Paquot, M. (2015) *International Corpus of Learner English (ICLE)*. Available at: <https://www.uclouvain.be/en-cecl-icle.html> {Accessed 25 Jul. 2016}

Paquot M (2015) *LOCNESS Corpus*
<https://www.uclouvain.be/en-cecl-locness.html> {Accessed 29 Jul. 2016}

Parkinson, J. and Musgrave, J. (2014) 'Development of noun phrase complexity in the writing of English for Academic Purposes students.' *Journal of English for Academic Purposes*, 14 (2014), pp. 48-59.

Perez, G. (2004) 'Textual dimensions in biological science corpora.' *Revista Virtual de Estudos da Linguagem*, 2 (3) Available at: http://www.revel.inf.br/files/artigos/revel_3_textual_dimensions.pdf {Accessed 25 Jul. 2016}.

Pienemann, M. and Brindley, G. (1988) 'Constructing an acquisition-based procedure for second language assessment.' *Studies in second language acquisition*, 10 (2), pp. 217-243.

Reid, J. (1992) 'A computer text analysis of four cohesive devices in English discourse by native and non-native writers.' *Journal of Second Language Writing*, 1 (2), pp. 79-107.

Richards, B. (1987) 'Type/token ratios: what do they really tell us?' *Journal of Child Language*, 14 (2), pp. 201-209.

[Riley –Jones, G. \(2012\) 'The Construction of Professional Identity.' In *English for Academic Purposes: A Renewed Call for an Engagement with Ideology*. BALEAP PIM, Durham, 2012.](#)

[Available at: http://qmlanguagecentre.on-rev.com/baleap/archive/media/uploads/pim/durham-june-9-2012/extended-abstracts/Durham_PIM_Extended_Abstract_Riley-Jones.pdf](http://qmlanguagecentre.on-rev.com/baleap/archive/media/uploads/pim/durham-june-9-2012/extended-abstracts/Durham_PIM_Extended_Abstract_Riley-Jones.pdf) {Accessed 18 Jul. 2016}.

Robertson, D. (2000) 'Variability in the use of the English article system by Chinese learners.' *Second Language Research*, 16 (2), pp. 135-172.

Sabatini, J., Albro, E. and O'Reilly, T. (2012) *Measuring Up: Advances in How We Assess Reading Ability*. Lanham, MD: Rowman and Littlefield Education.

- Scott, M. (2008) *Wordsmith Tools Version 5*. Liverpool: Lexical Analysis Software.
- Seidlhofer, B. (ed.) (2003) *Controversies in Applied Linguistics*. Oxford: Oxford University Press.
- Shaw, P. and Liu, E. (1998) 'What Develops in the Development of Second-language Writing.' *Applied Linguistics*, 19 (2), pp. 225-254.
- Sinclair, J. (1991) *Corpus, Concordance and Collocation*. Oxford: Oxford University Press.
- Spalding, E., Wang, J. and Lin, E. (2009) 'Analyzing Voice in the Writing of Chinese Teachers of English'. *Research in the Teaching of English*, 44 (1), pp. 23-51.
- Spinner, P. (2011) 'Second Language Assessment and Morphosyntactic Development.' *Studies in Second Language Acquisition*, 33 (4), pp. 529-561.
- Spoelman, M. and Verspoor, M. (2010) 'Dynamic Patterns in Development of Accuracy and Complexity: A Longitudinal Case study in the Acquisition of Finnish.' *Applied Linguistics*, 31 (4), pp. 532-553.
- Stanat, P., Becker, M., Baumert, J., Lüdtke, O. and Eckhardt, A. (2012) 'Improving second language skills of immigrant students: A field trial study evaluating the effects of a summer learning programme.' *Learning and Instruction*, 22 (3), pp. 159-170.
- Staples, S., Egbert, J., Biber, D. and Gray, B. (2016) 'Academic Writing development at the University Level: Phrasal and Clausal Complexity Across Level of Study, Discipline and Genre.' *Written Communication*, 33 (2), pp.1-35.
- Storch, N. (2009) 'The impact of studying in a second language (L2) medium university on the development of L2 writing.' *Journal of second language writing*, 18 (2), pp. 103-118.
- Tannen, D. (1982) 'Oral and literate strategies in spoken and written narratives.' *Language*, 58 (1), pp. 1-21.
- Tannen, D. (1985) 'Relative focus on involvement in oral and written discourse', in Olson. D., Torrance, N. and Hildyard, A. (eds.) *Literacy, language and learning: the nature and consequences of reading and writing*. Cambridge: Cambridge University Press, pp. 124-147.
- The British National Corpus, version 3 BNC XML Edition. (2007) Distributed by Oxford University Computing Services on behalf of the BNC Consortium. Available at: <http://www.natcorp.ox.ac.uk/> {Accessed 27 Jul. 2016}.
- Thewissen, J. (2015) *Accuracy across proficiency levels: A learner corpus approach*. Presses Universitaires de Louvain.
- Toglia, M. and Battig, W. (1978) *Handbook of semantic word norms*. New York: Lawrence Erlbaum.

Toutanova K., Klein, D., Manning, C. and Singer, Y. (2003) 'Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network.' In *Proceedings of HLT-NAACL 2003*, pp. 252-259.

Tribble, C. (2002) 'Corpora and corpus analysis: New windows on academic writing', in Flowerdew, J.(ed.) *Academic Discourse*. Amsterdam: John Benjamins, pp. 131-408.

UCU (2010a) '*Privatising our Universities. A UCU report on the new cross party consensus and the Americanisation of Higher Education.*' UCU London. Available at: https://www.ucu.org.uk/media/3791/Privatising-our-universities-Feb-10/pdf/ucu_privatisingouruniversities_feb10.pdf {Accessed 25 Jul. 2016}.

UCU (2010b) '*Challenging the Market in Education- Kaplan A UCU Briefing.*' Available at: https://www.ucu.org.uk/media/4601/Kaplan-A-UCU-briefing-Jul-10/pdf/ucu-cme-Kaplan-briefing_jul10.pdf { Accessed 25 Jul.2016}.

Universities UK (2014) '*Patterns and trends in UK Higher Education 2014.*' Available at : <http://www.universitiesuk.ac.uk/facts-and-stats/data-and-analysis/Pages/patterns-and-trends-uk-higher-education-2014.aspx> {Accessed 25 Jul. 2016}.

Vainikka, A. and Young-Scholten, M. (2006) 'The roots of syntax and how they grow: Organic grammar, the basic variety and processability theory', in Unsworth, S., Parodi, A., Sorace, A and Young-Scholten, M. (eds.) *Paths of development in L1 and L2 acquisition*. Amsterdam: Benjamins, pp. 77-106.

Vance, S. (2005) 'Avoiding the personal pronoun I.' *Language and learning online*. Available at: <http://www.monash.edu.au/lis/llonline/writing/arts/sociology/2.3.1.xml> {Accessed 25 Jul. 2016}.

Van Rooy, B. and Terblanche, L. (2009) 'A multi-dimensional analysis of a learner corpus', in Renouf, A. and Kehoe, A. (eds.) *Corpus Linguistics: Refinements and Reassessments*. Rodopi, pp. 239-254.

Van Waes, L. and Leitjen, M. (2015) 'Fluency in writing: a multidimensional perspective on writing fluency applied to L1 and L2.' *Computers and Composition: An International Journal*, 38 (A), pp. 79-95.

Verspoor, M., Schmid, M. and Xu, X. (2012) 'A dynamic usage based perspective on L2 writing.' *Journal of Second Language Writing*, 21 (3), pp. 239-263.

Warren, B., Mendoza, M., Cruz, F., Leao, R. and Caprioli, J. (2016) 'Reward and extinction memories in vm PFC.' *Journal of Neuroscience*, 36 (25), pp. 6691-6702.

Wells, R. (1960) 'Nominal and verbal style', in Sebeok, T. (ed.) *Style in language*. MIT Press, pp. 213-220.

Weston, J., Crossley, S. and McNamara, D. (2010) 'Towards a computational assessment of freewriting quality.' In *Proceedings of the 23rd International Florida artificial intelligence research society conference*. Menlo Park, CA: The AAAI Press, pp. 283-288.

Wolfe-Quintero, K., Inagaki, H. and Kim, H. (1998) *Second Language Development in writing: Measures of Fluency, Accuracy and Complexity*. HI: University of Hawaii Press.

Yamamoto, M. (2006) *Agency and impersonality*. Amsterdam: John Benjamins.

Yasuda, S. (2011) 'Genre based tasks in foreign language writing: Developing writers' genre awareness, linguistic knowledge and writing competence.' *Journal of Second Language Writing*, 20 (2), pp. 95-166.

Zamel, V. (1984) 'Teaching those missing links in writing', in Mackay, S. (ed.) *Composing in a second language*. Cambridge: Newbury House.