THE UNIVERSITY OF HUDDERSFIELD

Error correction through corpus consultation in EAP writing: an analysis of corpus use in a pre-sessional context

A thesis submitted to the University of Huddersfield in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Marcus Bridle

Revised copy with corrections submitted March 18th 2015

Abstract

This study investigates the effect of corpus consultation on the accuracy of learner written error revisions. It examines the conditions which cause a learner to consult the corpus in correcting errors and whether these revisions are more effective than those made using other corrections methods.

Claims have been made for the potential usefulness of corpora in encouraging a better understanding of language through inductive learning (Johns, 1991; Benson, 2001; Watson Todd, 2003). The opportunity for learners to interact with the authentic language used to compile corpora has also been cited as a possible benefit (Thurstun and Candlin, 1998). However, theoretical advantages of using corpus data have not always translated into actual benefits in real learning contexts. Learners frequently encounter difficulties in dealing with the volume of information available to them in concordances and can reject corpus use because it adds to their learning load (Yoon and Hirvela, 2004; Frankenberg Garcia, 2005; Lee and Swales, 2006). This has meant that practical employment of corpus data has sometimes been difficult to implement.

In this experiment, learners on a six week pre-sessional English for Academic Purposes (EAP) course were shown how to use the BYU (Brigham Young University) website to access the BNC (British National Corpus) to address written errors. Through a draft/feedback/revision process using meta-linguistic error coding, the frequency, context and effectiveness of the corpus being used as a reference tool was measured.

Use of the corpus was found to be limited to a small range of error types which largely involved queries of a pragmatic nature. In these contexts, the corpus was found to be a potentially more effective correction tool than dictionary reference or recourse to previous knowledge and it may have a beneficial effect in encouraging top-down processing skills. However, its frequency of use over the course was low and accounted for only a small proportion of accurate error revisions as a whole. Learner response to the corpus corroborated the negative perception already noted in previous studies.

These findings prompt recommendations for further investigation into effective mediation of corpus data within the classroom and continued technological developments in order to make corpus data more accessible to non-specialists.

Key terms: BNC; corpora; concordances; EAP; error correction; pragmatic; pre-sessional; reference tool; top-down; written feedback.

Acknowledgements

I would like to acknowledge the continued guidance and support of Professor Dan McIntyre and Professor Lesley Jeffries. My thanks also go to Doctor David McNulty and Doctor John Stephenson of the University of Huddersfield for their advice on methods of statistical testing.

Contents

Word count: 79,819

[1] Appendices are included on a disc attached to the rear inside cover of the thesis.

List of Abbreviations

?	Sense error
^	Word missing error
A	Article error
Agr	Subject-Verb agreement error
BNC	British National Corpus
C	Corpus correction
CG	Control Group
C/U	Countable / Uncountable error
Con	Contrary meaning error
D	Dictionary correction
EAP	English for Academic Purposes
EFC	Error Free Clause
EFT	Error Free T-unit
EG	Experimental Group
EGAP	English for General Academic Purposes
EOP	English for Occupational Purposes
ESAP	English for Specific Academic Purposes
ESP	English for Specialised Purposes
F/I	Register error
G	Grammar
GE	General English
K	Knowledge correction
O	Other online source correction

P	Punctuation error
P/S	Number error
Prep	Prepositional error
REP	Repetition error
SP	Spelling error
Sub	Subjective language error
SPSS	Statistical Package for the Social Sciences
Syn	Synonym error
T	Tense error
TT	Total T-units
WF	Word form error
WO	Word order error
WW	Wrong word error

List of Tables and Charts

1

Introduction

1.1 Background to This Study

In this thesis, I examine the effectiveness of a corpus being used as a reference tool to correct learner errors on an English for Academic Purposes (EAP) course for potential postgraduate students. A corpus is a body of texts which are compiled from original sources and is designed to be a representative sample of a larger population. These sources can be from a particular domain, such as newspapers or from texts related to a particular genre or area, such as the Gledhill corpus (1995, as cited in Krishnamurthy and Kosem, 2007, p. 60) which is restricted to articles about cancer research. They can also be larger, comprised of texts from a variety of different genres and sub-genres, such as the British National Corpus. Corpora that are stored in computerised databases are usually searchable through specially designed software. One of the key features of a corpus and corpus use is the concept of word frequency (Bruce, 2011, p. 33). Examining word frequency is one way of establishing the relative saliency of a word. Additionally, many corpus interfaces provide a way of examining corpus data through concordances. A concordance is a list which shows a given word occurring numerous times in numerous contexts and cotexts. Through the analysis of concordances, patterns in the language which might otherwise be unnoticed can be foregrounded and this can lead to a better understanding of the way in which the language operates in particular contexts. Yoon (2011) has recorded numerous useful applications for corpus use in second language (L2) learning research and has noted the potential for use in L2 pedagogic contexts. The aim of the study is to identify particular error types which might be revised successfully by learners with corpus data and to examine whether or not this is a correction method which might be more effective than other methods such as corrections

made with student knowledge or with the use of dictionaries The study is completed within the context of a six week pre-sessional course at a British university involving Chinese learners, with one of the over-arching considerations being that any benefits of corpus use must be evaluated within the framework of a real teaching environment.

What follows in this introduction is a brief contextualisation of the main issues relating to the goals of EAP, the role of written English and the place of corpora within EAP , these being discussed in greater depth in the literature review. The presentation of the research questions follows. I then summarise the methodology used and the construction of the experiment before outlining the basic findings which are discussed in chapters 5 and 6.

1.2 EAP and Pre-sessional Courses

Pre-sessional English courses are a common component of many British universities. Due to the current high demand for places on British degree schemes from foreign students, they have become almost *de-rigeur* (Archibald, 2001; Alexander et al, 2008; Hyland, 2009a). This is partly due to the need for the students to familiarise themselves with the conventions of academic English and the types of skills they will need to be successful on their degree courses subsequently, and partly due to the administrative purposes of both HE institutions and the UKBA which ensure that learners have a level of language which is fit for purpose (Banerjee and Wall, 2006, p. 50).

The nature of pre-sessional courses varies from institution to institution. Some have a length of an entire year and follow the pattern of the academic year but the majority are short courses of perhaps between 4 and 16 weeks during the summer. These are intensive courses which finish just before the start of the academic year proper, and learners, assuming they have achieved any course requirements which were set for them, almost immediately begin their degree courses.

The short, intensive nature of these courses results in difficulties for both instructors and students alike. The time factor means it is often impossible to teach or learn at the natural pace of the students. A further complication is what is included on the syllabus. Not everything can be covered in such a short time and catering to individual learner needs is difficult. Even the eventual learning aims of the courses can be difficult to establish; they could be for the learners to have more familiarity with academic English itself, or more of an idea of the language and conventions used within their departments, or to develop study skills or simply to pass an exit exam which will provide a score meeting the conditions of a departmental offer. More often than not, the aims of a course are likely to be a mixture of all these and more.

The fact that it is difficult to cater to all these varying outcomes and needs means that many university pre-sessional courses are alike in that they provide English for General Academic Purposes (EGAP) and attempt to cover academic English in as broad a way as possible. A major focus of most, however, is written academic English as this is not only likely to be an important component of any assessment of the pre-sessional course but is also the skill area in which many students will have to develop the most in order to be successful within their departments and to communicate any research they go on to produce (Alexander et al., 2008).

Again, due to the restricted length of the courses, written academic English at the pre-sessional level has to adopt a wide-angle approach (Widdowson, 1983, p. 6) in an attempt to cover a number of functions which are considered important within the framework of the preparation course itself and which will provide a useful foundation for writing within the students' departments. These functions therefore might be dictated by a knowledge of written genres and language that learners are likely to need later in their academic life, or they may be functions associated with the local assessment criteria of the pre-sessional course.

A combination of the two are likely to be common factors in the choice of materials and language incorporated into the syllabus. In practice, this often means an emphasis on the rhetorical and organisational features which can be taught to students on a whole class basis (Swales, 2005; Bruce, 2011). This may involve looking at some elements of sentence and paragraph level construction and cohesion but is likely, again due to factors of time, to quickly assume a wider focus on whole text construction and basic genres and discourse patterns. These would include, for example, how to produce discussion essays, argumentative essays and problem-solution style essays and might include how to describe a process. These particular patterns feature heavily in widely administered assessment procedures like the International English Language Testing System (IELTS) which are used by British educational institutions to regulate admissions onto full-time study within the UK (Banerjee and Wall, 2006, p. 51). Typically, the basic grammatical elements of the language are taught or reviewed in the context of these frameworks. Problem-solution essays, for example, might involve input on cause and effect language. Describing a process is likely to include work examining the use of passive and active structures, as in Hamp-Lyons and Heasley (2006).

At the same time as these generic elements are covered, EAP writing courses attempt to familiarise learners new to the British university system fundamental skills and conventions of academic English. Skills include summarising, paraphrasing and synthesising sources based on a process of critical reading. Academic conventions include citing and referencing sources and using the correct type of register within the text. These are coupled with a drive towards encouraging learners to develop cultural and inter-cultural awareness and a set of independent learning skills which will equip them for later study within their departments (Watson-Todd, 2003; Alexander et al., 2008; Gillet, 2011).

1.3 Problems Affecting EAP Courses

This one-size-fits all type of framework of these wide-angle, English for General Academic Purposes (EGAP) courses, whilst catering for some of the needs of students, leaves others either marginalised or unaddressed. A common problem is that pre-sessional courses are not necessarily suited to the needs of learners who are, after all, about to begin courses which are specialised in nature and which they are to become specialists in and who might benefit from instruction in ESAP, or English for Specialised Academic Purposes (Jordan, 1997; Hyland, 2002). Whilst some institutions are able to counter this problem to some extent by grouping students of a similar subject interest together in classes, it is often the case that classes will contain students with a wide range of specialisms and therefore a differing array of needs. Those going on to study for an MA in journalism have a set of language and skill requirements that probably differ greatly to those students going on to an MSc in Electronic and Electrical Engineering (EEE). Similarly, potential EEE PhD candidates will have needs which differ from those undertaking the master's course.

These differences might manifest themselves in the types of academic skills which are needed but are also likely to be a problem in terms of linguistics. That is, the actual grammatical constructions which might be commonplace within their area and, more importantly, the lexical elements they need, differ. On a short pre-sessional, time in the classroom is often not enough to focus on the specialised lexical elements which might be required by learners. EAP instructors face what might be considered an impossible task in these situations when trying to provide materials which will be of benefit to all their students in becoming of aware of or acquiring the necessary language (Meyer, 1996, p. 34).

A further problem which evolves from this is that, given the course components might not cater to the individual student (because time does not allow for this), meeting the language needs of separate learners is difficult. It is argued by some (Turner, 2004;

Hartshorn et al., 2010) that the focus on organisation of texts and rhetorical function means that student language (i.e. the proficiency of the language which is produced by the learner) is to a large extent side lined. Micro level errors of form and use of lexis are difficult to address when a class might have twenty learners, each making a number of errors which apply largely only an individual basis.

The fact that many courses assess the work of students based on a variety of criteria which involve band scoring confounds this problem. Band scores, like those used in IELTS, for example, judge student language production based a number of criteria which collapse individual areas of language into wider, generalised criteria which are inevitably subjective to some extent. Band scoring means that, if a student begins a course with an IELTS writing score of 6 and achieves a 7 at the end of the course, we can claim, ostensibly, that an improvement has been made. Unfortunately, the descriptors cannot address in specific terms where the language problems of that student were at the beginning or in specifically what way they have improved by the end. It is difficult to say whether proficiency has improved or to say exactly how or in what respect proficiency has improved; this could be in linguistic accuracy, linguistic complexity or linguistic fluency, or any combination thereof (Wolfe-Quintero, 1998). This means that pre-sessional courses might neglect the idea of linguistic proficiency, as it isn't being measured, and instead are likely to judge improved writing production over a course less on the linguistic evidence per se than on the ability to produce a text which replicates the frameworks presented over the course (Archibald, 2001; Basturkmen and Lewis, 2002). In short, attention to both the individual learner and their detailed linguistic needs can be marginalised on pre-sessional courses.

1.4 Addressing the Problems Through Learner Autonomy

As mentioned already, EAP courses aim to foster a spirit of learner independence. This is judged to be important for students going on to further study, particularly at research level, as they will be expected to be responsible for their own methods of studying and final output. In many cases, the learning background of foreign EAP students is different to that which is the foundation of the British university system, and such students often lack the necessary study skills which will allow them to be successful in the future. EAP courses try, therefore, to create situations in which the learner takes responsibility for their own development (Fletcher, 2004). This might, for example, be in the production of a lengthier text over the period of the semester which receives limited feedback at the draft stage from the teacher. This kind of activity encourages the student to not only take responsibility of time management but also for the actual language they encounter, acquire and use. In doing so, students become both more field independent and also have the opportunity to bridge the lexical EGAP and ESAP divide by themselves. Efforts are also made to ensure that learners become aware of the typical errors they are likely to make and for them to be able to recognise and amend these by themselves. This emphasis on tasks with an emphasis on learner cognition is key to understanding how EGAP courses might mitigate, to some extent, the paucity of certain areas of specific linguistic focus because learners can investigate language specific to their own needs rather than just that presented on a course of study. Inductive approaches like this, where meaning and usage are inferred, consciously or unconsciously, from examples of and exposure to the language itself are in contrast to more traditional deductive methods of learning, where, for example, a concrete rule is presented and then applied by the learner (Gollin, 1998, p. 88; Hedge, 2000, p. 160) . Though the two techniques are not mutually exclusive (Thornbury, 1997, p.138), inductive approaches to learning have been shown to aid in the acquisition of the target language (Benson, 2001).

Measuring the effects of an inductive approach to writing proficiency is complicated due to the number of variables which might be involved. Nevertheless, it is worth investigating as a way of discovering if pre-sessional courses can have any tangible effects on the written language proficiency of students and, if so, in what areas specifically it might be helpful. Additionally, to what extent this type of learning can be included on a course of such a short and restrictive nature, and how this can be facilitated, is another area worthy of examination.

1.5 Corpora and EAP

The use of corpora in the creation of language learning materials is deemed to have a number of benefits. One advantage is that large, searchable bodies of authentic text have is that instances of real language use can be investigated and used as the basis for instruction (Thurstun and Candlin, 1998; Biber and Reppen, 2002; Swales, 2002; Yoon, 2011). Thus, examples of language which were chosen for the classroom based on the instructor's previous experience or intuition can, with reference to a corpus, be verified or refuted, leading to a more realistic choice in the type of language presented to students. In the same way, the creators of materials, in the form text books and dictionaries, can draw both upon the frequencies of particular language items to determine which are incorporated into materials but also use the examples within the corpora to examine typical contexts and co-texts of individual pieces of language (McCarthy and O'Dell, 2008; Coxhead, 2011; Cullen, 2012, Hewings, 2012). This is one way of narrowing the gap between EGAP and ESAP materials and has allowed an increased focus on the lexical aspects of EAP, including areas of register (Biber, 2006) and subject specific language (Lee and Swales, 2006). A particularly prominent product of this type of research has been the Academic Word List (AWL), produced by Coxhead (2000). This is a list of frequently used academic words which, despite

some reservations and failings which are noted in the literature review, has been used as the basis for several widely used EAP text-books (Schmitt and Schmitt, 2011).

This use of corpora has lead to the creation of various large, general corpora which are searchable through a web-based interface and numerous programs allowing for the production and investigation of smaller, specialised corpora, often produced by HE institutions themselves. Thus, there are medical corpora which can be used by medical students or their instructors (Gledhill, 1995, as cited in Krishnamurthy and Kosem, 2007, p. 360). Another field of corpus construction is that of learner corpora. These are corpora which consist of, for example, of essays composed by EFL students (Leech, 1997; Granger, 2004). Software allows items to be tagged as various parts of speech but also allows for items to be tagged as errors. In doing so, a large body of student essays could provide information about the relative frequency not just of the target language, but of learner errors. This could be the basis of focused classroom instruction. Other research using corpora at an EAP level has contrasted learner corpora and 'expert' corpora (either experts in the field or native English speakers) to identify areas of language which prove problematic (Huttner, 2010).

It is not only the experts or researchers, of course, who make use of corpora. Learners themselves can access or build corpora and use tools and interfaces based on corpus information to inform their own language production. These interfaces, now largely web-based, are able to show concordances which can be searched and manipulated by students to examine how exactly particular language features are used, as with the Brigham Young University British National Corpus (BYU-BNC) (Davies, 2004). These types of student investigation are potentially of use in terms of fostering autonomy within EAP. They allow for independent study of the language and, as corpora and concordances merely present the language, rather than, as in a text book, presenting the rules of the language, the learner is

forced to think inductively about the rules behind the language for themselves (Hunston, 2002, p. 171). This can lead to hypothesis testing and experimenting with the language and it has been claimed that this extra cognitive step encourages the acquisition of the language as the learner takes active ownership of it, rather than merely being the passive recipient of information from and instructor or text book. There are many ways of using corpora in the classroom but two common approaches are using corpus materials as research tool or as reference tools.

Using corpora as research materials involves student investigating how language is used through the examination of concordances. It encourages learners to use the corpus as a tool for learning about new language items from the bottom up (as an alternative to them being taught inductively via a set of pre-prepared rules) and might result in the context or point of use of the corpus being compositional in nature (Cresswell, 2007). Using corpora as a reference tool is more akin to using them in a similar way to that of dictionaries or other reference materials, as in Frankenberg-Garcia (2012a). The perceived advantage of corpora over other reference materials is again that the information is both authentic and that it is there in vast quantities, again allowing students to perhaps test their own understanding of the language against that which is used in the corpora (Thurstun and Candlin, 1998; Yoon and Hirvela, 2004).

Corpora as reference tools may have a particularly useful role to play on pre-sessional courses given one of the potential deficiencies mentioned earlier, namely that the linguistic proficiency of the individual is sacrificed at the expense of having to treat classes as largely generic in their language needs. A particular beneficial application of corpora as reference materials might be then, as some previous studies have intimated, in terms of individual error treatment. Instructors could point learners towards a corpus with reference to the errors they make in their writing (Gaskell and Cobb, 2004; Gilmore, 2009). In doing so, they firstly

provide the learner with individual feedback in the form of the identification of linguistic items within the text that are specific to that learner. Secondly, by directing them to the corpus as an available correction tool, they provide the learner with the opportunity to investigate the language in an inductive way, potentially encouraging the development of a sense of autonomy within the student (Lee and Swales, 2006). Furthermore, this act of inductive learning may allow for better acquisition of language (Benson, 2001; Harmer, 2001). The use of corpora, then, could be seen as a potential tool in bridging the gap between EGAP and ESAP, in mitigating the lack of micro level linguistic input of courses and in encouraging some independent learning skills.

Whatever the perceived potential benefits, however, they may be confounded by a number of factors. The first is the complex nature of some corpora. Learners, and instructors, can be loath to the idea of learning not just how to use a new type of technology in the form of the interface, but also to the fact that the investigation of corpora and concordances themselves requires an adjustment of attitude and approach towards language learning (Harmer, 2001; Frankenberg Garcia, 2012b). On a pre-sessional course, the time taken to learn how to use and then adopt corpora might also limit their effectiveness.

A further issue is that, currently, research has not shown to any reliable degree which linguistic items might benefit from the use of corpora as reference tools, whether or not, in terms of particular linguistic items, they might be more effective than other reference tools in improving linguistic proficiency and to what extent, if any, they have upon the subsequent understanding of those linguistic items.

Finally, there seems to be a distinct gap between research as to the potential benefit of corpora as reference tools and the actual benefits of them in a real pedagogic context such as a pre-sessional course (McCarthy, 2007; Yoon, 2011).

In summary, the potential benefits of using corpora in class on EAP courses seem to be many; linguistic proficiency may be addressed, learners individual language needs could be met and there may be the potential of harnessing the inductive learning approach required when using corpus materials to develop learner autonomy. Whether or not these benefits can be realised in an actual classroom context, and particularly on a short duration pre-sessional course is less clear.

1.6 Research Questions

With the above in mind, this thesis investigates the following research questions:

RQ1.1 Are learners on a pre-sessional course able to correct errors in their writing by using corpus data and concordancing as reference tools?

RQ1.2. Is corpus use chosen as a method of error correction over other methods in particular circumstances?

RQ1.3. Are particular error types more or less effectively dealt with using corpora as reference tools than with other correction methods?

RQ2.1 Do learners who use corpora as reference tools show any measurable improvement in their written proficiency?

RQ2.2 What aspects, if any, of student writing improve and can this improvement be judged to be as a result of having accessed corpus data when making corrections?

RQ3.1 What are learner attitudes towards the use of a corpus as a reference tool on a pre-sessional course?

RQ3.2 Do learners begin to use the corpus independent of teacher instruction?

RQ3.3 Do learners' perceptions of corpus use corroborate the quantitative data and to what extent does this conflict with the existing literature?

1.7 Thesis Structure

I first look in more detail at some of the issues and arguments and issues surrounding EAP, writing in EAP and autonomy. These have already been outlined above, but particular emphasis is placed on the problems surrounding the definition of EAP and the pedagogical attitudes towards EGAP and ESAP. Autonomy is also a term which requires further explanation and a working definition in terms of EAP is provided. Writing in EAP is examined in the contexts already mentioned, with further examination of the schism between current pedagogical practices which place an emphasis or structure and organisation, or the macro-level, what effects this has had on the micro-linguistic level of practice and why these micro-linguistic elements are actually of some importance to learner proficiency.

One these areas are established, I examine the use of corpora in more detail. Current uses of corpora as both research and pedagogical tools are outlined, as are the reasons for the gap between these two areas. Examples of corpora being used as research and reference tools are examined and these latter provide some of the foundations for the methodology behind this study and the choice of the research tools. Description of the potential benefits is balanced with a review of some of the problems encountered in previous research.

Chief amongst these is the acknowledgment that successful use of a corpus with a group of learners on a course requires some sort of mediation between the reference source and the learners via the instructor. One obvious example is training in the use of corpora and

any interface chosen, whether on screen or off, but mediation is also present in the type of tasks the learners are given when engaging with the corpus data. I then go on to show that a practical form of mediation is available in error feedback. The arguments which dispute the merits of both error correction in its entirety and also the different methods of written feedback, including indirect feedback, direct feedback and meta-linguistic feedback are covered. The potential benefits of using meta-linguistic feedback in combination with a corpus being used as a reference tool are then explained.

Chapter 3 consists of a detailed description of the experiment itself, how the various elements were adopted or designed to meet the needs of the research questions and why particular research tools have been employed. In brief, the course of a six-week pre-sessional course for Chinese learners was available for research. 26 learners across two groups were assessed on their effectiveness in improving linguistic accuracy through a writing / feedback / correction process. In the correction process, both groups were allowed to consult reference materials, with the experimental group having the additional use of the BYU-BNC corpus interface, in which they were trained. Learners logged both their correction and the method they used for that correction, and this was in turn logged by me, along with whether the correction had been made effectively or not. This cycle continued four times and the results were collated and compared against a baseline and final, extended piece of work which the students produced.

An explanation for the choice of the corpus interface is provided, and the rationale behind the training learners received is given. I also explain the choice of qualitative research tools, which included learner type evaluations, questionnaires and interviews.

This is followed by an explanation of the meta-linguistic feedback code which was employed, partly of which was due to institutional requirements. The problems associated with meta-linguistic codes in various contexts are described and accounted for. The

methodology behind the inter-rater reliability tests which are employed is also outlined at this point.

The methodology section involves a description of measures of written linguistic proficiency. Measures of complexity, fluency and accuracy are discussed and it is accuracy in the form of the t-unit which I use to measure the proficiency of learners over the course of this experiment, as well as the log of error types at draft and revision stages as mentioned before.

Because the experiment deals with real learners in a real pedagogic environment, there are a number of confounding factors which reduce some of the controls available on the experiment. These include learner interaction, access to materials, learner type, age, gender and course requirements. These are all outlined and the relevant compensatory factors are given, the most significant of which is the type of statistical test employed to analyse the raw data. Binomial multiple logistic regression testing was employed throughout the analysis section due to both its ability to take into an equation the large number of independent variables mentioned in both the error types and limitations sections and for the fact that it not only provides evidence that certain variables correlate but also for the fact that the test is able to express the effect size of these correlations, meaning the statistical data it provides can also be used to predict the effect of different independent variables on the dependent variable.

Chapter 4 provides an account of the results which were recorded over the course of the experiment and outlines the main findings of both the quantitative data obtained from the texts produced by the learners and records the qualitative information supplied in the questionnaires and interviews.

In chapter 5 I evaluate the data. This suggests that corpus data can be used to correct errors in a pre-sessional environment. I also find that it may be more effective than

corrections made by students using their previous knowledge in situations where the language issue is pragmatic rather than semantic. However, this positive finding is balanced against results which suggest the use of the corpus is extremely limited in the areas to which it can be applied and the fact that evidence for it having contributed to overall accuracy and longer-term acquisition is equivocal. Student use of the corpus and opinion is also examined and, whilst some learners recognise that the corpus was effective in particular circumstances, it becomes evident that in this course the corpus was not a popular reference tool and that the factors which account for this are similar to those expressed in the literature already available. With this in mind, I go on to examine the limitations of the experiment, the potential implications for corpus use within pre-sessional contexts and discuss the extent to which corpus use might be considered an element of EAP courses in the future.

2

Literature Review

2.1 Introduction

This chapter begins by providing a definition of English for Academic Purposes (EAP) and outlining some of its goals both within and outside of the immediate contexts of the classroom, with particular consideration given to British pre-sessional courses. Concepts of learner autonomy are discussed before the position of writing within EAP is considered. With this background information in mind, I then examine the use of corpora within EAP. This use falls into two broad areas: (i) corpus in research and the production of materials, and (ii) corpus in pedagogy. Research into experiments in the classroom is then sub-divided into two further categories: corpora as research tools and corpora as reference tools. The relative merits and problems of using corpora within the classroom are considered as are some of the possible ways of putting corpora to pedagogic use. In doing so, I review previous pieces of research which employ corpora and outline their main findings. Once I have established the validity of using corpora as reference tools within the classroom I go on to examine the importance of both learner training and mediation of corpus data. This involves examining methods of written feedback, both in terms of their practicality and effectiveness, before I go on to suggest that the use of corpora on pre-sessional courses might be facilitated via the route of metalinguistic written feedback.

2.2 EAP

In section 2.2.1, I examine the concept of EAP and the increasing need for academic English courses tailored for a variety of international students in British universities. Section 2.2.2 considers definitions of autonomy and describes the way in which these have influenced the

goals and teaching methods involved on EAP courses. Section 2.2.3 goes on to specifically look at the role of writing within EAP and examines some of the problematic areas which are created, on pre-sessional courses in particular, due to differing stances on whether organisational aspects of written English or linguistic elements should be regarded as learning priorities.

2.2.1 Defining English for Academic Purposes

EAP is one of the two main branches of what has been termed English for Specialised Purposes (ESP), the other branch consisting of English for Occupational Purposes (EOP), (Flowerdew and Peacock, 2001). Whilst EOP is generally associated with a particular industrial or business area, EAP is the English language in an academic context aimed at aiding learners to study or research at various phases of their academic careers; they may be pre-sessional, trying to gain admission into higher education in an English speaking environment, or in-sessional, meaning they are already studying for foundation, first or post-graduate qualifications. As well as being associated with these traditional levels of achievement, EAP services offered by institutes of higher education are also accessed by those wishing to publish to a global audience, which to a large degree uses English as its lingua franca, and by those who wish to teach or research in English speaking higher education institutes. With a rising number of international students attending British universities in the last few years (Alexander et al., 2008), a predicted increase in these numbers of between fifteen and twenty per cent by 2016 (www.ukgov.bis, 2013), and a growing propensity for traditionally vocational EOP subjects to be handled in academic contexts, there is a complementary growth in the call for research into and understanding of academic English, (Hyland, 2009a).

The broad definition of EAP supplied above is rather simplistic and belies some of the complications involved in describing what it actually is, how it is to be learnt and how it is to be taught. One important consensus seems to be that EAP is different from General English (GE) in that EAP is goal or needs driven (Hyland and Hamp-Lyons, 2002; Jordan, 2002; Watson Todd, 2003; Alexander, Argent and Spencer, 2008; Bruce, 2011). Whilst GE courses and instruction could be described as being centred around an often rather vague notion of progressing from one perceived level of achievement to another, elementary to pre-intermediate, for example, EAP instruction is designed to help students reach a particular tangible point; on a pre-sessional course, for example, learners might work to meet the academic criteria which will allow them entrance to the department in which they will major. Other examples might be preparing a thesis for publication or giving formal academic presentations.

What is taught and learnt in an EAP environment, therefore, tends to be specific to a particular group or individual aim, to be largely inflexible in terms of time and to have an outcome which will have an explicit summative result, which in its simplest terms will be pass or fail. This means EAP courses are based upon knowledge of learner needs and aims, and these will vary from institution to institution, course to course, and learner to learner.

A complicating factor in defining EAP is that it is not considered as one homogenous discipline. Widdowson (1983) originally described ESP itself as being divided into two separate strains: narrow-angle and wide-angle. Narrow-angle courses are those associated with a range of knowledge and skills which are limited to a specific area and where a greater degree of understanding or proficiency might be considered superfluous. EOP originally fitted this description, and Air-Speak, the common language of air traffic controllers and pilots, is an extreme example of this. Wide-angle courses, on the other hand, are those which aim to equip students with a degree of English which will allow them to operate in any given

number of future situations (Widdowson, 1983, p. 6). Narrow-angle courses have come to be associated with Widdowson's term competence, a knowledge of the language system adequate for pre-defined contexts. Wide-angle courses are associated with the term capacity, or the ability to exploit competence in response to a variety of changing contexts (Widdowson, 1983, p. 7) . EAP was originally placed at the wide-angle end of the spectrum as learners might be expected to have to cope with a broader range of undefined situations.

However, since the initial idea of the narrow-to-wide-angle scale took root, EAP itself has been divided into what are now recognised as two related but distinct areas. Jordan (1997) named these English for General Academic Purposes (EGAP) and English for Specific Academic Purposes (ESAP). EGAP, as the name implies, is a wide-angle approach applied to those aspects of English considered academic regardless of subject. These might include the ability to think critically, or to write in a particular academic register. ESAP, on the other hand, is academic English applied in a specific subject area. Jordan (1997) and Hyland (2002, 2009) are amongst the researchers who have argued that there should currently be a greater emphasis on ESAP within EAP teaching. This is based on the premise that a learner who wishes to enter into a department of Mechanical Engineering, for example, is more than likely to have language and skill needs which differ significantly from a learner who is going on to research nineteenth century literature. It is arguable whether capacity is needed for those learners who might go on to a subject which makes particularly narrow use of English, although given the increasing emphasis in EAP on academic culture in general, including the interpersonal relationships between students and staff, (Jordan, 2002, p.73), this might increasingly be an exception to the rule . Whilst some institutions and materials are able to provide ESAP, particularly once students have joined a specific degree course, many EAP practitioners often find themselves in the position of having to deliver EGAP and ESAP to the same group of students. This is typical of pre-sessional courses. As will be seen later,

these two poles have had a significant effect on the way corpora have been approached in EAP.

The wide variety of different motivations for learning EAP and the subsequent differences in approach required, coupled with the differences in approach implied by EGAP and ESAP mean that any one definition is likely to be deficient, particularly when it might be argued that these definitions appear to stem largely from the point of view of instructors rather than learners. However, it has also been argued by Watson Todd in reviewing the methodologies applied to the area (2003, p. 151) that there are some widely accepted approaches towards practice across the EAP spectrum. These are:

1. A focus on inductive learning.
2. Using process syllabuses.
3. Promoting learner autonomy.
4. Using authentic materials and tasks.
5. Integrating technology in teaching.
6. Using team teaching.

None of these are mutually exclusive and Watson Todd (2003, p. 152) argues that practice is likely to be a combination of these. It is important to note that these tenets are not purely language related, but suggest rather an emphasis on the skills and procedures needed to acquire language. Adding to these, Gillet (2011) charges EAP with being responsible not just for language and skills development but also for developing cultural and inter-cultural awareness. My thesis will take into account the perceived importance of inductive learning, the promotion of autonomy in learners, the concept of using authentic language and the use of technology in and out of the classroom. The following sub-sections will look closely at what autonomy is and why it has been singled out as a goal in EAP and how writing is seen as central to EAP practice and the development and goals of learners.

2.2.2 EAP and Autonomy

As with EAP itself, providing an exact definition of autonomy is problematic. As will be seen in section 2.3., however, it is mentioned so often in relation to corpus use and EAP learning goals that concepts of autonomy need to be examined in some detail.

Benson (2001, p. 47) argues that autonomy is multi-dimensional and its definition may vary according to the type of learner, task, stage of development and context. However, there are a number of learner and methodological attitudes which have been identified as key to becoming autonomous.

Holec (1981) outlines autonomy as a series of organising principles adopted by the learner. These involve the learner firstly being able to decide upon the objectives of his or her learning. Following this, what is to be learnt and the methods and techniques to be employed need to be identified by the learner. The procedure and progress of achieving the objectives need to be monitored and finally an evaluation of the process has to be made by the learner. Hedge (2000, p. 84) has pointed out that this is a description of learning as a management process, with the learner undertaking a series of concrete, organised steps. Little (1991) goes beyond defining it as a series of organisational stages and focuses on the cognitive and psychological elements which are needed to progress through these stages, namely the development of the ability to think objectively, to think critically, to make decisions and to act independently.

Whilst responsibility for learning as defined here rests on the student to a greater degree, the cultivation of autonomy is not simply a case of the practitioner abandoning the learner. Both Holec (1981) and Little (1991) concur that autonomy requires an initial structure or environment for its promotion to be successful. Little (1990) had already emphasised the importance of the teacher and is careful to point out that autonomy is not the same as self-instruction.

At the same time, Little (1990) notes that autonomy in itself is not an actual methodology which is applied to learners. Finally, Little suggests that, as with language learning in general, autonomy is not easily defined and that it is not helpful to see autonomy as a final goal reached by all learners in the same way. Rather, it is an approach which, if adopted, augments every stage of the learning process. As touched upon earlier, how autonomous a learner is judged to be may depend upon a variety of conditions; a learner who has the capacity to show autonomy in the way they approach the development of their reading skills may not show the same capacity when it comes to writing. This suggests implications for the measurement of autonomy. Just as the acquisition of a language has interlingual stages, autonomy may manifest itself in a series of regressions and advances rather than a steady line of progression.

The concept of autonomy in EAP itself exhibits itself in particular ways. The EAP course is commonly seen as one in which learning the language tends to be seen not as an end in itself, at least by its practitioners, but as part of a more holistic drive to develop independent academic competence in general. Jordan (2002) has noted that this is evident in the way EAP encourages learners to increase their critical thinking abilities and their awareness of academic culture itself. Beyond the traditional goals of language acquisition and fluency which might be encountered on GE courses, the aim is to create active learners and foster a critical disposition towards both materials, assumptions of the instructor and peers and the language itself (Fletcher, 2004; Perez-Parades and Cantos-Gomez, 2004). This environment for the cultivation of autonomy is often created by placing an emphasis on explorative, inductive learning which follows the 'organic approach' outlined by Nunan (1998). This approach challenges some of the traditional teacher–student relationships and highlights a need for materials which encourage learners to discover the language for themselves. Grammar knowledge and acquisition, for example, might occur via a pattern of

noticing/theorising/testing by the learners, rather than the systems being introduced in the more rigid and traditional present/ practice/produce form which places the instructor in the position of the font of knowledge. It has been suggested that the mental process of discovering the meaning of language is as important as the actual language itself (Thurston and Candlin, 1998, p. 278). As Little (1991) has pointed out, this does not involve the teacher disappearing. Rather, it involves a shift in role towards that of the facilitator. By providing a framework which promotes a culture of hypothesising about language and meaning, EAP courses aim to equip learners with both the tangible aspects of the language and a level of autonomy which will "ensure continued learning beyond the lifespan and location of EAP instruction, whether classroom or self access based." (Alexander et al, 2008, p. 271). Fostering this mentality has been identified by Banerjee and Wall (2006, p. 57) as being particularly important for pre-sessional students going onto a post-graduate course because they tend to struggle with the emphasis that is placed on learner-independence within the departments themselves.

The connection between autonomy and inductive learning is an important one as early studies by Johns (1991) suggested that Data Driven Learning (DDL), including the use of corpora and concordances in the classroom, provided an arena in which to develop both. The following section explores the role of writing in EAP. This includes examining the ways in which the concepts of autonomy discussed above have been incorporated in writing pedagogy and will also look at areas which have been identified as problematic.

2.2.3 Writing in EAP

Arguably, EAP has a tendency to place a greater emphasis on reading and writing skills than those of listening and speaking. Alexander et al. (2008, p. 178) claim it is the most important

of the skills because successful production of written text acts as the "currency" which admits entry to higher education. Subsequent assessment of writing at all stages of study mean that competency in academic writing is at the core of most learner needs (Bruce, 2011, p. 239). The type of writing varies considerably given the needs of different students. EGAP and ESAP writing needs may differ widely. Likewise, undergraduate and postgraduate courses are likely to cover different areas of writing. Below is a list, not by any means exhaustive, of typical writing genres which EAP might have to cover:

Critical essay	Dissertation	Business report
Examination essay	Project	Poster presentation
Case study essay	Lab Report	Team assignments
Literature review	Research Proposal	Management brief
Reflective assessment	Case notes	Website content

Figure 2.1: *EAP writing genres. Reproduced from Alexander et al (2008, p. 181).*

Meyer (1996) has proposed that because of this variety, EAP practitioners, particularly those involved on pre-sessional courses, may find themselves in a "disciplinary vacuum" (p. 34) where they are trying to cater for many different learners, often without a knowledge of the eventual discipline specific target language and writing skills those learners will need on their eventual academic courses. It is unlikely that all of these genres will be covered over a course of study. This has led to the increasing adoption of structural move analysis as outlined by Swales (2005), where potentially useful or important rhetorical systems are taught to students. Published materials might break language down into a series of smaller rhetorical moves which may later be incorporated into a whole. Thus, learners become familiar with how to organise, for example, cause and effect patterns, or problem solution

patterns, or argumentative patterns in the hope that they will be able to press these into service into specific contexts once they finally enter their chosen academic department. At an EAP level, learners also have to become acclimatised to academic registers and academic style.

Again, precisely what constitutes academic style varies. According to Biber's corpus study of register used in university language (2006), this variation is multi-dimensional and depends on factors such as purpose, discipline and context. As already mentioned, pre-sessional courses, for example, have a particular purpose and may have definitions of academic register which differ from those in other areas.

Bruce (2011, p. 240) has noted that EGAP courses, meaning those aiming to increase the general, rather than subject specific, language knowledge and skills of students require this kind of wide-angled take on writing as they seek to develop capacity, that quality mentioned earlier as one in which the learners can respond effectively in a variety of situations.

Broadly speaking, increased competence in academic writing is described as a progression through various rhetorical functions and moves from knowledge telling to knowledge transforming. At the knowledge telling end of the spectrum, learners are able to operate on a level of description, using narrative, spatial, instructional or sequential organisational patterns. Progressing, learners might move on to the macro function of explaining, of which micro-functions might include definition, comparison and cause and effect. Finally, the macro stage of persuasion is reached, which is composed of the micro functions of resolving problems with solutions, arguing, evaluating and providing evidence and conclusions. Of course, progression to the persuasive stage does not mean the description phase is abandoned, but it is possibly the most cognitively complex macro stage of the three.

How to learn and teach these moves varies but can be seen as falling into two broad but identifiable categories – top-down and bottom up. Often, reading materials serve the function of models for written work. This is a top-down, product based approach with students attempting to learn from an example genre which serves as a means of giving students prior or schematic knowledge of a form before they attempt reproduction (Hedge, 2000, p. 195). An alternative method employs a bottom-up approach. This implies students drafting and redrafting texts according to particular linguistic elements and systemic knowledge associated with a specific rhetorical move and gradually combining these to produce an example of a particular genre of writing (Hedge, 2000, p. 189). These methods are not mutually exclusive but a way of entry into understanding or using the language often implies adopting one or other of the approaches initially (Flowerdew, 2009, p. 402). Either of these approaches may involve the further breaking down of texts into finer micro aspects such as the use of particular discourse markers or common patterns, general to specific, topic sentence to supporting sentences for example. Hamp-Lyons and Heasley (2006) present language in the context of a particular function. For example, writing about processes involves instruction in the passive and nominalisation (2006, p. 89-99) and this is presented under the genre of the methodology aspect of a text. Giving definitions presents relative clauses (Hamp-Lyons and Heasley 2006, p. 53) whilst the genre of a discussion section looks at the language of argument, with the text book providing a series of set phrases to report on positive and negative aspects of a study. Jordan (1999) also attempts to contextualise language items as elements which might be helpful as part of a greater rhetorical context.

Archibald (2001) has noted that pre-sessional courses in Britain largely focus on the discourse and organisational aspects of writing, rather than the language per se. In his experiment, in which he tested the effects of pre-sessional course instruction in writing

proficiency, only two of the seven traits of assessment related directly to proficiency in the target language itself: (i) linguistic accuracy, defined as grammar, spelling and punctuation which did not impede communication and (ii) linguistic appropriacy, defined as the strength of grammatical and lexical choices. The remaining five traits related to organisation and structure of the text. Over the course of eight weeks, he found that improvements in linguistic accuracy and appropriacy were statistically less likely than improvements in any of the other areas (p. 162). Cho (2003) conducted an experiment in which proficiency of pieces of writing made under test conditions and those made after lengthy instruction were measured. Again, it was the linguistic aspect which was statistically less likely to show any improvement (p. 179). Basturkmen and Lewis (2002) reported that teachers administering a twelve week pre-sessional regarded improvement in grammatical accuracy and linguistic proficiency as an unrealistic aim and so emphasised the organisational and critical thinking elements of the course rather than learner proficiency in the language.

There seems then for there to be some justification for the emphasis of structure and organisation over the language itself on EAP pre-sessional courses; the time available and, in the cases mentioned above, the lack of a significant linguistic improvement. Banerjee and Wall (2006) have noted that, in the long term, departmental supervisors do not necessarily value accurate language over aspects like content and correct referencing.

The weighting towards forming the correct organisational patterns above all else has attracted some criticism, however. One issue is that there may be a discrepancy between how important teachers, and perhaps by extension the syllabus, consider wider rhetorical functions and learner perception of their own needs. These are often more linguistic in nature (Burgess and Etherington, 2002). Hartshorn et al. (2010) have also suggested that whilst one of the main goals of English teaching is to increase levels of linguistic accuracy, intensive courses often find themselves not addressing aspects of proficiency. Turner (2004) has

argued that the precision and accuracy of language production itself has suffered at the expense of the emphasis on being able to communicate within these rhetorical frameworks and that this has a subsequent negative effect on the content itself, potentially undermining learner produced texts as a whole. Firstly, Turner suggests that perhaps the reason that linguistic elements are overlooked on courses and in departments is because "academic success is rarely attributed to good language use" (p. 99). However, as she points out, whilst good language use goes unmarked and unnoticed, bad language use is visible and creates issues. She argues that EAP instruction in writing concentrates on the macro-level without realising the importance of proficiency at a micro-level. Whilst a student might be academically aware and capable of critical thinking, their potential for expressing arguments effectively can be undermined by their poor language knowledge. Thus she concludes that because students need to be able to manipulate their language in order to express arguments and rhetorical functions with any hope of success, "language proficiency is as important as content knowledge" (p. 104). In particular, she suggests that increased understanding of lexis will help learners to operate at Widdowson's (1983) *capacity* level in order to negotiate the varied types of academic discourse they will be exposed to.

A further issue is the extent to which autonomy can be developed in written work. Alexander, et al. (2008) argue that self and peer correction of writing are elements which encourage independent, evaluative critical thought in learners, but at the same time stress the role of the teacher in making stages of progress explicit to learners (p. 189). Whilst learners, particularly at postgraduate level, will be judged on their ability to produce a piece of independently written work, on EAP courses and even in their in-sessional stage they may need a substantial amount of teacher and classroom centred support. The relationship between teaching effective writing skills to learners and enabling them to develop a sense of autonomy rests, therefore, on a delicate balance.

2.2.4 Summary

EAP, and particularly EGAP, has a number of objectives which exist at a concrete, linguistic level and a cognitive level. Practitioners are on the one hand expected by their institutions and their learners to deliver instruction which can show explicit development of written competence through the ability of learners to produce ever more sophisticated academic texts covering a range of academic genres. They are expected to guide students to and through the appropriate materials and language. Given the short nature of many courses and the potential issues regarding EGAP and ESAP needs, lexical and linguistic proficiency are areas which are often overlooked in favour of broader knowledge of organisational structure. In addition to development within the classroom, instructors at EAP level are charged with the task of encouraging learners to adopt a less teacher and support based learning style and to develop independent learning skills, not necessarily language focused, which will aid them in their studies within their eventual program of study.

The following section investigates the place of corpora within EAP, both in terms of research informing our knowledge of academic writing and learner writing and in terms of the classroom and learners themselves. It goes on to suggest ways in which the use of corpora may help to aid in linguistic proficiency and promote elements of autonomous learning.

2.3 Corpora

Section 2.3.1 describes the influence of corpus research on EAP. 2.3.2 identifies the benefits of using corpora with learners in the classroom, whilst section 2.3.3 examines the problems associated with using corpora.

2.3.1 Corpora in EAP

Perhaps one of the most firmly established uses of corpora is in the development of materials. Specifically, corpus research has been used to identify which linguistic token or combination of token should be taught and learnt. COBUILD (Collins Birmingham University International Language Database) is a prominent example of a corpus project which has gone on to directly influence the content of materials, courses and the classroom; this collection of modern texts resulted in the compilation and publication of the Collins COBUILD English Language Dictionary (Hanks, 2009) which in various editions has been a feature of classrooms since 1987. The perceived advantage of this type of dictionary is that its content is based on authentic language use, although Frankenberg-Garcia's study (2012a) brings into question the potential effectiveness of the presentation of this information. These dictionaries are based upon an objective examination of the frequencies of words and collocations rather than the intuition or subjective notions of experience writers. As such, they might be considered as a more realistic, accurate representation of the real language learners need.

In the same way, corpora have influenced the design of various text books. Swales (2002, p. 152) remarks upon the differences between the information presented in his own pre- and post-corpus informed books. He found that some of the rules and examples expounded in his earlier text books were, when compared with substantial corpus data, presenting learners with an unrepresentative sample of the language. This deficiency was attributed to the fact that the language previously included in his text books was based more

upon writer instinct than any tangible evidence of actual usage, a practice also noted by Biber and Reppen (2002). Other studies corroborate this. Thursturn and Candlin (1998, p. 270) note that examination of concordance data has highlighted that some of the rules in language books contradict the actual use. Romer (2004) notes the same, finding for example that idealised form of 'if' clauses prescribed by textbooks and learnt wholesale uncritically by students is at odds with the actual use shown in concordances. Yoon (2011) has also made the distinction between the authentic language used in real world contexts and that found in the fabricated and contrived contexts of textbooks, presenting the former as more desirable in terms of accurate exposure to the language. As such, textbooks are increasingly informed by corpus data. McCarthy and O'Dell (2008), Cullen (2012), Hewings (2012) and Hewings and Thaine (2012) and all draw upon the Cambridge English Corpus and the Cambridge Learner Corpus to identify both the target language employed in texts and exercises and some of the errors which learners might regularly encounter.

Another influential tool which has emerged from the examination of corpus data has been Coxhead's (2000) Academic Word List (AWL). This was generated from a corpus of 3.5 million academic words and presents the 570 most common word families organised into nine separate groupings. The fact that it was generated from a corpus of academic texts may be seen as a way of narrowing the divide between EGAP and ESAP needs on the wide-angled courses mentioned above by virtue of it equipping learners with a broad range of lexical items. It has directly influenced dictionaries, textbooks and online materials in use at the moment (Coxhead, 2011). Whilst the AWL has drawn criticism for both the underrepresentation of certain lexical items within particular subject areas and the misrepresentation of items within some co-textual scenarios (Hyland and Tse, 2007), it continues to inform modern EAP publications. McCarthy and O'Dell (2008), for example,

employ the AWL alongside corpus derived target language, whilst Schmitt and Schmitt (2011) use the list to generate the core of their vocabulary text book.

Other research has involved the production of smaller corpora encompassing a more specific range of language. Krishnamurthy and Kosem (2007) have produced a quite extensive catalogue of specialised corpora created for academic research purposes, many of which are no longer available for viewing and were designed for specific research projects or institutional use. These are often attempts to create more specialist corpora which might be of use to ESAP practitioners. The Gledhill Corpus from1995 consisted of 0.5 million words generated by the compilation of 150 cancer research articles, whilst the Hyland Corpus (2000) of 1.4 million words encompasses a wider range of disciplines (8) but restricts itself to 240 academic journals. Corpora have also included those built from postgraduate learners' written texts created by native speaker students, like MICUSP, the Michigan Corpus of Upper-Level Student Papers (2009). Through the examination of these, researchers might be able to identify subject specific language and structures which may then be used to inform discipline specific materials.

Another area of research has been in the creation of learner corpora. By collating learner texts, researchers can examine corpora not only for accurately produced language but may be able to examine the frequency of particular errors. This, again, may inform materials and approaches to be used in class which target those errors. Granger (2003) has suggested that a comparison of native speaker corpora with those of learners can highlight discrepancies in language use and may have applications in syllabus and materials design, testing and methodology. Leech (1997) also argued that analysis of errors found in corpora, or items over- and under-represented, could provide a guide for addressing these issues in the classroom, and Huttner (2010) successfully produced a genre analysis of corpus data which consisted of student and expert written examples of conclusions, with discrepancies being

found at both macro and micro-levels. Tan (2005, p.133), however, has cautioned that whilst the examination of learner corpora for errors may provide worthwhile insight, the current and increasing prevalence of localised Englishes means that authoritative judgments on what constitutes authentic English need to be taken with great care.

2.3.2 The Perceived Benefits of Using Corpora with Learners

Just as corpora have been influenced the production of materials and research outside the classroom, there are arguments that they have a number of applications in the classroom. Perhaps one of the more obvious ones is an extension of that used by researchers in the compilation of specialised corpora. Cobb (1997), for example, created a corpus for his students from reading texts from their course which enabled them, through the use of a Computer Aided Language Learning (CALL) suite called PET.200, to examine the key words they needed to improve their vocabulary. In terms of this language acquisition, "a small but consistent gain was found for words introduced through concordances" (p. 301). There is also the potential for learners or researchers to create their own corpora. Freely available concordancing programs such as Wordsmith Tools (Scott, 1996) and AntConc (Anthony, 2010) mean that researchers, teachers and learners might be encouraged more to create their own corpora. In the case of learners themselves, these could be constructed from examples of their own written work. They could also be created using texts from the disciplines the learners eventually aim to study. In doing so, it might be possible to overcome that particularly problematic area encountered on EGAP courses that was mentioned earlier, namely the inability to cater for a number of disparate disciplines. Lee and Swales (2006) showed that learners who built their own specialised medical and education based corpora were, with a significant amount of training, able to use these corpora to "inform their own writing" (p. 68). There is the potential, then, for learners to develop their own resource bank

of lexical and grammatical structures relevant to their own language specific language needs, something which Renouf (1997) has identified as "teaching to establish resources". Autonomy and corpora will be dealt with more below, but it seems appropriate to reiterate that one of the goals of EAP earlier defined was to equip learners for later study, and this seems to offer a concrete example of that.

In terms of more generalised, widely available corpora, one of the apparent advantages for learners, as with researchers and materials designers, is the access to potentially more authentic language than that shown in text books and other traditional classroom materials. The sheer volume of the language available is also potentially beneficial. Whilst traditional materials might be informed by corpus research, physical constraints of size dictate the amount of information which can be incorporated into paper dictionaries and text books and online corpora are able to display a comparatively large amount of text information, running into millions of tokens. The type of information available is also different in nature. In corpora, issues of context and co-text can be addressed almost simultaneously. On even a simple word search using the Brigham Young University British National Corpus (BYU-BNC) interface (Davies, 2004), learners are able to see the frequency of a word, its context in the source references and concordances, and its co-text in the concordances. With a few other relatively simple steps, learners can also use the interface to examine the corpus for various types of collocation, synonyms, comparisons of synonyms, KWIC (Key Word In Context) displays and visual displays of genre. Thurstun and Candlin (1998, p. 278) have claimed that the examination of this "rich context" gives a great degree of scope for students to enhance their awareness of the language they are studying. One of the aims of this thesis is to establish how this might actually be shown to be the case and how it may be measured, as reflected by research questions RQ1 and RQ2 outlined in subsection 1.6 of the introductory chapter.

The fact that language can be investigated and manipulated like this using corpus tools means that they have become central in the drive for autonomy. Advocates such as Harmer (2001), Stevens (1995) and Thurston and Candlin (1998) have suggested that the inductive learning which goes on when looking at concordances means the learners are potentially at the centre of an inductive learning process which is highly motivational. There is the potential for them to interact with the language independently of the teacher, to formulate their own hypotheses about the language and generally to become involved in learning as research. Benson (2001, p. 138) has claimed that any potential control the learner has over the language promotes not only linguistic awareness but also encourages the development of metacognitive skills which might be applied to other fields of study. The use of corpora may encourage the use of those elements of autonomy, critical thought, independent thinking, and decision making that were outlined above as part of the remit of the EAP course in section 2.2.1 . Again, measuring the benefits of this potential control is part of the gap in knowledge I investigate within this thesis.

2.3.3 The Potential Problems of Using Corpora

There is not a "two-way street" (McCarthy, 2007, p. 565) between research and pedagogy; academics shape the way corpora are used by suggesting a theory rather than demand from the classroom dictating suitable or useful approaches and this results in a discrepancy between the supposed benefits of corpus use as seen from controlled studies and the potential benefits gained amidst actual realities of EAP courses.

One of the current problems facing EAP practitioners and students is the lack of familiarity with corpora. McCarthy (2007) has noted that whilst teachers may be acquainted with the terminology, they do not fully understand the application of corpora and have fears about the possible implications of their use. It seems that of particular concern is that they

will need more highly developed computational or linguistic skills than they currently possess in order to take advantage of them, something also pointed out by Harmer (2001, p. 175) who suggests that complicated software and the format in which the language is displayed has "deterred many teachers and students from going anywhere near them." Frankenberg-Garcia (2012b) has also noted that the success of student interaction with corpus materials is less likely when teachers are themselves unaware of how to approach or use corpus data. A survey by Jarvis (2004) suggested that only about ten per cent of UK EAP courses incorporated the use of corpora. Part of this problem seems to be that there is a distinct gap between the theory of corpora use and its practical application. He asserts that the idea of computers being not just an accessory to learning but informing pedagogy itself is already well established but that actual use of IT equipment can be problematic. Jarvis' research reported comments from teachers avoiding the use of computers which included a lack of physical space, a lack of hardware, and a lack of time. The use of ICT (Information and Communications Technology), some concluded, is better outside rather than inside the classroom, where face-to-face interaction is the important thing. Nearly 80% of teachers showed little interest in delivering computer based lessons and over 40% thought that the institution they worked for had no expectation of computers being used on a course. This is despite evidence to the effect that almost 70% of EAP providing universities believed that students arrived in the UK with a working knowledge of computers, although not necessarily for the purposes of language study. This points again to a potential gap in instructors' beliefs of what the learners need and are familiar with and how the learners perceive the situation themselves; 95% of students surveyed expected to be using computers as part of independent study. Jarvis suggests that the key to more successful incorporation of computer based learning involving material such as corpora is that the tasks provided are linked to the syllabus and made relevant to the context of the learners' own academic context and needs.

Whilst it has been noted above that the potentially vast amount of language data available for use by practitioners and learners is a boon, it has been shown to generate its own problems. Mauranen (2004) examined the use of corpora in the development of spoken English in the classroom and identified that one of the key difficulties found was that, when trying to employ corpus based materials, the sheer volume and range of information available to the teacher meant that the amount of time and energy employed in trying to devise entirely new exercises was significantly more than that involved in using other teaching methods (p. 200). This is echoed by Breyer (2009) who again notes the discrepancy between research activity and pedagogical use and the problems teachers encounter in putting what academics have hailed as enormous amounts of valuable data into practice. In her study, she reported that less than 40% of teachers had been able to find corpora or data from corpora which could be applied to their class objectives (p. 162). Perhaps more significantly, Breyer noted that the chief fear of teachers was that, by using corpora, their authority in the classroom would be undermined. That is, whilst authentic language has been regarded by some researchers as a virtue of corpora, in a classroom situation it may not necessarily correspond to the rules (which may actually have been previously taught to learners) found in text books and grammar references. This contrast was highlighted as a major reason for teachers being unwilling to make use of corpus based materials and activities in class. Breyer (2009) points out that this concern illustrates a key obstacle for the acceptance of corpora; more traditional, entrenched and widespread pedagogies involve the teacher, or syllabus, identifying the target language from within language as a whole first and subsequently engineering teaching materials which work towards the acquisition of that target language. Examination of the scheme of work pages of many EAP textbooks, even those originally informed by corpora, will show just how prevalent this approach is (Cullen, 2012; Hewings, 2012). This is, of course, at odds with the ideas of autonomy and language discovery which, as has been shown,

theorists have hoped to develop through corpora, as corpora and concordance data start with the language itself as the focal point from which the target language is acquired by careful examination and testing on the part of the learner.

Students of English have been found to experience similar negative feelings to teachers when confronted with corpora. 'Concordancing Burnout' (Lee and Swales, 2006, p. 57) is a problem associated with the amount of information which is presented to learners in the form of corpus interfaces and corpora. Just as teachers or instructors might have to learn new approaches and skills, so students may have to adopt an entirely new system and way of thinking about language and learning itself. Thurstun and Candlin (1998, p. 278) noted that asking students to hypothesize on language usage and rules from concordances can have the negative effect of tiring them. The act of learning inductively places an increased cognitive load on the learner and this can be seen as a negative affective factor. Though inductive learning might have long term benefits in terms of language acquisition and attitudes towards autonomous research, it may not be immediately obvious to a student why they are being asked to work through lines of concordances in order to formulate their own rules when text books and grammars can supply it to them in what they might perceive as an easy to digest format. This could lead to a loss in motivation and self-confidence.

In direct contrast to one of the main perceived benefits outlined earlier, Sun (2007) recorded rather that one of the main weaknesses of corpora from his students' point of view was the volume of data available to them and the fact that any effective use of it required a shift away from a more comfortable, deductive learning mode. This echoes research by Frankenberg-Garcia (2005, p. 194), who reported that the use of mono-lingual concordances resulted in learners complaining that "there were too many [concordances] and it took too long to read them all". Similarly, whilst learners in Yoon and Hirvela's (2004) investigation into EAP student attitudes towards corpus use could identify some of the benefits, such as the

display of collocations of prepositions, for example, an issue amongst some participants was that it simply took too much energy to extract the relevant information from the concordances, especially when compared to using a dictionary (p. 274). The abundance of information available to users can, rather than being seen as a potentially rich mine of information, be interpreted as noise (Dudeney and Hockly, 2007, p. 109) which can be so distracting that the benefits of using corpora are lost on both learners and practitioners. Again, this could partly be due to the range of learner interests and needs; it may be that students who are going on to study corpus linguistics are more receptive to the use of corpora for their own language work than engineers or mathematicians, who may not see them as having any future application.

These problems bring into relief one issue which questions the desirability of autonomy itself, at least from a learner's point of view. There are problems to be encountered when the teacher has a reduced amount of control and when the information in a corpus contradicts that which is displayed elsewhere (Hunston, 2002, p. 176). In devolving authority about the language to a corpora and responsibility for learning to the learners, practitioners may be presenting an entirely new and, based on previous educational conditioning, undesirable mode of learning (Harmer, 2001). Benson (2001) has suggested that contact with the concept of autonomy, as defined as learning which is not completely centred on the teacher, may initially be quite traumatic. Others (Ho and Crookall, 1995; Jones, 1995) have argued that if the concept and objectives of autonomy are not managed carefully, then it could be described as a type of cultural imperialism when employed with learners from a system where it is non-existent or largely marginal. Even when taking into account the counter argument (Pierson, 1996; Yang 1998) that it is a mistake to judge the autonomous capabilities of learners based on their previous educational and cultural experience, introducing the use of inductive learning through corpora might be accepted much more readily by some learners than others. There is also the argument that only certain

levels of learners, typically advanced rather than lower levels, are able to derive any great benefit from learning with corpora (Kenning, 1996).

Finally, institutional constraints need to be considered. Time is a restrictive factor. Adapting to new technology and methods can take time for both learners and teachers, and Johns (1991) admitted that training in software can be a barrier towards using corpora and concordances. Cobb (1997) further noted that the time needed to train may not be available anyway or might be better spent in other areas. The very nature of concordances in the way they have been discussed as part of autonomous learning means that realistically computer, and probably internet, access is required to use them and this is likely to be most effective when individual students have access to the technology. Admittedly, concordances can be printed out from a computer accessed by the instructor and presented to learners for inductive learning, but this removes some of the potential access to wider autonomous language exploration. Whilst most British institutions are likely to have IT systems and computer access for students in some form, practical considerations remain; there might be the logistical issues of large class sizes, many classes being dependent on just one computer lab, if the correct software is accessible, if the internet connection is fast enough for searches to provide information relatively quickly and how long the course is. All these have a bearing on deciding if corpora can be used on an EAP course and, secondly, how much they should feature within a course and how they can be integrated with the other parts of the timetable and the other methodologies and materials.

2.3.4 Summary

It seems that current research into corpora could be described as following one of two paths. From a research perspective, that is, outside the classroom, corpora are relatively well established as tools for selecting the type of language which should be used in reference

materials. They have also been identified as tools which can be employed in bridging the divide between EGAP and ESAP through the manufacture of specialist corpora which can be used with ESAP learners. Studies have also shown the potential for both the compilation of learner corpora by researchers with the aim of identifying common errors or discrepancies with expert or native speaker writing, again with a view to informing the materials subsequently used in the classroom, and the construction of specialised corpora by the students themselves for analysis of the specific language they need.

The second path is perhaps a more obstacle strewn one and involves the actual use of corpora as an explicit course tool. In terms of benefits, the use of corpora and concordances in the classroom has been identified as something which supports the development of learner autonomy. The stress on inductive learning and the development of cognitive processes which turn the learning process into a research process is seen as one of the great potential boons of corpus use. The amount of real language available in corpora has also been held to be of a benefit to both students and practitioners, giving them access to more authentic language than traditional materials. At the same time, exactly the same factors have been shown to be potential drawbacks. Inductive learning places a greater stress on learners and can result in the creation of negative affective factors such as feelings of confusion and a resulting loss in motivation. The inductive, autonomous approach to learning is also something which may cause serious difficulties for learners form particular backgrounds and of particular levels of English knowledge or competence. Additionally, the amount of readily available authentic language can actually inhibit the use of corpora, with practitioners expressing nervousness at the thought of their position as expert being challenged and the learners being overwhelmed by the information available.

The next section will examine the ways in which corpora have actually been used in the classroom and will examine how effective the use of corpus data and concordances has

judged to have been in benefitting learners and addressing some of the potential problems mentioned above.

2.4 Corpora in the Classroom

Subsection 2.4.1 looks at examples of corpora being used in the classroom as either research or reference tools, both of which are defined below. Section 2.4.2 examines the extent to which the practitioner needs to mediate between corpus based resources and the learners and what forms this mediation can take.

2.4.1 Corpora as Research or Reference Tools

Hyland (2003) has described the use of corpora in the classroom as falling into two different categories, research tools or reference tools. The use of corpora and concordancing as research tools involves raising linguistic awareness of rules and underlying patterns by exposing them to concordance data and allowing them to make and test their own hypotheses. As a reference tool, corpora and concordances are used by learners to deal with particular language problems as and when they are encountered in place of or alongside more traditional forms such as dictionaries. Following the example of Yoon (2011), the following examination will separate studies into these two categories.

Turnbull and Burston (1998) conducted an experiment in which two learners were assessed on their ability and willingness to make use of concordance searches to investigate problematic areas in their own writing. Using a small corpus derived from corrected versions of their own written work and after a short training course, the students were left to use the corpus software as often as they wished, making language explorations independent of supplementary instructor intervention. A qualitative piece of research, the main finding from this study was in the attitude the learners expressed towards using the corpus and what this

may say about the effectiveness of the method with different learner types. One learner, already comfortable with independent, inductive work, was happy to make use of the corpus and found she had benefitted from analysing the concordances. The other learner, however, was of a type who favoured a teacher-centred, deductive mode of learning and found that corpora and concordances provided little benefit and were a source of frustration. The authors concluded that the effectiveness of corpora as a tool for language learning and promoting autonomy may be reliant both on the type of learner and the type of learning style the learner has developed as a result of previous experience. A study by Cresswell (2007) seemed to concur with this. Two groups of students were examined, one with access to a corpus and the other without. The students were allowed to explore patterns of use of various connectors, such as 'on the contrary', before giving a meta-linguistic account of the rules and usage governing the target language. A quantitative measure of effectiveness was whether or not the learners' rate of accuracy in the use of these connectors in their subsequent written work was higher in the group using the corpus. The results were largely equivocal, providing no conclusive evidence to suggest that one group had performed better than the other. Again, though there were improved accuracy rates in the group using the corpus, these were attributed only to a small number of the learners, leading the researcher to posit that individual learning styles may have a strong influence on whether or not corpus tools can be effective.

In the 2006 study by Lee and Swales already mentioned in section 2.3.2, graduate students were trained carefully in the use of corpora over the period of a semester. At the same time, the students were asked to compile two corpora. One of these was a corpus of expert writing in the field, the other the students' own writing. The learners then compared these corpora inductively, trying to identify discrepancies between the use of language between the two and identify problem areas in their own language. The sample was very

small, but the students reported that having used the corpora, they felt an increased level of confidence in being able to produce accurate texts without the aid of a native speaker. Although there was no statistical evidence to suggest a measurable improvement in the use of the target language, the researchers decided that the use of corpora had decentred the language, meaning that sole authority did not now lie just with the teacher and in the classroom but was something which could be investigated by learners themselves, and that this decentering promoted autonomy.

The previous three studies resulted in largely being awareness raising exercises for the learners. Cobb (1997) looked at using concordances as a vocabulary research tool for language acquisition. Two groups of students engaged in a series of computer generated tasks designed to help learn vocabulary. The difference between the two was that the experimental group had an inductive learning task to complete using concordance lines which the other group did not. The students were then given spelling tests and text gap fill tests. No significant difference was noted in the accuracy of spelling between the two groups, but when completing the text exercise, the students who had consulted concordances as part of the course showed a 12% higher success rate. Whether or not this was statistically significant given the sample size and other variables is unclear, but it does point towards a possible benefit of using corpora in terms of language acquisition.

Studies involving the use of corpora and concordances in language teaching as reference tools have thus far been more common than as research tools. A reason for this may perhaps be seen in the similar research methodologies shared by many of them, which involve a collaborative effort between the practitioners and learners upon which stricter controls can be placed than in the examples involving the use of corpora as research tools.

Watson Todd (2001) presents a study in which Thai students were asked to produce texts and then hand them to the researcher who then highlighted where errors had been made.

The learners then selected one of these errors and investigated it online using ten concordance lines, from which they inductively drew their own conclusions as to how to correct the error. The study found that 72% of the corrections made were accurate, although unfortunately there is no control group to draw a comparison with so it is unclear as to whether the correction was the result of the information supplied by the corpus and the process of inductive investigation, or merely the awareness of the error and prior-knowledge based correction. As will be seen, this lack of differentiation between different correction methods amongst learners who have the option of using a corpus is something which this thesis attempts to address.

Gilmore (2009) produced a study with a similar mechanism in which students were asked to produce a text which was then handed in and had the errors highlighted. Students then used corpora to make corrections within ninety minutes. Again, the study lacked a control group, prompting the same question about how the errors really were corrected, but the researcher reported successful correction rates of 61% and a positive attitude towards the corpus based exploration from the students.

Gaskell and Cobb (2004) made a study in which students were trained in corpus use and then asked to correct their written errors using corpus consultation. At first, access to corpora was heavily regulated by the researchers as students were supplied with a particular URL which took them to a set of pre-defined concordances. Later, this feature was removed and students performed independent searches based on the errors which had been marked up. The overall results again show a favourable success rate. Accuracy rates were initially 80%, dropping to 60% when the URL signpost was made unavailable, but then rising to 70% as learners became familiar with the concordances towards the latter part of the experiment. However, when pieces of work outside the experiment conditions were assessed, in this case pre- and post-course timed writing, there was little evidence that the use of corpora had

affected sentence level error rates. Corpus use, whilst being of some use in the actual revision phases, had not necessarily helped to affect language acquisition in the long-term. This may be a result of the fact that the number of iterations was limited by the length of the study.

Chambers and O'Sullivan (2004) again used an error based system with postgraduate learners. After an initial corpus training period, learners were asked to correct their work using concordances as a reference point. 75% of the corrections made were accurate. In a follow up experiment (O'Sullivan and Chambers, 2006), undergraduate students underwent the same process and a similar result was observed, with approximately 73% of corrections made being accurate. One of the conclusions drawn was that corpora seemed to have been particularly useful in correcting idiomatic phrases and prepositions and it was suggested that it was unlikely these errors would have been corrected as effectively using the dictionaries and textbooks available, although, once again, the lack of a control group or any statistical evidence for whether other methods of correction might have been just as effective means conclusions must be drawn tentatively. Sripicham (2004) employed a similar method where the use of corpus data was applied to students' individual work in order to address their errors and reported a positive reaction from the learners.

It should be reiterated here that it was not always clear in the above studies how exactly, or if at all, the researchers had tried to differentiate between a correction made by the learner on the basis of corpus evidence, or if they had simply corrected the error through knowledge after having it pointed out to them. Although Gaskell and Cobb (2004) employed a system which disclosed if a concordance had been consulted, they did not examine the frequency or effectiveness of these searches in relation to other reference tools. This seems to be an important area to address if research is going to highlight accurately the extent to which corpora can aid in written accuracy. Likewise, perhaps more information about the

actual type of errors being corrected and the success rate would be helpful in understanding the usefulness of corpus data for particular grammatical or lexical areas. These concerns form the basis for research questions RQ1 and RQ2.

The evidence from these studies suggests that corpus use is beneficial in several ways. Both research and reference practices have suggested that student autonomy can be promoted. Turnbull and Burston (1998) and Cresswell (2007) have demonstrated, however, the need for caution in assuming that the use of corpora is a magic bullet which will turn students into successful independent learners, as learner type and background seem to have a significant effect here. Yoon and Hirvela (2004) also urge caution and found that whilst learner reaction on the whole was positive, criticism of corpora in terms of complexity and time taken in use was by no means absent. Research and reference practises also suggest that immediate language awareness may be affected positively by the use of corpus data, particularly it seems in a framework which allows learners to work inductively on their own work. This seems to work particularly well when students are directed towards correction activities, but it is unclear how much proficiency is affected as a whole in other pieces of writing. Sun (2000) actually suggests that whilst students perceived the access to authentic language through corpora as a benefit, they held a largely negative view in terms of what it would do for proficiency and language acquisition.

Training learners has so far only been touched upon and this is because the idea of training and mediation between the practitioner and the student is a large area and constitutes the following section. However, it is clear from the description of the studies above that learners received some form of training and, in the form of the draft/feedback/correction method, a form of mediation of the data. The importance of this is outlined below.

2.4.2 Mediation of Corpus Data and Activities

Yoon (2011) has proposed that the high rate of success in error corrections in Watson Todd's study described above (2001) may be largely influenced by the researcher guiding the students towards their searches and limiting the concordance lines to ten, thus avoiding that fatigue and confusion mentioned in section 2.3.3. In the other studies, it seems that key factor in a beneficial experience for the learner is that some form of initial or continuing mediation from the instructor needs to be present. That is, the assumption that just giving students access to a corpus interface and expecting them to use it effectively is unlikely to produce a happy outcome. As already intimated, this may be why convincing quantitative data has been difficult to extract from studies using corpora as less-guided, and therefore less controlled, research tools.

It has already been established that new software and concepts can be daunting to learners and so, not surprisingly, the studies mentioned in section 2.4.1 all employed some form of initial training. This varied from brief introductions to the concept of corpora and a particular interface to relatively lengthy training programs which examined in detail particular functions or strategies. This could be regarded as the primary phase of mediation and may be quite heavily teacher, rather than learner, focused. At this stage, Charles (2012, p. 100) notes that it is crucial that the practical benefits of using a corpus are made clear to the users, if only for motivational purposes. Students with a lack of interest or no understanding of the purpose of the activities or tools they are using are as unlikely to benefit from corpora as they are from any other type of learning. Beyond this initial training, however, varying degrees of mediation can be provided.

Gaskell and Cobb (2004) suggest that there are three principles in employing a corpus based grammar resource. The first is that its use should be applied to areas that learners have already shown a need for rather than in presenting new ideas. Secondly, learning should be

driven by the data rather than by declarative presentation of rules. That is, the repeated examples of the language itself should form the basis for inductive learning. Finally, concordance data, at least in the early stages, should be selected by the practitioner and a period of training should be offered. All these point to substantial mediation of the corpus data available. The first principle suggests that corpora are used to investigate language that the practitioner or learners have already identified. This implies the corpus will be approached with a concrete, specific idea of the language to be investigated. Allying this principle to the second means that learners need to be directed towards the specific language in order for inductive learning to take place and the third principle implies a significant amount of control on behalf of the practitioner. The studies which used corpora as a reference tool achieved a level of mediation by making students look at a particular example of their language, in all cases their individual errors. This mediation does not contradict or undermine the idea of autonomy. In section 2.2.2 it was pointed out that learner independence does not equate to teacher absence. Charles (2007) visualises the teacher as a facilitator here. The mediation is necessary in making corpus based tasks practical and she describes the control the facilitator has over what language is examined as providing a necessary "jumping off point" (p. 297) which might then enable the learner to proceed on a more independent route.

Sripicham (2004) found that a problem his students had was that of overgeneralisation. Students were examining the corpora and in an effort to establish concrete rules or patterns from the concordances, they were often oversimplifying in their theories. He offered two solutions. The first involved vetting the concordances before the learners reached them. It should be pointed out that this vetting did not involve a manipulation of the language itself. Rather, it involved the deliberate selection of particular concordances for students to examine. Whilst this solution still engages learners in the act of inductive learning and still uses

naturally occurring examples of the language, it does seem to be a little removed from the long term goals of autonomy in EAP. If all the idiosyncrasies of the language are removed, then students get a curiously unauthentic display of authentic language. In the long term, this heavy mediation is likely to be harmful as learners will be unable to cope with divergent examples of the language. The second solution offered is simply to balance independent discovery learning on the part of the students with collaboration and advice where needed from the teacher. This seems to strike a balance between inductive, autonomous investigation and the realities of the classroom and the learning expectations of the students.

The importance of a compromise to what might be thought of as full autonomy, or a hands-off approach, and teacher guidance, is corroborated by a study undertaken by Perez-Parades et al (2011). This was an investigation into guided and unguided corpus consultation. Two groups were provided with the same tasks to be carried out using corpus searches and concordances. Though both groups received training in the use of the BYU-BNC, when undertaking the tasks, the control group was given no advice on how to search for particular items. The experimental group, however, received sign-posting from the teacher directing them towards the type of search they should be performing in order to access the language. The results of the experiment do not go as far as to suggest the guided corpus group were able to produce statistically significant more accurate responses to the tasks they performed. However, the group whose interaction with the corpus was mediated to an extent by the teacher accessed the corpus much more readily than the control group and spent longer performing more complex manipulations of the language and a greater number of searches overall. These results suggest that the mediated group had a much more positive experience of using the corpus than the other group. The conclusion drawn was that "skills and guidance are necessary when teachers take a corpus to the classroom" (p. 249).

2.4.3 Summary

It seems clear from the studies mentioned above that in order for corpus use to be successful and for autonomy to be encouraged, some form of mediation has to occur. This mediation should emphasise the needs of the learner and direct them to specific language items. From there, the learner can begin to engage cognitively with the language. Successful mediation initially may lead to a greater degree of autonomy later, thus meeting one of the requirements of EAP courses in equipping the learner with tools and skills for later use. I think it is also of note that these corpus based studies, in focusing on specific elements of the language, are addressing the elements of linguistic emphasis that some EAP courses have been charged with lacking.

In the following section I will look at error feedback as a particularly suitable form of mediation in terms of corpus use. This is partly based upon the evidence provided by the studies already examined, but I will also show that error feedback, despite the on-going debates about its efficacy, is something that can hardly be avoided in the prosecution of an EAP writing course.

2.5.Written Feedback

Truscott (1996) rejects the idea that error correction has any effect upon learner development. Because language acquisition is a long, developmental process passing through interlingual stages that are difficult to measure, the argument was made that there is no real evidence to support its efficacy. On the other hand, Polio, Fleck and Leder (1998, p. 47) suggest that interlingual problems can actually be addressed by error correction of written work as this is an awareness raising activity and asks learners to examine their own language. At the same time, however, the study they conducted was unable to show any significant difference in the improvement of writing between a group which received no indication of where their errors

lay and an experimental group which received extra help with the errors they were producing. There are several other sources looked at below which contradict this information, but beyond the actual effectiveness of error correction is the issue of perceived need and effectiveness in the eyes of the learners.

Even if longitudinal effectiveness of error correction has not yet been reliably measured, practitioners still need to respond to student writing in some way (Yates and Kennel, 2002). If a written task has been assigned and the learner has performed the task and submitted it, at the very least a response in terms of a grade is usually given. Chandler (2003) goes further and says that learners expect specific feedback on their errors. In the first place, not providing feedback might imply negligence on the part of the instructor or institution. This would have the obvious negative effects of demotivation. Secondly, when teaching a group which contains several learners with different abilities and needs, the teacher is often only able to apply a broad brush stroke in terms of the aims of lessons and the materials used; some classes are going to be more relevant to some learners than others, and it is difficult to tailor a course to meet the needs of every learner exactly. As already mentioned in section 2.2.2, this can result in courses where there is a 'disciplinary vacuum', with some EGAP needs being addressed at the expense of ESAP, particularly on a lexical level. In this environment, written activities become one of the only vehicles for an instructor to be able to examine the specific strengths and weaknesses of students. In turn, the feedback on written errors might constitute one of the only occasions a learner is directed towards language focuses relevant to them specifically. A further, practical consideration which stems directly from this is the fact that practitioners are generally expected or required to provide some form of error feedback for learners.

What form this error feedback takes further divides the EAP community. Although Gass (1983) and Makino (1993) indicated that students had the potential to deal with errors

when they were highlighted for them, exactly what form this highlighting should take is again a matter of debate. The following section will review common forms of error correction and suggest which might be more appropriate in regard to the discussion already made on autonomy and the use of corpora.

Direct feedback is that which is highly detailed. If the student has made a grammatical error within a sentence, for example, the instructor's feedback may take the form of a fully written out correct version of the structure. If the learner's error is lexical, the instructor might simply replace the wrong word with the correct one. One of the advantages of this method is that the students receive explicit and clear guidance on how their error differed from the target language. They could, therefore, repeat or modify the correction in an attempt to aid language acquisition. Ferris (2002) suggests that the direct feedback model is most suitable and effective for learners of a lower level. Sheen (2007) also noted that this method encouraged the acquisition of particular structural aspects of the language, although again this was with reference to only lower level students.

Direct feedback, however, becomes problematical within the scope of the aims of EAP courses. The method encourages little, or indeed no, cognitive participation of the learner by engaging them with their own analysis of the error. In terms of this thesis, direct feedback would be an inappropriate mode of feedback as, by providing the learner with the language, the need for investigation with corpus materials is rendered entirely redundant.

Indirect feedback, as the name suggests, is the polar opposite. This consists of the instructor alerting the learner to the fact that there are errors in the text, but does not attempt in any way to reformulate any of the language. In fact, it gives learners the barest of clues as to what the error actually is and might include circling a word to show it is incorrect but offering no other information, or simply providing a mark in the margin of the text which indicates that there is an error somewhere within that line. The rationale for this technique is

that it requires thought and work on the part of the learner in analysing their own writing. The approach has obvious benefits in relation to cognitive learning and autonomy; the learners do all the work. Ferris and Roberts (2001) suggest that, in the long term, this approach is much more likely to lead to accurate language use, but reliable longitudinal evidence to support this is lacking.

One of the problems, however, is the cognitive load that this processing places on the student. Initially, learners may not actually be able to identify the error itself or understand why it might have been an error in the first place if they do not have enough knowledge of that particular structure or lexical item. A further problem is one of time; this approach implies a greater investment of time by the learner (although probably not the instructor) and this might not always be feasible on all EAP courses or in in-sessional contexts. The method could, then, have entirely the opposite and unintended effect of causing the student to avoid engaging in error correction at all. In terms of corpus use, the methodology also has a weak spot in that, if the teacher is aiming to encourage the learners to use corpora to aid in their error corrections, it might be that further mediation or greater guidance is needed so that students have some idea of where to start their searches.

An approach which offers a compromise, although one which is generally classified as indirect rather than direct feedback, is one which uses meta-linguistic correction codes. With these, learner writing is marked up with symbols representing a series of meta-linguistic terms which the student can then examine. If they have made an error in tense, for example, the instructor might highlight the error by marking it with a 'T'. This means that the student is made aware of the nature of their error, but is still obliged to perform further analysis of the language in order to perform the correction. Ferris (2006) showed that 77% of errors marked up with a correction code were revised correctly by students. Using the approach also provided immediately interpretable data about which type of errors had been targeted more

effectively than others. It is perhaps also of benefit to the student that they might have a relatively simple diagnosis of which errors are particularly frequent and which might need further though or research; a page with many 'Sub/Vb' annotations might highlight a problem with a student not understanding subject-verb agreement or make them aware that they are not concentrating on their accuracy enough whilst writing, meaning a relatively straightforward error is being repeated.

The conclusions can be drawn then that, given the lower levels targeted by direct feedback and its inappropriateness in the light of autonomous learning, either indirect feedback or indirect feedback with meta-linguistic coding seem to be most relevant to EAP if students are expected to engage in inductive learning. When considering the use of a corpus, it seems there are two arguments for the use of correction codes. Firstly, whilst indirect feedback might be more convenient for instructors in the amount of time spent marking, an inverse effect might be experienced by the students, possibly resulting in unwarranted lengths of time being spent on revision of what might only be local level, or minor, mistakes rather than general errors. Worse, this could result in an abandonment of the correction procedure altogether. With correction codes, a greater deal of time and perhaps thought is required by the instructor to mark up the errors properly, but in doing so the student is guided to the relevant area of language that needs to be examined. This has the secondary effect, then, of offering a mediated way into a corpus, with students already having a particular language item and the nature of the problem in mind for investigation. Correction codes might also have the advantage in terms of data analysis in allowing an instructor to see which types of error were corrected more or less effectively through corpus work and this could in turn feedback in the future into materials and methods to target these consistent errors.

If targeting particular error types through a correction code, the number and type of errors targeted needs to be considered. Bitchener (2008) proposed that this type of feedback

should be restricted to a few or even just a single type of error. His study, for example, involved highlighting the misuse of articles. This has the advantage of singling out particular areas for learners, and indeed the instructor, to focus on. It has the added psychological benefit of avoiding the potential confusion and demoralisation of a student having a great number or large variety of errors to consider.

However, there are several drawbacks to this approach. The first is in how to establish which error or set of errors should be targeted. Again, this is a result of having classes of mixed ability or mixed needs students, which is likely to occur in a majority of classes, particularly within an EAP pre-sessional course. Whether it would be immediately obvious or certain that all learners were afflicted by the same error type might be difficult to identify. Whilst it might be feasible to suggest that this could occur at lower levels of language learning, it is likely that learners of EAP will have a greater diversity in their linguistic needs.

A second problem is that identified by Bruton (2009). He suggests that by focusing on a single error, the feedback process is no longer focusing on the learner's writing as a whole. Rather, the writing has become the vehicle for focus on grammar within a written context (p. 139). The danger here is that whilst one error might be tackled effectively, whether this would contribute towards overall written accuracy or proficiency might be in doubt. Expanding on this, Bruton asserts that in order for this meta-linguistic feedback to be of any use, students must be given some freedom of expression. That is, they must be free to make the type of errors they as individuals are susceptible to make in order for their needs to met effectively in feedback. This makes idea of targeting just one error or a very limited set of errors appear less useful and Bruton implies that feedback should not focus on "one language item at a time"(p. 140).

Another feature of error correction which Bruton notes which seems to be particularly relevant in the light of the corpus reference tool experiments described above is that errors should be tracked accurately. He suggests (p. 140) that research design involving error feedback and its effectiveness must take into account the following: Identification that particular L2 features produce language errors; that students demonstrate their ability to deal with these errors and produce a correct version; that targeted features are tracked in subsequent writing. These features fit neatly into the methods of draft/feedback/correction modes used in the corpus as reference studies. A meta-linguistic code which facilitates the identification, analysis and correction of a fairly wide variety of errors should be sought in order to meet these demands and examples are looked at in the methodology chapter of this thesis.

2.6 Conclusion

EAP writing courses differ in the contexts in which they are taught, meaning the content can vary widely. This is largely due, however, to a unifying feature which is to place learner needs, often defined by a particular goal such as an exam score or publication date, at the centre of course planning. Another feature which many EAP courses share as a result of specific student needs is a finite amount of time in which to achieve goals. This means that courses which have a large number of learners, representing a wide variety of needs, may struggle to address the specific language issues of individual students. This is particularly evident on EAP pre-sessional courses and may result in a lack of emphasis on certain aspects of written EAP. In the literature review, I have highlighted two areas which seem to be under-addressed on pre-sessional courses; ESAP needs, possibly in the form of a specific range of lexis, and wider linguistic proficiency needs which have been judged by some researchers and instructors as either less important than structural elements of written work or as simply impervious to the effects of course input. However, at the same time arguments are

made to the effect that a lack of linguistic proficiency can both undermine the organisational aspects of learner written work and lead to an inability to articulate an effective academic argument in subsequent writing.

Not unaware of these virtually intrinsic short-comings of EAP writing courses, institutions and practitioners seek to foster an attitude of autonomy amongst learners. Whilst definitions for this are subject to a variety of conditions, it is not unreasonable to say that within EAP it is regarded as a move towards learner independence and the idea of learning as research. That is, practitioners seek to provide learners with tools which will encourage them to learn inductively and take responsibility for their knowledge and acquisition of the target language they need.

Theories of DDL and the increasingly widespread availability of corpora have encouraged interest in the use of corpora with students as a way of providing a route into inductive learning techniques. Because of the inherent linguistic focus of corpus tools, they could be a way of addressing the lack of focused linguistic input on EAP courses and may offer a way for the EGAP / ESAP divide to be bridged. Certainly, the research mentioned involving corpora as reference tools suggests that there may be some positive short term benefits to the use of corpora in improving proficiency.

However, there are some gaps in the research into the effectiveness of corpora. It has been noted that several studies suggest an increase in accuracy when students have access to corpus data for error correction tasks. Whilst these types of task seem sound in meeting the requirements of mediating between the student and potentially overwhelming amounts of information, there does not seem to be evidence to suggest that using corpora as reference tools for correction is any more effective than using other forms of reference tool, such as a dictionary, within a realistic teaching context. Perhaps more importantly, the studies which have involved learners making corrections using corpus data do not always seem to have

made an attempt to identify if a correction was really made as result of corpus consultation. That is, the assumption in the research mentioned above seems to have been that if a learner is presented with ten errors and access to a corpus, then he or she will have used the corpus to correct that error. This is a problem, because in section 2.4.1 it was shown that the result of using corpus data for some learners was simply to raise awareness. Similarly, providing error feedback can also be an awareness raising activity in itself which does not necessarily call for the subsequent use of a reference tool. This may mean that those ten errors were corrected using the corpus data, or it may mean that any number of those errors were resolved simply because the learner's attention was drawn to an error which they actually recognised automatically how to correct.

The aim of research question RQ1 is to improve knowledge about when the corpus is chosen over other correction methods when being used as a reference tool as part of a draft/feedback/revision cycle; which particular errors, if any, are targeted using corpus data more consistently than others? A natural extension to this question is whether or not the corpus offered a more effective method of correction than other reference tools.

A further problem that exists in the research is identifying whether or not the use of a corpus has any measurable effect on the subsequent production of elements of the target language. This is particularly pertinent in light of the fact that many EAP courses are short in nature. Research question 2 (RQ2) asks, then, if learners who used the corpus showed any significant improvement in writing proficiency and to what extent this could be judged to be as a result of using the corpus. The implication of any positive or negative result here relates to the implementation of corpus use on short courses; if there is little or no measurable improvement in written proficiency as a result of corpus use, questions need to be asked about the appropriateness of incorporating it on courses.

Finally, qualitative reports of learner reaction to corpora are contradictory. This may be due to the types of learners and contexts of learning which studies have previously involved. There are reports of learners finding the access to large amounts of authentic data a boon. Some studies have shown that learners respond positively to the inductive aspect of corpus use and the ability to become independent researchers of the target language. Equally, these factors have produced negative attitudes amongst learners who feel the time and energy taken to work through corpus interfaces and concordances is not warranted based on the results and who would rather have the target language presented to them in familiar formats from textbooks and the teacher. Research question 3 (RQ3) adds to this body of research in identifying the attitudes of a very specific group of learners in a very specific learning context and seeks to identify any relationships between the qualitative responses of the learners and the quantitative nature of RQ1 and RQ2.

The following chapter explains the nature of the learning environment in which the experiment took place. I outline the nature of the experiment itself and go on to describe the research tools which were used, justifying my choice of the error feedback method used and the measurements of accuracy. I also examine in detail the reasons for the types of statistical tests chosen for use as part of the data analysis.

3

Methodology

3.1 Introduction

This chapter outlines and justifies the methods employed to collect and analyse data used to answer the research questions described in Chapter 1, section 1.6**.** First, a description of the environment in which the data were collected is given, including a description of the sample, an overview of the types of data being gathered and a data gathering timetable.

Following this, the tools used to produce and collect the data are examined in detail. A justification for the use of the BYU-BNC is given and an outline of the training which the learners had in the use of the interface is given. The actual training tool used can be found in appendix 4. The choice of meta-linguistic error correction coding is examined as are some of the problems generated by particular error types and the solutions employed. The method employed to measure accuracy in the form of EFT (Error Free T-units), which are defined in section 3.4.2, is then explained. For these last two points, I employed a method of inter-rater reliability testing. The rationale behind this is explained, as is the method itself, and I also supply the results as these dictated some of the subsequent decisions made in the construction of the experiment.

The collation of the data is then briefly outlined before the confounding factors affecting the analysis of the results are discussed. Finally, with the confounding factors in mind, as they influenced to some extent the tests employed, the type of statistical analysis employed is explained in detail.

3.2 Data Gathering

Subsection 3.2.1 gives details of the participants involved in the experiment and section 3.2.2 outlines how the experiment was embedded within a pre-sessional course context. Sections 3.2.3, 3.2.4 and 3.2.5 describe the type of data which was collected and the methods of collection.

3.2.1 Participants

The University of Sheffield's six week intensive pre-sessional course provided the opportunity to examine the effects of the use of corpus data on learners within a realistic pedagogic context. The framework of the course allowed me to gather data from two groups of similar configurations studying the same materials and following the same syllabus for equal lengths of time. These groups were allocated to me by the University of Sheffield as students to teach, rather than purely as study samples, meaning the sample was one of convenience. However, I reduced the number of non-linguistic variables as much as possible in the design of the experiment. In addition to this, the basic composition of the groups had the effect of limiting some potential variables; all the learners were from mainland China and all were studying in order to enter onto postgraduate study in either the University of Sheffield management or economics departments. All the students had completed undergraduate degrees in mainland China and all shared the same mother tongue of Mandarin. These groups were typical of the intake on this particular course. The confounding factors themselves are examined in detail in section 3.7.

The control group (CG) consisted of fourteen students, ten female and four male (gender being a potential variable in initial statistical analyses). All had received conditional offers which would be converted to an unconditional status based on their performance on the pre-sessional course. The experimental group (EG) initially consisted of thirteen students but

was reduced to twelve due to illness. Eight of the remaining group were female and four were male. Data initially supplied by the thirteenth member who did not complete the course was discounted. The group was subject to the same offer conditions from the university, although one member of the EG group had already obtained an unconditional offer.

The basis of placement within groups was on similarities of previous IELTS scores and the nature of the learners' intended areas of study. In the case of both the CG and EG, the learners all intended to progress onto courses which were management, finance and economics related. Placement was made by the University of Sheffield English Language Teaching Centre and was made independent of myself. These conditions of placement aimed to create groups consisting of students of a similar level of English and of similar subject interests.

3.2.2 Description of the University of Sheffield Pre-sessional Course

The six week pre-sessional course consisted of nineteen and a half hours contact time with instructors. Figure 3.1 provides an overview of a typical week's input. Note that the actual order of when students received input on a particular skill varied according to instructors' timetables.

Over the period of each week, contact time was divided into ten 90 minute reading, writing, speaking, listening and project classes in the morning and three 90 minute exam classes in the afternoons. All students also attended a 60 minute lecture on an academic topic once a week. Students also received three short tutorials over the period of the course; the first in the initial week of arrival, one in the third or fourth week with the final tutorial occurring in the last week upon receiving results.

	09.15-10.45	**11.15 – 12.45**	**13.45 – 15.15**
Monday	Writing	Reading	Exam Preparation
Tuesday	Listening	Speaking	Exam Preparation
Wednesday	Extended Assignment	Extended Assignment	Academic Lecture
Thursday	Writing	Reading	Exam Preparation
Friday	Listening	Speaking	N/A

Figure 3.1. *Six week course weekly timetable.*

Summative assessment of the students was by means of the University of Sheffield English Proficiency Test (USEPT), which was modelled along the lines of the common International English Language Testing System (IELTS) and provided a score which was a direct equivalent. Other assessment was through an extended written assignment, the project, which will be discussed in greater detail below as it constituted one source of data for the experiment.

I worked with two other teachers. These teachers were responsible for listening and speaking input whilst I was responsible for reading and writing of the CG and EG, thus maintaining control of the written input. All students were issued with, and used, the course ESAP textbook (Corballis and Jennings, 2009), which was essentially an EGAP textbook with reading texts drawn from a management or finance oriented source. A second teacher was responsible for any extended assignment input for the CG but it was agreed that this would take the same form as the EG as much as possible. Input for this session covered summarising, paraphrasing and referencing skills, with the rest of the time being independent research time for students with the instructor available to answer any queries.

Initial needs analysis was given in the form of a questionnaire. This included administrative data recording names, country of origin, contact details, intended area of study,

history of learning English and perceived strengths and weaknesses. A copy is included in appendix 2. Students were also given a learner type questionnaire (Mumford and Honey, 1992). Again, this will be discussed in more detail subsequently as it provided data towards the research. The final part of needs analysis was the pre-course writing task. This was administered as a matter of course in order to get an overview of the basic writing skills of the learners but the examples were subsequently used by me both for the inter-rater reliability testing and as the source of the data for the baseline cycle of the experiment. The specific nature of the task is related below alongside a discussion of the other writing tasks which were performed as part of the experiment.

3.2.3 Data Gathering Timetable

Figure 3.2 outlines the data gathering timetable which was employed during the course of the experiment.

	Learner Activity	**Researcher Activity**
Week 1	• Needs analysis questionnaire completed. • Learner type questionnaire completed. • Pre-course (baseline) writing. • Corpus training (EG only). • Essay 1 task received.	• Inter-rater reliability test conducted.
Week 2	• Essay 1 draft submitted. • Essay 1 draft revised and re-submitted. • Essay 2 task received.	• Essay 1 task draft received / correction code applied / returned to learner. • Essay 1 revision received. Data logged.
Week 3	• Essay 2 draft submitted. • Essay 2 draft revised and re-submitted. • Submission of extended writing draft. • Essay 3 task received.	• Essay 2 task draft received / correction code applied / returned to learner. • Essay 2 revision received. Data logged. • Structure / cohesion / content of extended writing commented on.

Week 4	• Essay 3 draft submitted. • Essay 3 draft revised and re-submitted. • Extended writing draft comments received. • Essay 4 task received.	• Essay 3 task draft received / correction code applied / returned to learner. • Essay 3 revision received. Data logged. • Extended writing draft returned.
Week 5	• Essay 4 draft submitted. • Essay 4 draft revised and re-submitted. • Exams completed (entirely independent of experiment). • Final submission of extended writing.	• Essay 4 task draft received / correction code applied / returned to learner. • Essay 4 revision received. Data logged.
Week 6	• Course questionnaire completed (CG and EG). • Interviews completed (CG and EG).	• Extended writing submissions coded and logged. • Questionnaire and interview data logged.

Figure 3.2 *Data gathering timetable.*

The four weeks saw the students complete four written pieces of homework. These were set by the researcher but conformed to the expectations of the course; they were short essays (intended to be between three and four hundred words) with the question performing two tasks: (i) to provide students with an opportunity to exercise particular rhetorical and genre based functions, largely dictated by the types of functions they would be expected to be aware of for the exit exam and (ii) to provide students with a writing task broadly related to their field of study. The titles given, including the initial pre-course task which served as the source of baseline data, are given below:

- Pre-course task: Is it better to save money for old age or make use of it when young?

- Essay 1 task: What balance do companies need to strike between profit and expansion on the one hand, and responsibility towards their workers and local communities on the other?
- Essay 2 task: What are some of the main causes and effects of unemployment in Europe?
- Essay task 3: The gap between the rich and poor is increasing. Are there any solutions to this growing problem?
- Essay task 4: Whilst unemployment rises and economic growth slows, directors of companies (which are often performing badly) continue to award themselves huge financial bonuses. Discuss.

These essays were assigned on Fridays, written over the weekend, submitted on Monday, marked up with meta-linguistic error coding and returned to learners on the Wednesday, with the corrected drafts being returned on the Friday, whereupon the cycle was repeated. The exception to this was the pre-course task which was assigned and completed by all students within a 40 minute period in class during the first week.

On receiving the initial draft, I logged all the error types and details of the errors. Drafts were marked up in MSWord using the meta-linguistic code described later in section 3.5 of this chapter. This was then returned to the student for revision. Once the corrected version had been submitted, I then logged the method of correction, the details of the correction and whether it had been successfully made or not. Texts were returned to learners with tick marks for corrected errors and feedback to address matters of organisation, critical thought and content, but in order to preserve the level of control over the language, specific feedback was not given on the errors themselves.

In correcting the drafts, the learners were asked to perform two tasks. The first was to correct the error on the hard, paper copy based on the code. The second was to mark the

method used to correct the error using a simple code. The CG had the options of knowledge (K), monolingual dictionary (D), bilingual electronic dictionary (ED) and other / online source (O). The EG received the same, but with the additional option of a corpus based correction (C). These codes were employed to aid in answering RQ1.2, providing information on which method of correction had been applied to particular types of error and RQ1.3, providing information on whether or not this method had been successful.

One issue complicating the comparison between dictionary and corpus effectiveness in this experiment is the dictionaries which were employed. Those available to the learners were learners' personal electronic dictionaries and monolingual English dictionaries by Collins. In the case of the latter, corpus consultation directly informed the dictionary entries. It is unclear as to whether the electronic dictionaries being used by the students were informed by corpus data although this is entirely possible. The implication is that even students who were using dictionaries rather than the BYU-BNC itself were accessing data which had been mediated by a corpus. The distinction is made in this experiment, then, between using dictionaries which may have been created through the use of corpus data, and the act of accessing a corpus and concordances in a raw from through the BYU-BNC interface. The data utilised by PEDs (Personal Electronic Dictionaries), as compared with paper dictionaries and corpora in their raw form is an area which would benefit from further research, particularly considering the technological implications discussed in chapter 5.

3.2.4 Extended Writing

Students began work on a piece of independent extended writing in week 1. The guidelines were dictated by the course itself rather than me: learners were to produce a 1,500 word text on an area related to their target discipline. Students therefore chose their own focus. In the

case of both groups, as the learners were all aiming to progress to a small number of related disciplines, the subject matter varied but all had a common management or finance related theme. Students received little instruction on these pieces beyond an initial tutorial detailing the nature of the assignment and how to go about choosing the topic, and some input on summarising and paraphrasing techniques and referencing conventions. For both groups, the input was identical and taken from Bailey (2011). Learners were then left to complete the first draft by the end of week three. When submitted, I commented on the drafts only in terms of organisation and academic conventions such as referencing, and overall content. That is, they were not subjected to feedback on individual language components in the same way that the essays were. This was not only for the purposes of the experiment, but was common practice across the course both as a result of time management and to emphasise the fact that the assignment was specifically a tool for promoting researching, synthesizing and organising skills at university level.

Final submission of the projects occurred at the end of week five. The submissions were then individually logged for language errors in the same way as the four in-course essays, but the text did not go through the revision cycle. This was done in order to examine whether the observations of the mediated error correction cycle had any correlation in terms of accuracy and error production with a piece of work produced independently of instruction or language input by the instructor and served as the mechanism for answering RQ2.2.

As mentioned above, the final method of assessment of written language was the university's USEPT test. The results for this were not collected for the experiment. In a practical sense, it was not possible due to the confidential nature of the papers. Test conditions were also entirely unlike the conditions which the four essays of the experiment were conducted in. Finally, the results themselves as a score had little bearing on whether the language was accurate or not; the score is the result of marking according to a number of

generalised criteria, rather than a detailed analysis of language accuracy. This collapsing of the language into general bands is not considered an accurate measure of achievement because it does not employ actual measurements of proficiency, as detailed in section 3.6 of this chapter, and instead produces a general impression of the language produced overall through the use of prescribed language level descriptors, some of which insufficiently reflect the target language and constructs of pre-sessional EAP courses (Banerjee and Wall, 2006, p. 54).

3.2.5 Qualitative Data Collection

In the initial week, learners were asked to complete the Honey and Mumford learner type questionnaire (1992). Copies of these can be found in appendix 3 but I will provide a summary and rationale for its use in the experiment here.

The questionnaire consists of a series of 80 statements which relate to potential personal attitudes of the learner. These attitudes might be towards a general moral or ethical issue, such as statement number 1, "I have a strong beliefs about what is right and wrong, good and bad" to more specific and concrete statements related to work, time and personal management, such as statement 77, "I like meetings to be run on methodical lines, sticking to laid down agenda, etc". To complete the questionnaire, the participant simply marks any statement they agree with using a tick and any they disagree with using a cross. Once this has been completed, the participant refers to a scoring system which divides their responses into four categories which correspond to a learner type – activist, reflector, theorist and pragmatist. These scores can then used to complete either a simple table or a quadrant diagram which show which of the four learning types the participant leans towards. These then correspond to a set of detailed descriptions of the characteristics of each learner type. The outcome of these questionnaires is reported in chapter 5, section 5.2.2 , but it is pertinent

to note now that because some results expressed equal weightings between two learner types, compounds of learner types were created as a result and factored into the analysis. This follows the practice of Poole (2006).

In terms of the learners themselves, and beyond the actual experiment, this survey was offered as an awareness raising activity to help them think about their learning habits and how these may or may not be adaptable to the learning environment of British universities. In terms of the experiment, the scores were used as a variable to ascertain whether or not a particular learning style might have an effect, for example, on proficiency, or frequency of use of a particular correction type, or a correlation with attitudes expressed in the interviews and exit surveys. As such, the survey did not form the basis of the entire experiment as with Poole (2006), but was rather employed as a potential way to triangulate other findings.

In order to examine the attitudes of students in both groups towards the exercises they had been involved in and the tools they had used, two data collection tools were employed. The first was a questionnaire administered in the final week of the course, once the last piece of work had been submitted. This was based on the questionnaire used by Yoon and Hirvela (2004) with local modifications which took into account the course and the two different groups. A different questionnaire was given to the EG and CG, the former consisting of 44 questions and the latter of 42 questions, all marked according to a 5-point Likert scale. These were administered through a Google form and can be viewed in appendices 11.1 and 11.2. The survey targeted the attitudes of the student towards feedback, the methods of error correction and thoughts on proficiency and, in the case of the corpus group, if they would continue to use corpora in the future. The adaptation of an existing questionnaire was used with RQ3 in mind, part of which examines whether or not the attitudes expressed by learners expressed in this experiment corroborate those found in earlier pieces of research.

Learners from both groups were then invited to take part in two separate interviews, one for the CG and one for the EG. In the event, only five members of the EG and two members of the CG were willing to submit to an interview. The EG interviews aimed to look in more detail at the attitudes of the students and also attempted to examine how the students had engaged with the corpus; the types of searches they had conducted and if they had used it at all during the composition of either their essays or the extended piece of writing.

A year after the completion of the course, members of the EG were issued with a further, shorter survey based on the first to find out if they had continued with the use of the corpus and if their attitudes had changed at all.

3.3 Corpus Selection and Training

Sub-section 3.4.1. examines why I chose the BYU-BNC as the corpus interface for the learners to interact with, whilst section 3.4.2 details the training program which was used.

3.3.1 Corpus Choice

As already mentioned in the literature review, there are a wide variety of corpus interfaces available. One of the main criteria that the chosen corpus had to meet that was that it be freely available online. As one of the observations was whether learners made use of the corpus independently, as opposed to only addressing the errors highlighted in the experiment writing cycle, access had to be available both inside and outside the classroom. This prohibited the use of some corpora currently available, including the large (550 million word) Collins Wordbank and the International Corpus of English which can either only be accessed by paid subscription or downloaded by license. Lack of availability, small sizes, and copyright issues (Krishnamurthy and Kosem, 2007, p. 359) meant no availability of a management ESAP corpora specific to the students in question, nor did time permit the

construction of an appropriate ESAP corpus. This in itself was not deemed a problem based on observations by Li and Pemberton (1994, p.184) which suggest that because students have an interest in or previous knowledge about their subject area, it is not necessarily specific language so much as the language which is common to all areas of EAP which learners have most difficulty with. Lee and Swales (2006, p. 71) also suggest that large corpora might be more suited to EGAP work, as found on a pre-sessional course, as the types of query made are likely to be answered more successfully when larger quantities of data are consulted.

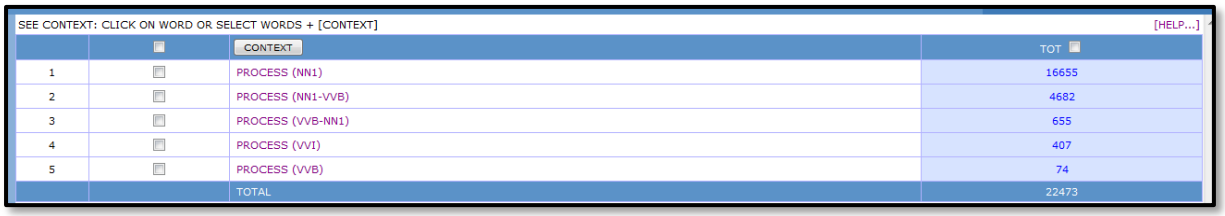
Three interfaces were found freely available which were judged potentially useful. IntelliText (Stephenson, 2011) is an interface operated by the University of Leeds which permits queries to access a number of corpora simultaneously. This might offer significant opportunities for the linguistics researcher but had the potential to overwhelm a learner whose discipline or interest lay elsewhere; based on the information encountered in reviewing the literature, I deemed that this might encourage negative attitudes before the experiment had even begun. There were also certain problems in ease of use. Certain functions, tested on several computers, didn't seem to work or froze the page on consecutive occasions (the 'genres' function, for example). A further problem was that of the interface, which is quite complex and requires a lot of user movement due to the number of panes which need to be navigated between. There is difficulty, for example, in observing concordances and frequency data simultaneously.

Another freely available interface was the BNCweb (CQP-edition) Version 4.2 provided by Lancaster University (Hoffman and Evert, 2006). This is based on the BNC corpus, as is the interface which was eventually chosen, but has a number of more advanced features. These include the ability to upload sub-corpora and self-tag tokens. Again, whilst these features might be of benefit to linguists, it was felt that they would be superfluous for

the aims of this particular experiment and for the needs of these particular learners. Time training in its use may also have presented an issue.

The BYU-BNC (Davies, 2004) was eventually selected for use for a number of reasons. One was the interface itself. Krishnamurthy and Kosem (2007) have claimed that of all the elements of an EAP corpus meant to be used with learners, the interface is the most important part. They noted the dauntingly complicated nature of research corpora and themselves cite the BYU-BNC as an example of a user friendly interface (2007, p. 369). Not only does it display all information in three simple frames (query frame on the left, data frame at the top right and concordance frame bottom right), but it also displays information on any query in three basic ways which can be selected by the user. The first is as a list of tokens through which the relevant concordances can be accessed.

This is illustrated in figure 3.3. The second is the query represented in the form of graphs which show the distribution of the word across seven general genres. Again, concordances can be accessed by clicking on these genres and sub-genres are available for display as well. This display was considered to be particularly relevant as it gives an immediate clue as to whether a language item might be suited to an academic context or not and is shown in figure 3.4. A third display provide colour coded KWIC (Key Word In Context) data to provide co-textual grammatical information, as displayed in figure 3.5.


SEE CONTEXT: CLICK ON WORD OR SELECT WORDS + [CONTEXT] [HELP...]

		CONTEXT	TOT
1		PROCESS (NN1)	16655
2		PROCESS (NN1-VVB)	4682
3		PROCESS (VVB-NN1)	655
4		PROCESS (VVI)	407
5		PROCESS (VVB)	74
		TOTAL	22473

Figure 3.3: *BYU-BNC display of tokens for 'process' by frequency.*

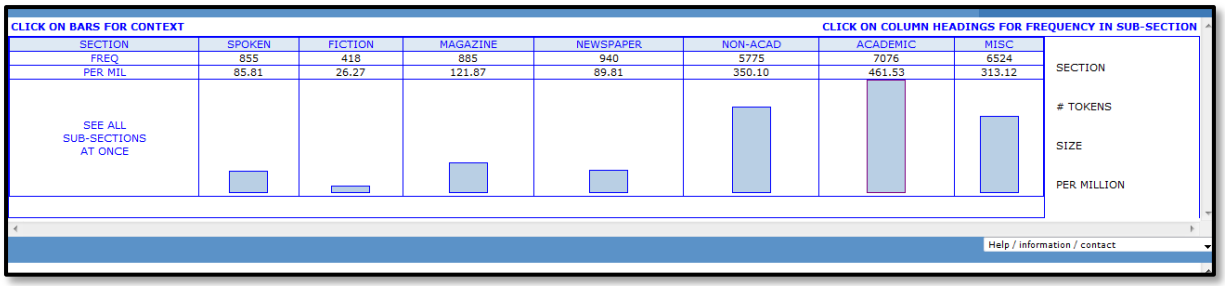


Figure 3.4: *BYU-BNC display of tokens for 'process' by genre.*

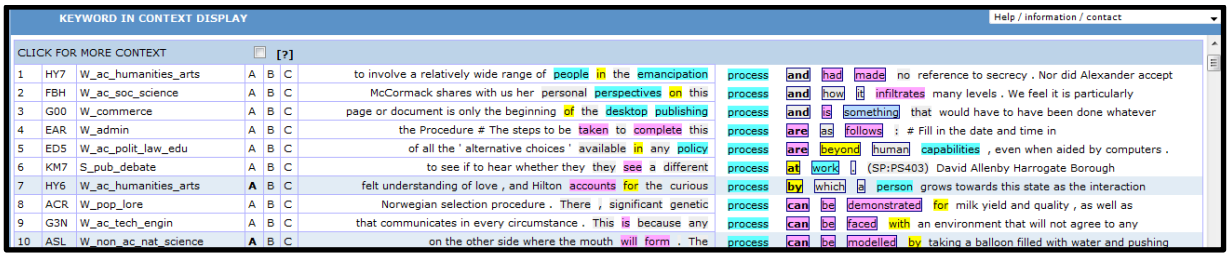


Figure 3.5: *BYU-BNC KWIC display for 'process'.*

A second reason is the relatively large size of the BYU-BNC and the range of sources from which the language has been extracted. Consisting of just under 100 million words, the BNC is monolingual and consists of British English which represents the language at the end of the twentieth century, being completed in 1994 with a modified version available from 2007. 90% of the corpus is compiled of written English, with 91% of this from the period between 1985 and 1993. This written component is divided into six different genres: fiction, magazine, newspaper, non-academic, academic and miscellaneous. These are further compiled of a number of sub-genre, including different academic disciplines in the academic strand. Biber (2006, p. 252) has noted the importance of a corpus being composed from sources derived from a variety of registers and topics.

In addition to being freely available for use after initial registration (also free), there are limitations on the number of searches conducted but I judged these sufficient for the pre-sessional EAP learner: 100 separate searches in any 24 hour period with 1000 retrievable

KWIC queries available. Once registered, the site maintains a user history recording recent searches and access dates.

A potentially problematic area is the age and, possibly, size of the BNC. Compilation had been achieved by 1994, twenty years before my thesis was completed. No new language has been added to the corpus since, and the period 1994-2014 has seen the rise of much new vocabulary related to various domains but particularly technology. The BNC can be compared to the COCA (Corpus of Contemporary American English), which represents American English from 1990 to 2012, a more contemporary offering of English. However, at the time the pilot trials went ahead, this was not available through the BYU interface. It was incorporated into the interface, along with a number of other features, as the experiment ran. This corpus is also relatively large, at 450 million words. The COCA interface also now links to Google Books, which provides a corpus of 155 billion words. However, The BNC's size and age have been deemed adequate by other relatively recent investigations (Hoffman et al, 2008; Chambers, 2007; Yoon 2008) for it to seen as an acceptable source for making investigations into the use of contemporary British English.

3.3.2 BYU-BNC Training System

The literature review highlighted several studies which have incorporated and noted the importance of a training system of some description to be given to learners before they embark upon corpus based searches of their own. This is both for the efficacy of the use of the corpus but also for motivational purposes. Because of the time limitations dictated by the short nature of pre-sessional courses, this training system, which can be found in full in appendix 4, was designed to be completed within a 90 minute session (or one lesson slot) in a computer lab with learners who had never used a corpus before. In part, it was based on an online training system developed for the English learners at the University of Jyvaskyla,

Finland (Korhonen et al, 2007). This consisted of a series of online tasks to be complete over a period of approximately three hours by those new to the BYU-BNC. Having half this time available meant modifying this course and restricting instruction to a particular number of functions. The training system was piloted and modified in four different sessions with four different groups of EAP students at the University of Sheffield in the period between July 2011 and May 2012, and consisted of a series of tasks illustrated with screen shots of search and analysis procedures.

The training system had to achieve a number of objectives. The first was to guide the learners through the use of the interface. The second was to provide students with an awareness of the meta-linguistic codes and terminology used on the BYU-BNC. The third, and perhaps most important in light of the literature review findings, was to make it absolutely clear to learners how the BYU-BNC might be used in a way which was going to be of potential use to them in their further studies. What follows is an outline of the session with a description as to why the certain steps and functions were selected.

After confirming that none of the learners had encountered a corpus before, the concept of a corpus and its uses were explained to the group and they signed up to the site. Students then conducted their first search by looking for information about the word *process*. They were shown how to identify frequency, what the concordance lines were and how to access the original context, and then they were asked to hypothesise about usage of the word *process* based on the first set of concordances. The aim here was to help negotiate a basic search, show that patterns of language use could be identified independently and that this identification could be done without having to scour through every single concordance line available.

The second stage involved a more detailed manipulation of the interface, with learners being taught how to combine a word search with information about the parts of

speech commonly associated with the item using the *Group By* function. Again, the students were asked to generate hypotheses about the word *process* and its frequency of use as a particular part of speech. At this stage, they were also introduced to the *Chart* function of the interface and how to investigate the genre and sub-genre information of *process.* They repeated this procedure with the words *function* and *contrast.* Based on the information that they had gleaned so far, students were asked briefly to hypothesize about whether the searches for information about an individual word might combine to suggest a general linguistic trend within academic English i.e., constructions using noun forms are particularly prevalent.

The third stage moved on to investigating how the BYU-BNC could be used to address lexico-grammatical issues. The issue examined was that of prepositions, and the learners became familiar with how to search for frequently occurring prepositions with particular word types by using the *POS list*. At this point, learners were asked to make themselves aware of the codes used in the *POS list*, many of which were self-explanatory. For those which weren't, the learners queried the instructor and were also shown where to find the full explanation of the *POS list*. Students then used the same section of the interface to experiment with other types of collocation and were also shown the wild card function in relation to fixed expressions. The KWIC display was also introduced at this stage as something which might be useful in helping illustrate the relationship of the target language with various supporting linguistic items within the concordances.

Finally, the class was introduced to the feature of the BYU-BNC which allows for the generation of synonyms and the *Compare* function which enables users to hypothesise upon the suitability of one word combination over another similar word combination. The tokens *large* and *big* were chosen for comparison, with the students again being asked to hypothesise

about which one, and why, might be more common in an academic context based on the lists and concordances which were generated.

Once the program had been worked through, the learners had some time to familiarise themselves a little more by making further informal, self-directed searches and to raise any queries and troubleshoot any issues. To consolidate what had been learned in the session, students were provided with a homework task which involved a series of corpus and concordance queries, the completion of which would mean utilising all the functions looked at in the session again.

3.4 Accuracy and Errors

This section details the data measurements which were used. Section 3.4.1 explains the choice of a meta-linguistic correction code as a tool for the mediation of corpus data. 3.4.2 discusses the method which was used to measure a particular aspect of written proficiency, namely accuracy. Section 3.4.3 describes the inter-rater reliability tests which were used to ascertain the reliability of accuracy measurement and error recognition. Based on those tests, section 3.4.4 and its subsections provides the definitions of particular error types used in the experiment.

3.4.1. Error Correction Method

Error correction was adopted as the method of mediation between the learners and the corpus itself. The literature review has already discussed the importance of guiding the students into the corpus. Research also stresses that effective use of corpus based tools is more likely when students see that it is directly related to their own language. In terms of the institution, error correction was expected of the instructors by both management and the learners themselves.

It was necessary for the experiment to balance the amount of control involved in the mediation and a need to be able to measure accuracy quantitatively with the principles of trying to encourage autonomy in the learners. A high degree of control could be achieved by restricting the data the accessed by the learners. As with Cobb (1997) and Cresswell (2007), both groups could be given the same tasks and told to work through them using different methods and / or tools. Measurement of success rate in the tasks might then give some idea as to the effectiveness of one method over the other. Similarly, the EG could be guided by the instructor dictating the specifics of searches or restricting the amount of language examined, as in Watson Todd (2001). Though this would provide a highly controlled environment, the literature review has already outlined some of the problems with these heavily controlled approaches.

The first is the potential for the focus to shift away from the holistic process of writing to the work on an atomistic level. Any measure of success would then be on a given element of the language rather than on the writing as a whole. Secondly, the higher the level of control, it could be argued, the less chance there is of successfully fostering a move towards autonomous learning; mediation to the extent that the instructor is supplying the students with all the possible language items and dictating the types of queries undertaken is far removed from prompting learners to undertake their own investigations. Thirdly, tight controls over the language examined have the potential to move the focus away from learner needs. As mentioned in section 2.4 of chapter 2, by pre-selecting language to examine, the instructor is by default treating the class to some extent as a generic whole. This might be justifiable when trying to meet the goals of a syllabus, being able to produce a text in a certain genre, for example, or perhaps when working with a group of learners whose needs are likely to be more uniform, such as at a lower level. It is harder to justify such a tight control over the language when taking a varied class of advanced learners into consideration.

Needs are likely to range quite broadly. A class wide, rather than individually focused, approach towards investigating language is likely to benefit only a proportion of the students. In terms of corpus use, it might also mean that students who have knowledge of a particular language feature would not need to resort to using the corpus for examination. In this case, the researcher would be unable to identify whether or not the corpus or some other factor was responsible for the production of accurate language, something already noted as problem in other studies and one which I am trying to address through research question RQ1.

The experiment required, then, an approach which allowed students to tackle their specific needs rather than teacher led dictation of target and language whilst at the same time providing some form of guidance as to how to do this. Metalinguistic error coding was chosen as the most suitable way to execute this. It provides learners with a definite area to focus on without dictating how exactly the error can be corrected. It is immediately obvious to learners that the exercise is in some way beneficial as it targets their individual writing. By covering a number of different error types, the system addresses the writing as a whole rather than just one or two pieces of grammar. In terms of the research aims RQ1 and RQ2, using linguistic coding also meant that it was possible to track different error types over the course, establish which errors were more or less effectively dealt with and also allowed for investigation as to which type of error correction method worked with particular types of error. Whilst the results of Chandler's study into error correction methods (2003) suggested that the most effective was actually direct correction, there was a discrepancy between this result and the perception of the students, who felt that metalinguistic coding of errors was more beneficial. A similar attitude had already been noted in Ferris and Roberts (2001), where underlining coupled with a description was the method preferred by students.
The University of Sheffield provided a metalinguistic correction chart to be used with the students, and this code is shown in figure 3.6. This chart is of a nature similar to that of

Ferris' (2001) sixteen point code and Jordan's (1999) twenty point code in that it attempts to cover a range of specific errors rather than grouping them into broader categories such as 'Grammar', 'Vocabulary' and the like.

All errors were marked up on student feedback, but not all were included in the subsequent collation and analysis. Spelling errors were ignored due to auto-correct facilities and the difficulty of distinguishing between an actual error in student spelling and a simple mistake or typo that had gone undetected or ignored by the student. In the event, possibly again due to spell check and auto-correct, manuscripts were largely free of these items. Elements of repetition were similarly not counted if it seemed obvious that a mistake had been made rather than an error, as in the example "because the the factory produced little…" Punctuation errors were not included in collation for the same reason and for the difficulty of being able to correct these using corpora or dictionaries.

Symbol	**Meaning**	**Example**
SP	SPELLING	I need to do more practise
P	PUNCTUATION	I need scissors paper clips paper and pins
T	TENSE	I go there yesterday
WW	WRONG WORD	He likes searching for information on the interweb
WO	WORD ORDER	Where you did go last night?
WF	WORD FORM	Wind energy is very safety.
G	GRAMMAR	If he had seen me, he will have stopped to help.
F/I	FORMAL/INFORMAL	There has been loads of research on this topic.
^	WORD MISSING	She shouted out ^ him but he did not hear.
Agr	SUBJECT-VERB AGREEMENT	Fred like ice-cream.
P/S	NUMBER	It takes two hour to travel to London.
C/U	COUNTABLE/UNCOUNTABLE	Bob likes chips; in fact he loves it.
REP	REPETITION	And he marvelled at the marvellous place and he also marvelled at…

/	EXTRA WORD	And thus, therefore, he realised…
?	SENSE/MEANING UNCLEAR	And I knew his sound was good sound
//	NEW PARAGRAPH NEEDED	Final advantage. Now, we will look at…
≠	PARALLEL CONSTRUCTION NEEDED	Managing your time and to study hard are important.
Frag	FRAGMENT	He left his job. Because he wanted to go travelling.
R-O	RUN-ON-SENTENCE	London is a famous city, many tourists go there.
A	ARTICLE	A man wearing a blue hat knows the answers.

Figure 3.6. *Meta-linguistic error correction code supplied to the learners on the 6 week pre-sessional.*

It is perhaps obvious that the labelling of some errors is, to a lesser or greater degree depending on the context, a matter of some subjectivity on the part of the marker and judgment of lexical items in particular is likely to suffer from human error (Witalisz, 2007). For this reason, inter-rater reliability tests were undertaken both for errors and measures of proficiency. These tests and the results are outlined in greater detail in section 3.4.3 of this chapter, following an outline of exactly what constitutes proficiency within this study.

3.4.2 Measuring Written Proficiency Through Accuracy

Wolfe-Quintero et al. (1998) define written proficiency as a combination of three components: fluency, complexity and accuracy. Fluency is the ability to produce language at a particular rate and this may be measured, for example, by production of a text within a particular time constraint. Complexity is measured through the judgment of how complex or varied a text's structure might be. Accuracy is error-free production. Wolfe-Quinetro et al. (1998) suggest that these are inter-dependent on each other and explain that "trade-offs" (p39) between the three elements are likely to occur; an increase in fluency might be made at

the expense of accuracy as the learner pays less attention to error production, whilst an increase in accuracy might be at the expense of complexity, perhaps where the learner decides to use less complex language but which they are confident in expressing without problems.

However, measuring the inter-dependent nature of these components is beyond the scope of this study and, as in research done by Polio (1997), Polio, Fleck and Leder (1998) and Storch and Tapper (2009), accuracy was used as the tool to judge the increase in written proficiency. Referring again to the Wolfe-Quintero et al. survey of proficiency studies (1998), accuracy is important as a measurement of proficiency because it "shows the conformity of second language knowledge to target language norms" (p. 4). That is, accuracy is a measure of how much control learners have over the production of the correct target language.

In order to measure accuracy, each script was analysed for number of (TT) T-units (terminal-unit) and number of error-free T-units (EFT), with the corresponding ratios reported as percentages. The T-unit as defined by Hunt (1965) consists of any main independent clause and its dependent clauses. Below is an example of a T-Unit:

"Because there are too many people looking for water, these wells becomes [sic] dry after one or two months."

In this example, because of the error in subject – verb agreement (these…becomes), the T-unit would be classed as incorrect.

Learner essays produced Total T-Units (TT) and Error-Free T-units (EFT). This system gives an overview of how accurate the essay appears to be, though it does not address

the errors themselves, or why exactly the T-unit is incorrect. The more EFT present, the more accurate the piece of writing. One issue brought up by Bardovi-Harlig and Boffman (1989) is the discrepancy between number of errors and number of T-units. An example from this experiment is provided below.

"If the government will increase [sic] benefits, the poverty [sic] people would suffer less."

In the revision process, this was corrected to the following:

"If the government increases [sic], the poor would suffer less."

In the original draft, there is one T-unit which is incorrect, but there are two errors. During the revision process, the learner has reduced the number of errors, but the T-unit remains incorrect and has the same EFT value of 0 as in the original. For the purposes of this study, this was not regarded as a problem as overall accuracy was being analysed with correct T-units. If a revision remained inaccurate due to errors being left in a T-unit, then that T-unit was still inaccurate and was to be measured as such. The individual errors themselves were all accounted for when looking at error types and success of different correction methods.

For the same reason, the decision was made to use this measure and not to make an additional count of error free clauses (EFC). These were considered superfluous and potentially confusing. EFC would not necessarily have supplied a more detailed description of accuracy. In the example above, the count of EFC in the revision would be 1, but the unit of meaning, or the T-unit, would still have a value of 0. It was felt therefore that the calculation of EFC may have complicated any descriptions of accuracy. That is, in theory a learner may have doubled the number of EFC in a revision, prompting the conclusion that

accuracy had increased, and not have increased the number of EFT at all, which would generate exactly the opposite conclusion. If this was the case, then the question would be whether or not that piece of writing had increased in accuracy by half, because one of the clauses of each T-unit was correct, or not at all, because the T-unit as a whole were still incorrect. Additionally, the volume of information being generated by the accuracy measurements had to be taken into account, particularly when these addressed only part of the study. The measurement of EFC is considered as a measure of linguistic proficiency, but from the perspective of linguistic complexity rather than accuracy (Polio, 1997; Inoue, 2010). This study is concerned with proficiency as measured through accuracy.

T-units were examined twice, once at the initial draft stage and once after the revision. In addition to this, the more detailed phase of the error analysis generated a much greater amount of information which had to be collated and analysed. In the cases of the baseline and final extended writings, only one count needed to be made as there was no revision phase.

3.4.3 Tests of inter-rater reliability

The point mentioned above in section 3.4.1 regarding the accurate classification of errors also applies to the accurate identification of TT and EFT. Subjectivity may play a part in the classification of either, and so a statistical test was employed to measure inter-rater reliability. That is, when two raters looked at measures of accuracy and numbers and types of error independently of each other, was there a difference between the two raters significant enough to suggest that the measures were unreliable or that particular areas were likely to cause problems which could undermine the subsequent stages of the experiment?

Once the baseline writing task had been written by both groups, it was used as the basis for an inter-rater reliability test to determine if any aspects of grading the pieces in terms of TT, EFT or types of error were likely to prove problematic. The rest of the writing

was not marked by two raters, largely due to the volume and timings involved in the turnaround between draft and revision, meaning the main objective of the reliability rating was to examine whether my rating corresponded to a second party's measure, establishing whether or not my ratings were objective and reliable enough for the remainder of the experiment.

In order to do this, 4 separate measures were taken by each rater: total number of T-units (TT) and error free T-units (EFT), to establish reliable reports of accuracy, and number and type of errors. Once these were collated, they were subject to two-tailed independent t-tests with the IBM Statistical Package for the Social Sciences (SPSS) software (IBM, 2012). These tests established whether or not there was any significant variation away from the mean value in each case. SPSS is a statistical testing package which is widely available through higher education institutes. As such, it is currently the software most often employed in studies relating to social sciences (Acton and Miller, 2009). It is the software package which was used to conduct all the statistical tests in this thesis.

The t-test itself, not to be confused with a T-unit, is a statistical test which can be used to examine whether or not the mean averages between two sets of independent samples are significantly different or not (Acton and Miller, 2009; Garner, 2005). For the purposes of my context, any indication that there is no significant difference between means suggests that inter-rater reliability can be assumed. The test employed here is non-directional, which means my null hypothesis predicts a difference between the inter-raters, but I make no prediction about whether inter-rater A or B diverges from the mean. Statistical significance is judged to be present at a level of lower than 5%, or <.05, as is common throughout the social sciences (Bryman and Cramer, 2005, p. 135; Butler, 1985, p. 73).

Error	Levene's Test For equality of Variance Sig.	t-test for equality of means		
		t	df	Sig. (2-tailed)
TT	.993	.339	50	.736
EFT	.247	.188	50	.241

Figure 3.7: *t-test for inter-rater reliability tests on accuracy (TT and EFT).*

The null hypothesis for each test is that there **is a** significant difference between the marks provided by each rater. Figure 3.7 shows the inter-rater reliability results for accuracy in terms of TT and EFT. The t-test for number of TT showed a non significant result in Levene's test for equality of variance (Sig.993, >.05), so equality can be assumed. The subsequent t-test for equality of means produced a result of (t (50) = .339, p.736 (>.05)). This represents a non-significant result, so the null hypothesis can be rejected. In other words, there is no significant difference in the marking of each rater.

In terms of EFT, Levene's test for equality of variance was non-significant at (Sig.247, >.05) so I can assume equality. The t-test gave a result of (t (50) = .188, p.241 (>.05). Again, a non-significant result indicates that there is no significant difference in the way EFT are being calculated. These results satisfied me that the way t-units and error free t-units were being judged was reliable and valid for the remainder of the experiment.

The table in figure 3.8 summarises the results of the inter-rater reliability tests for differences in error evaluation. It shows type of error and frequency per rater.

	WW	F/I	^	G	T	Agr	P/S	C/U	A	WO	WF	?
Rater A	96	122	43	29	9	12	7	10	12	6	34	6
Rater B	121	142	39	32	9	19	3	11	29	4	43	8
Rank Rater A	2	1	3	5	8	6	9	7	6	10	4	10
Rank Rater B	2	1	4	5	9	7	12	8	6	11	3	10

Figure 3.8: *Error frequencies and rankings between inter-raters.*

Counts between raters was similar, though rater B generally noted a higher number of errors than rater A. Rankings were also similar but not equal. Reasons for this are discussed below the summary of statistical tests table. Again, the test employed was the independent two-tailed t-test, with the null hypothesis being that there would be a significant variance in marks between the raters, as shown in figure 3.9.

Error	Levene's Test For equality of Variance Sig.	t-test for equality of means		
		t	df	Sig. (2-tailed)
WW	.141	-1.242	50	.220
F/I	.776	-1.017	50	.314
^	.111	.357	50	.722
G	.332	-.253	50	.801
T	.554	.000	50	1.000
Agr	.055	-1.414	50	.164

P/S	.040*	.988	34.719	.330
C/U	.418	-.198	50	.844
A	.380	-2.098	50	.041
WO	.167	.693	50	.491
WF	.040*	-.836	50	.407
?	.327	-.488	50	.628

* Equal variances not assumed.

Figure 3.9: *t-test for significant variance between error ratings*

Of all the error types, the only area in which there was a significant difference was with article errors. For every other error, the null hypothesis could be rejected, so I could assume that the identification of TT, EFT and errors themselves was reliable for the remainder of the test.

Following the collation of figures, the raters discussed the results and areas which, though not suggested as problematic by the tests, were nevertheless noted as potentially confusing. A similar procedure was employed by Polio (1997) and was used, as it is here, to more clearly define the parameters of particular errors. What follows is a set of identification rules which were employed for the remainder of the study.

3.4.4. Further Definitions of Error Types

This section discusses the distinctions between tense (T) and grammar (G) errors, missing words (^) and article errors (A), wrong words (WW), errors of register (F/I) and errors of word form (WF).

- **Tense (T) and Grammar (G)**

Mark ups for tense referred to time and aspect rather than the incorrect form of the verb. That is, the following sentence, the error was marked as one of tense:

When I was young, I will go to the countryside.

'Will' directly contradicts the time and aspect of the sentence as a whole. Grammar (G) errors are more accurately thought of as construction errors. The following sentence, however, is an example of a grammar error:

If the government reduces taxes, it would be more popular.

The if-clause requires the complement to use will + infinitive as part of a first conditional. The use of would still refers to future possibility, so it is not an error of tense, but is instead the wrong clause construction, indicative of the second conditional.

- **Missing words (^) and articles (A)**

Missing words included missing articles. Only if an incorrect article had been used, rather than no article, was the error marked A. The following example is a missing word error:

'Different parts of economy'.

Whilst the next error is an article error:

'Individuals are part of the society.'

- **Wrong Word (WW), Word Form (WF) and Formal or Inormal (F/I) errors**

Completion of inter-rater reliability tests showed that the distinction between simply a wrong word (WW), word form (WF) and a formal or informal register (F/I) was not statistically significant, but discussion produced a number of potential issues for the analysis section. The issue was that all of these errors had the potential to be defined as wrong word, and that whilst distinction between many of them could be made, with some of them the distinction was extremely subjective. Word-form and formal informal could effectively became subordinates of the superordinate wrong word.

A particular type of informality was the use of pronouns, in particular when forming a structure which expressed an opinion or evaluation of some sort. These were readily identifiable as a particular type of wrong word and consistently marked as 'formal / informal', examples being 'I think' and 'we want to'.

More ambiguous and subjective were those words which were could be labelled as informal precisely because the word itself which had been selected by the learner was deemed to be wrong because of its register.

For example, 'to battle for' could be judged as both informal (F/I), in that it perhaps sounds too emotive, or simply as a wrong word (WW), in that it contains a similar meaning to the word which it was eventually corrected to (compete for), but has merely been placed in an inappropriate context.

Similarly, use of phrasal verbs was judged to be both informal in error type, or simply just the wrong word. In terms of learner response, I did not deem this to be of particular importance; the correction would be to an alternative word whatever code had been applied. However, this did highlight an area that needed more thought in terms of the analysis,

because it suggested that for these lexical items, the coding was too broad. This may have meant that the future analysis was not precise enough in identifying lexical sub-categories.

Potential problems in marking items as wrong word or word form was also noted. 'Producing activity', or the use of the noun rather than adjective in 'productive activity', for example, were identified generally by one marker as 'Word form error', presumably because the meaning was clear from the context and only the form itself was wrong. On the other hand, it could also be marked simply as the wrong word with just as much justification. Another incidence was 'works' and 'workers'. 'Responsibility to their works' rather than 'responsibility to their workers' was marked by one marker as word form because of the shared root of the two words, and wrong word by another marker due to the fact that the meaning actually changed ('responsibility to works' implying a workers responsibility towards their factory' and, on the other hand ' responsibility to their workers' implying a company responsibility to the staff).

In summary, the correction chart, whilst unlikely to have had any effect on the learners' response to corrections was not deemed entirely sufficient to either be able to accurately differentiate between some types of error, and didn't distinguish between a number of evident subcategories which are detailed below and which were used in parts of the analysis. This area is also addressed in chapter 5 when discussing the limitations of the experiment.

3.4.5 Analysis of Wrong Word (WW) Errors and Formal / Informal (F/I) Errors

All wrong word and informal / formal errors were analysed in further detail to try and ascertain more data as to specifically what type of error had been made and how effectively or ineffectively these errors had been corrected. This action was taken because, as described above, I became aware of the potential for the analysis to become too blunt if these lexical

categories were examined in a broad sense and also because, as data was collated, it became clear even from a review of the raw figures that it was precisely these lexical areas which were the cause of the greatest number of errors.

These lexical errors were therefore subsequently split into the following categories:

- F/I: Lexical errors which were register related - the use of statements including subjective pronouns and statements which contained explicitly inappropriate constructions - 'an old seer once said that the wise man'.
- Emotive: Lexical errors using language which, whilst correct, was too emotive or subjective in nature.
- Prep: Lexical errors which were prepositional in nature.
- WF: Lexical errors which were errors that could clearly be identified as those of word form.
- WW/ CON/ SYN: Lexical errors which were wrong word errors. Of these, there were three types: an entirely wrong word (WW): something which had no place in the construction); a word which supplied, within the construction, a meaning contrary to that which was being sought by the learner (CON); a word which conveyed the correct meaning but was incorrect in terms of context or co-text and required a synonym for correction (SYN).

The results of this analysis are detailed in chapter 4, 4.7**.**

3.5 Confounding factors

I mentioned in the introduction to this chapter the presence of a number of confounding factors. These are perhaps to be expected in an experiment which is conducted with real students on a real EAP pre-sessional course, rather than under conditions with more stringent

controls. Below, I detail these confounding factors and look at both their implications and the measures which were taken within the experiment design, in terms of controls, and in the analysis, to mitigate any negative effects upon the results as much as possible.

3.5.1 Participants

The subjects were all part of a purposive, non-randomised sample. Several factors contributed to the impracticality of obtaining an entirely random sample. The main factor was that of the institution allocating learners to particular groups. As already noted, this was done based on a consideration of student level at entry, the eventual target department of the student and the total numbers of students, both male and female.

The institution had knowledge of the student level prior to their arrival in the UK based on their IELTS (International English Language Testing System). In the case of this study, this had little effect on the placement of students as all entered the course with either an overall band 6 or 6.5 score. In other words, a score which was half a band away from their course requirements. As such, all students had similar levels in terms of IELTS. It could be argued that IELTS itself is not a good measure of actual ability due to it being composed of various general descriptors which are collapsible and do not provide a detailed analysis of why, for example, a student might have writing problems. However, this was the institutional mechanism of judging ability in English.

The target department was of much greater significance in the grouping of students. Whilst students might have been studying different individual courses, all were grouped according to the department they would be entering and the level of study, be it MA or PhD. This allowed for the allocation of materials and teachers which would be most suitable for those students and was the institution's attempt to mitigate the EGAP/ESAP divide.

The sample was purposive, then, in the sense that I knew that the students would be studying MScs with the Management English School. Summer intakes from previous years suggested that the classes would be formed mostly, if not entirely, of Mandarin speaking mainland Chinese learners, and this turned out to be the case. This meant that before the course began, I was able to form an idea of a target group to study and the context of the experiment itself – that of mainland Chinese Mandarin speakers engaging on a six week pre-sessional course in order to enter management, finance or economics departments for the study of one year taught masters.

Several factors resulted from the profile of the students themselves. Age varied, although all students were in their early twenties. Educational background also varied, although all students had completed a bachelor's degree in mainland China.

Learner type was another consideration taken into account. As explained, each learner was given the Honey and Mumford test (1992) and their responses recorded to factor into later analysis. In the event, there was not a huge amount of variation in learner type.

Exposure to English was yet another confounding factor which had to be taken into account. For all students, this was the first time they had ever lived or studied in an English speaking country. However, all students had experience of using English within a university environment; as part of their first degree studies, students had taken compulsory English classes, though these were separate from their major area of study. That is, their majors had been taught in the mother tongue and English lesson were supplementary to this.

Prior to university, all students had learnt some English. The length of time spent studying English varied. The majority of students had been studying English for about ten years, equating with their entry into middle school or junior high. Other students, however, had had several more years and appeared to have been studying English in junior school. All

the students had arrived into the UK at the same time, the weekend before the study commenced.

The nature of these nuisance factors meant it was possible select and structure statistical tests which accommodated them as independent variables.

3.5.2 Instructors

Ideally, the experiment would have involved the researcher observing and performing the research without being involved in the teaching process itself. This was not possible and I had to research and teach, meaning measures had to be taken to reduce teaching bias as much as possible. One of these was being in control of both the groups' reading and writing input. Identical lesson plans and materials were given to both the CG and EG, and all learners received identical writing and reading homework. The only timetabling difference was the point at which the experimental group was taken to be trained in the use of the corpus. Practically speaking, total duplication in the lessons was impossible due to interaction with different students, but all lessons had the same learning goals and target language.

Speaking and listening classes were managed by two other teachers. These colleagues were aware of the experiment and the fact that I was trying to limit the amount of confounding factors as much as possible. They too kept to the pattern of delivering the same lessons to both groups of students. In fact, this pattern of lesson repetition was a standard operating procedure on the pre-sessional course and the timetable was arranged to accommodate it in order to reduce the workload of teaching staff. These teachers were asked to direct, as far as was reasonably possible, any queries students may have had about their writing to the researcher. It is to be expected, however, that planned or incidental language input relevant to learners' writing ability occurred during the listening and speaking lessons.

Three afternoons per week, students were assigned to exam classes with the three teachers. A level of control was lost here as students were assigned at random to different exam groups. Here, they received input on exam skills related to IELTS and USEPT (University of Sheffield English Proficiency Test), at the time itself heavily based on the IELTS exam. Confounding factors were limited by teacher rotating the groups and repeating the same lessons again.

There were other factors which could not be accounted for and remain limiting factors. As all students were eventually going to enter the management school, they were all grouped for their English classes in the same building. This meant at break time and lunch time there was ample opportunity for students to socialise with each other and, possibly, discuss the content of their classes. Transitional meetings arranged between the English Language Teaching Centre and the department meant that students were also able to meet students whom they would be studying with for the next year. Similarly, the accommodation provided for students consisted of halls of residence where all students were grouped together, again meaning interaction of members of different groups was all but unavoidable.

The experiment meant that students drafted a piece of work as homework and submitted it to the researcher, and then revised the errors before submitting it to the researcher. Students completed these tasks to the timetable outlined in section 3.3.3. Homework was issued on Friday. The draft was submitted the following Monday. Copies marked up with errors were provided to the students on Wednesday and the revisions were returned to the teacher on Friday, whereupon the process started again. The periods when the students were writing and, perhaps more importantly, correcting, were unsupervised by the researcher. Initially, I had considered supervising the drafting and correction process in class time but this was rejected. Whilst this approach would have provided a high degree of control, the two main drawbacks were time and, crucially, ethical considerations.

There were two main time considerations. The first was that the students had a syllabus to cover in the weeks that they were on the pre-sessional course. There was simply not enough time in the programme to engage in in-class drafting and revision sessions. The second factor was that giving a time limit dictated by the length of the class, which would have been a maximum of ninety minutes, would have for all practical purposes changed the drafting sessions into something resembling exam practice sessions, not something the research was investigating. It may also have resulted in, of course, many hundreds of unfinished drafts or unfinished pieces of revision, risking the collection of any data at all.

Ethical considerations are linked to those of time. Whilst a degree of control was sought by the researcher, the experiment existed as part of a course and it would have been entirely unethical to force the students to spend class time on the stages of an experiment at the expense of other areas of input and learning. Again, lessons resembling, from a student's point of view, test conditions may have prompted feelings of stress and other negative affective factors.

In terms of academic writing and the aims of EAP as a whole, outlined in the literature review, staging several timed sessions would have been contrary to the course goals. Whilst an element of control was lost, then, and students performed the drafts and revisions in their own time and at their own, unmeasured pace, this was an unavoidable factor. As far as was possible, control was maintained by having (a) the cycle adhered too within a timetable of a maximum of 72 hours (the weekend draft period) and (b) errors marked with the method of correction by the student.

The same can be said of the independent assignments. The very nature of these and their focus on student research skills, autonomy and the use of the university facilities meant that these pieces were almost entirely in the hands of the students. In week 3, there was some feedback from me but this was limited to comments on structure and progress of research in

terms of reading and synthesis of ideas and students were told that the responsibility for accuracy of language output was theirs. Extended assignments were, as a matter of departmental protocol, entered into the Turnitin programme to check work for plagiarism.

Whilst these factors need to be taken into account when reviewing the reliability of the results, I would suggest that they were largely unavoidable, particularly when it is remembered that I established at the beginning of the thesis that the experiment should gather information from a practical and realistic teaching context rather than a more artificially controlled environment. These limiting factors will be returned to in the discussion section, where their potential effects on the measurements of corpus use, and effects on corpus use itself, will be addressed.

3.6 Collation of Data and Analysis

Each feedback and revision cycle created several distinct sets of data. For every draft, a count was made of TT and EFT as accuracy measures. The number and types of error were also logged. Once the revision had been completed by the learner, TT and EFT were counted again to log any change in accuracy. The error corrections which had been made were also recorded and logged as correct, incorrect or not-attempted. Further to this, the method of correction which had been utilised for each correction was logged. Figure 3.10 shows a typical log of one learner at an individual cycle stage.

TT	*EFT*	*Error Type*	*Error Details*	*Correction Attempted*	*Method of correction*	*Correction correct?*	*Why Incorrect?*
16	Draft 10	P/S	series of setback	setbacks	K	Y	
	Revision 12	^	downturn implies unemployment rate will rise	Addition of 'the'	K	Y	
		F/I	totally miserable	dreadful	C	N	Also too informal
Word Count 311		A	The social disorder…	A social disorder	K	N	No article needed
		A	The global attention	A global attention	K	N	No article needed
		WW	notorious actions	uncontrollable actions	C	Y	
		^	would have worse impact	Addition of 'a'	K	Y	
		WW	with 'displeasure at the prospect' are	who have displeasure at the prospect	K	Y	
		WW	abyss of despair	NOT ATTEMPTED	N/A	N/A	
		WW	miss the future	NOT ATTEMPTED	N/A	N/A	

Figure 3.10: *Data log for participant EG1, cycle 2.*

This process was repeated for each cycle for each learner. The pre-course task, or baseline data, and extended assignment, or final stage data, were treated in the same way, though with omission of the revision details as these only existed as drafts. Once this data had been logged, collations were made to give a summary of data for each learner across the period of the experiment. From these tables, the data could be manipulated for analysis and for input into the SPSS statistics package. These tables provided the basis for examining research questions RQ1 and RQ2, whilst the data collected from the surveys and questionnaires mentioned earlier provided the data for RQ3.

3.6.1 Statistical Analysis: Regression Models

This final section will examine the statistical method which I employed in the analysis, describe how the test works and explain why it was suitable for the type of questions, context and data in the experiment. I also provide an example of SPSS output as a working model to illustrate how the features of the test are being used to explain the data.

The research questions themselves, the purposive nature of the sample taken during the experiment and the range of data collected suggested that a particular type of regression analysis was the most appropriate form of statistical analysis to employ. Regression, in its simplest form, can be explained as type of correlation analysis. Correlation allows the degree to which two variables are connected to be measured and can provide a statistical account of how strong the relationship between these two variables is in the form of the coefficient of determination (Urdan, 2005).

However, whilst simple correlation allows the identification of a relationship between variables, it does not allow for the analysis of whether or not there is a causal relationship between them. In other words, basic correlation tests, dependent of course on what is being tested, cannot necessarily be used to suggest a reason for the relationship between two

variables. Regression analysis overcomes this issue by distinguishing between independent and dependent types of variable. In doing so, regression analysis allows the prediction of one variable, the dependent, based on the values of another, the independent variable (Lowie and Seton, 2013). With a simple correlation, for example, a question based on the data in this study might be simply 'Does learner type correlate with the number of errors made?' Regression analysis allows a conceptual shift, changing the question to 'Given the learner type is known, how many errors are likely?' (Urdan, 2005). Regression implies, therefore, that the analyst is interested in one particular aspect, the dependent variable, or the outcome, and wants to investigate how this outcome is affected by the independent variables.

3.6.2 Multiple Regression

In the example above, there is, according to Urdan (2005, p. 146) little difference in the outcome when using either a simple correlation test or a simple linear regression test, as long as the variables are interval or ratio, because the data involved are relatively limited. The advantage of performing the linear regression analysis in this case would be that one variable, the dependent, could be calculated, or predicted, given the value of the independent variable. However, the limitation which exists here is the number of independent variables being examined. Examining only one independent variable might imply that this is the only factor affecting the value of the dependent variable. Multiple regression methods allow for an extension where the dependent variable can be calculated based on more than one independent value. This is an important consideration given the number of variables involved.

Continuing the example from above, it could be argued that examining the likelihood of errors based purely on the basis of one contributing factor is rather naive. There are likely to be other factors which contribute to the outcome of the dependent variable. This is

important to note, as in section 3.5 I noted a number of confounding factors which multiple regression can take into account as independent variables. Age of the student, for example, might be a contributing factor to a result. Multiple regression formulas allow for the introduction of this next variable in addition to the original. This extends the scope of investigation, because as well as the questions 'Given the learner type, how many errors are likely?' and 'Given the age of the learner, how many errors are likely?', further questions can be asked: 'Do learner type and age, combined, predict the number of errors which are likely?' (calculated through the multiple correlation coefficient); 'When one independent variable is controlled for (when the effect of the other variable is accounted for), is the other still statistically significant? ' and 'Is learner type or the age of the learner a more significant factor in predicting the number of errors which will be made?' Multiple regression allows for the relationship between the interdependent variables and the dependent variable to be examined but also enables the relationship between the interdependent variables themselves to be analysed. When using SPSS, this last is important to remember because significance in the 'variables in the equation' output is calculated by examining each interdependent variable whilst controlling for the effects of the others. Simply put, in the example above, when SPSS supplies the significance values of each independent variable, it is doing so after already allowing for the other. So the value of significance for the number of years spent learning English is given regardless of the learner type. By calculating this SPSS allows both questions 2 and 3 above to be addressed – whether each variable is significant or not and which is more significant than the other. Nisbet, et al (2005, p. 102) also use multiple logistic regression in their study of the relationship between various learning styles and the effects on writing proficiency, citing the usefulness of the test in question is due to its ability to provide an estimation not only of the statistical significance of a number of given variables but also

because of the fact that it provides an estimation of the magnitude of the effect of these variables.

Theoretically, the list of independent variables can go on indefinitely, although, given the number of relationships which can be expressed, this makes the resulting output more complex and considered selection needs to be applied (Erickson and Nosanchuk, 1992).

3.6.3 Logistical Regression

In the examples of normal linear and multiple linear regression given above, the dependent variable has been number of errors, a nominal variable. However, it has already been explained in section 3.4 that I am interested in examining the accuracy of learner language in terms of EFT and the positive or negative effect of particular correction methods used on different error types. This presents a problem, as examination of both the above calls for a dependent variable which is categorical and dichotomous. That is, a T-unit is either correct or incorrect, and likewise a correction which was made with the corpus on a particular type of error is correct or incorrect. In both cases, the test cannot accept any deviation from the absolute.

The solution to this is the logistic regression test. In logistic regression, the questions which can be asked in previous examples still hold true, but the solution is expressed in terms of the probability of whether a case with a particular set of independent variables comes under the category of interest (YES / 1) for example or does not (NO / 0).

3.6.4 Regression and This Experiment

As mentioned in the introduction to this section, both the quantitative research questions themselves and the nature of the sample taken suggest the use of regression. The sample is one of real learners with a variety of traits, backgrounds and personalities, and this in itself

means that these variables need to be taken into account during analysis. Multiple regression allows for age and years spent learning English to be input as scalar (numerical) data, and gender and learner type to be entered as nominal (categorical) data. In addition to this background data, the additional data from the experiment, can be added such as group, point in time (baseline, draft, revision, final) number of errors, total number of t-units, type of error and method of correction as further independent variables, depending on the types of question being asked of the test.

Selection of binomial, or binary, logistic regression was dictated by the research questions themselves. That is, the category of interest. In all cases when dealing with data generated from student writing, this was dichotomous, falling into one of two categories: CORRECT (1) or INCORRECT (0). In the case of looking at accuracy of T-units, these were either error free or not (1 or 0 in categorical terms). Similarly, when examining the effect of a corpus corrections, the category of interest was CORRECT (1) or INCORRECT (0) and the independent variables were the error types. This allowed for the examination of whether the corpus was more or less effective on particular types of error. Numerous variations of the dependent variable are possible and, where made, these are noted during the analysis. What follows is an example of a binomial logistic regression test made upon data from the experiment in order to illustrate the parts of the output which I used to interpret the data. It must be stressed that the data used here is for illustrative purposes only; it has been collated to give an example output and is not representative of the actual results and nor is it repeated in the analysis.

3.6.5 Example Test and Analysis

For this test, I have decided to examine what is a predictor of an accurate T-unit (EFT), and whether one factor might have more an effect than another. Therefore, my dependent variable

is EFT – Yes or No. The independent variables, those factors which might predict the outcome of the dependent variable, are:

- Learner type: reflector or pragmatist.
- Group: CG or EG
- Gender: Male or female
- Cycle: Baseline or final

Figure 3.11 is a summary of the information I have submitted to the SPSS database. The percent value of 100 tells me I have not omitted any data in error. As this has to be the case for every single test, in the analysis this information is taken as given and is not reported upon, other than the appendices.

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	993	100.0
	Missing Cases	0	.0
	Total	993	100.0
Unselected Cases		0	.0
Total		993	100.0

Figure 3.11: *Case processing summary*

Figure 3.12 simply tells me I have coded my dependent variables correctly, with 'Yes' being the positive value; figure 3.13 simply supplies the encoding for the independent variables. The frequency is the number of items in that category counted. Again, this encoding is not reported in subsequent analysis as it is merely descriptive of the data which has been input.

Original Value	Internal Value
No	0
Yes	1

Figure 3.12: *Dependent variable encoding*

		Frequency	Parameter coding
			(1)
Learn_Typ	Reflector	815	1.000
	Pragmatist	178	.000
Group	Control	472	1.000
	Experimental	521	.000
Gender	Female	780	1.000
	Male	213	.000
Cycle	Baseline	181	1.000
	Final	812	.000

Figure 3.13: *Categorical variables codings*

Before the results proper are supplied, the test supplies information on the data before the test is applied. Figure 3.14 shows 'Variables not in the Equation'. This is a simple correlation table. It shows the significance of each independent variable on the dependent variable. However, this is the significance of each individual variable without adjustments being made for the other variables. The main purpose of this output is to simply highlight any changes made once the logistical regression test has been applied, thus justifying the application of the

test. For example, at this stage, age is not significant (p>.05). Once the test has been applied, as will be seen below in figure 3.15, this changes. Again, because this does not relate to the eventual statistical data, this aspect is not reported on in the analysis section.

	Score	df	Sig.
Cycle(1)	56.368	1	.000
Group(1)	5.670	1	.017
Age	.019	1	.891
Gender(1)	.062	1	.804
Learn_Eng	.559	1	.455
Learn_Typ(1)	15.902	1	.000
Overall Statistics	79.468	7	.000

Figure 3.14: *Variables not in the equation*

The final figure 3.15, 'Variables in the Equation', is the analysis proper. These are the results once the logistic regression test has been applied. This now shows the significance of each independent variable whilst factoring in the effect of the other variables at the same time. Throughout the chapter detailing the results, it is this information which is shown and interpreted.

	B	S.E.	Wald	df	Sig.	Exp(B)
Cycle(1)	-1.432	.203	49.784	1	.000	.239
Group(1)	.255	.276	.850	1	.357	1.290
Age	.404	.141	8.243	1	.004	1.497
Gender(1)	-.254	.192	1.753	1	.186	.776
Learn_Eng	.014	.038	.140	1	.709	1.014
Learn_Typ(1)	-.897	.235	14.621	1	.000	.408
Constant	-9.140	3.353	7.430	1	.006	.000

Figure 3.15: *Variables in the equation*

First, I will deal with the constant. This has no meaningful interpretation in terms of this analysis, or any of those made in the subsequent results section. The constant represents the value the dependent variable would be given if every other independent variable simultaneously had the value of 0. It is only included in the model because this is a default setting in SPSS and is included by way of common practice (Acton et al, 2009: 261), and is ignored in subsequent reporting.

The Sig. column shows that three independent variables have a significant effect upon whether a T-unit is accurate or not: If it has come from the Cycle (1), p.000, the age of the participant, p.004 and learner type (1), or reflector, p.000. All of these are significant at the p.<.05 level. The other independent variables, for now, can be rejected as factors in significantly affecting accuracy.

Note that the df column, degrees of freedom, is 1. This is because the independent variables are either binary (gender, for example), or the number of variables has been restricted (there are many learner types, for example, but in this analysis I only utilised two). Because of the type and arrangement of data collected, this is a common feature of the results section. The S.E. column presents the standard error of each independent variable. A result

which was significant at p.<.05 would suggest that the sample may differ from significantly from the results expected to be found in a larger population and would invite caution as to any claims which could be subsequently made as to the reliability of interpretation (Garner, 2005: 111 – 115).

The Wald value is a chi-square value and can be used in conjunction with the Sig. value to help interpret the importance of the independent variable. Acton et al (2009, p. 267) suggest employing this as a ranking device: the higher the Wald value indicating a greater significance. This is obviously useful in the case of 'cycle' and 'learner type' here, as both have identical Sig. values. The Wald values, however, suggest that 'cycle (1) ' is a more significant predictor (49.784) than 'learner type (1)' (14.621).

The beta co-efficient, column 'B', is chiefly of importance because it indicates whether the effect of an independent variable is positive or negative. Understanding this is crucial, because a misunderstanding will supply exactly the opposite interpretation to the correct one. For example, in the categorical variable codings table (figure 3.13), 'cycle' has two values: baseline = 1.000 and final = .000. These are simply the codes SPSS has assigned. If one now looks at the variable in the equation table (figure 3.15), cycle is labelled cycle (1). This refers to 1.000, or the baseline. Looking immediately right into the Beta column provides the figure -1.432. The minus sign is the important part. This lowers the odds of that variable affecting the dependent variable. In other words, 'cycle (1)' is the most significant predictor of accuracy according to the Sig. and Wald values. However, the minus sign tells me that it is at the final stage, not the baseline, that the effect occurs. In other words, learners were more likely to produce EFT at the final stage than at the baseline stage.

The Exp (B) value in the right hand column is directly related to the Beta value and is an alternative way of supplying the same information. If the value in the Exp (B) column is more than 1 means that independent variable has greater odds of influencing the dependent

variable. Lower than 1 means it has less chance. Because of the importance of noting the positive or negative value of the B value, and how this related to the coding of the independent variable, I chose to report the B value rather than the Exp(B) value in my results.

In summary, interpretation of the logistic regression results involved me using the following procedure:

1. Interpreting the significance of the independent variables' p. value using the Sig. column.
2. Evaluating the relative importance of the p. value using the Wald co-efficient.
3. Ensuring that results were reported accurately in terms of having more or less effect on the dependent variable by checking the coding of the independent variable and its positive or negative value in the Beat column.

As mentioned, the variable in the equation tables are included within the results chapter and examined. The full outputs can be found in appendix 10.

3.7 Conclusion

The key features of the experiment are as follows: (i) Two similar groups of Chinese learners in a working 6 week pre-sessional environment participated. There were non-linguistic confounding factors and these have been explained and accounted for; (ii) Quantitative data was obtained from a timed in-class baseline task, 4 separate draft/feedback/revision cycles and an independent piece of work which would operate as the final stage in order to answer RQ1 and RQ2. This data was supplemented with nature of a quantitative nature to address RQ3 ; (iii) The experiment measured accuracy through error free t-units (EFT) and errors through a meta-linguistic coding system. Tests were conducted to ensure the reliability of the measures and, although there remain potential limitations due to the types of error being identified, the measures were judged fit for purpose; (iv) Both because of some of the

confounding factors mentioned, the range of data and the desire for a predictive component to the statistical analysis, binomial multiple logistic regression was used for statistical testing of results.

The following chapter details the results of the experiment, using percentage data to illustrate the changes in accuracy and error-type production and statistical analysis at key points to evaluate the results for significance. This is followed by a detailed summary of the information supplied by the participating students in the form of the questionnaire and interview responses.

4

Results

4.1 Introduction

This chapter is organised into four main sections. The first provides an overview of the effects of the course on the accuracy of both the CG (Control Group) and the EG (Experimental Group). Initially this is given as a statistical summary of the differences between accuracy levels at the baseline and final stages. This is followed by a more detailed description of how accuracy developed between the four cycles of the draft/feedback/revision process. The aim here is to both establish a broad overview of the experiment and to help answer part of RQ2, which asks specifically if the EG group showed any sign of quantitative improvement in accuracy over the course.

Once this overview of accuracy development across the course has been given, section two goes on to examine the types of error which were produced by both groups at different stages of the experiment. In this section, errors which were particularly problematic for learners are identified, and the results from this section serve as the foundation for the third section which examines how particular errors were corrected and the relative effectiveness of the different correction methods employed by the learners. This is in answer to research question RQ1, which asks if the corpus is used in response to particular types of error and, if so, if it is relatively more effective than other correction methods. This third section goes on to examine the prevalence of these errors in the extended writing in order to address RQ2, investigating the extent to which any improvement in the production of accurate text might be judged to be as a result of corpus use.

The fourth section reports on the quantitative data supplied by the learners in response to the questionnaires and interviews which were given in order to provide a basis for

answering research question RQ3, which compares student attitudes with the quantitative evidence and the attitudes expressed by learners in previous studies.

4.2 Initial Comparison of Groups

This section provides an overview of the progression of both the CG and EG over the course of the experiment. Part 4.2.1 describes the accuracy levels in terms of error free t-units (EFT) at the baseline and final stages. 4.2.1.2 then examines this data statistically, incorporating the available variables to examine the statistically significant predictors of improvement over the course.

Cycle	TT n.	Total EFT n.	Mean	95% Confidence interval		Std. Deviation	Skewness	Kurtosis
				Lower	Upper			
Base	384	75	2.88	1.93	3.84	2.37	.480	-.784
Cycle1 Draft	372	75	2.88	1.89	3.88	2.56	1.234	1.743
Cycle 1Revision	-	270	10.38	8.97	11.80	3.51	0.31	.807
Cycle 2 Draft	400	158	6.08	4.86	7.29	3.0	-.061	-.606
Cycle 2 Revision	-	328	12.62	10.89	14.34	4.27	-.946	1.617
Cycle 3 Draft	404	164	6.31	5.16	7.45	2.84	-.179	-.230
Cycle 3 Revision	-	329	12.65	10.98	14.33	4.26	-.379	.113
Cycle 4 Draft	316	155	5.96	4.36	7.56	4.0	.239	-.509
Cycle 4 Revision	-	265	10.9	8.16	12.22	5.02		
Final	1760	894	34.38	28.61	40.16	14.31	.125	-.608

Figure 4.1: *EFT production at all cycles*

Figure 4.1 provides an overview and shows the general accuracy in terms of EFT of the sample as a whole across each cycle of the experiment. Note that the mean accuracy increases from draft to revision stage in each cycle, but that mean accuracy does not increase

from draft to draft. A comparison of cycle 3 and cycle 4 shows that the mean average in the draft and the revision fell at the latter stage. This is also reflected in figure 4.4, which examines accuracy in terms of error free T-units as a percentage of total t-units (TT). The higher mean average of EFT is the final stage is likely to be indicative of a higher production of TT in the extended assignment. In the following section, figure 4.4 shows the difference in accuracy between the two groups is measured in EFT as a percentage of TT. Using a percentage allows a direct comparison to be made between all cycles, including the extended assignment which involved the production of a greater number of T-units.

4.2.1 Summary of Course Effects on Accuracy

As described in the methodology chapter, the pre-course task involved both the CG and EG writing an essay under timed conditions of 40 minutes. The number of words produced varied between 188 and 503 words. The mean average word count for the CG was 237 words. The mean average word count for the EG was 289 words, providing an overall mean of 261 words. The EFT (error free t-units) data was calculated as a percentage of TT (total t-units) in order to measure accuracy. A higher degree of accuracy was shown in the EG at this baseline stage. The EFT rate of the CG ranged from between 0% to 31.25%, with an average of 11% whilst the EG rate was higher, with a 29.18% mean average and a range of between 5.9% and 58.3%. If the scores of both the groups are ranked together, the median EFT value is 19.1%. Slightly less than a third of the CG subjects rank above this median, compared with three quarters of the EG.

These findings appear to indicate that, at the outset of the experiment, the experimental group possessed a higher degree of accuracy and this needs to be taken into account when making observations about increases in accuracy between the baseline and

final stage and, potentially, on the results of the draft and revision procedure. In order to evaluate whether or not whether there was a statistical discrepancy between groups at the outset of the experiment, I applied a logistic regression test accounting for as many variables as possible. This is outlined below in figure 4.1, following a description of the accuracy found in the extended writing at the final stage.

In the extended writing, the CG shows a mean accuracy in terms of EFT of 45.27%, with a range of between 23.3% and 78.10%. This is rather better than the pre-course task, which had an EFT mean of 11%, and is higher than the accuracy rate for all the drafts, examined in more detail in sub-section 4.3, though it is substantially lower than the accuracy rate of revisions.

The EG has a mean accuracy of EFTs at 55%. Again, this is higher than the accuracy rate across drafts throughout the course but not higher than any of the revisions. The range is between 19.10% and 78.9% correct, a greater range than the CG

When the two groups are ranked together, the median EFT is 50.45%. Slightly over one third of the subjects in the control group scored above the median, whilst three quarters of experimental subjects scored above this. Whilst the gap in accuracy between the two groups seems to have closed, the EG appears the most accurate of the two groups.

4.2.1.2 Statistical Analysis of Course Effects

The first set of statistical tests examines three key queries. The first is whether or not there was a significant initial discrepancy between the accuracy of the groups. The second is to examine if there was a learning effect over the course and the third is to establish which independent variables were potentially of significance in causing this effect.

The first test, in figure 4.1, investigates what, when examining the baseline piece of writing and the final piece of writing, had a statistically significant effect on whether a t-unit

produced was error free or not. The dependent variable therefore is 'EFT – Yes or No (1 or 0)'. A full range of independent variables was examined and included cycle, group, age, gender, years learning English and learner type.

As noted in the methodology chapter, elements of the output which SPSS reports on before the 'Variables in the equation' table are not shown here, but are retained in appendix 10.

The table of 'Variables Not In the Equation' is, as discussed in the methodology section 3.8.4, a correlation table of all the independent variables before the logistic regression test has been applied. That is, how each independent variable is affecting the dependent variable but without taking into account the other independent variables. It is shown in appendix 10 rather than here but is worth mentioning as it is perhaps some justification for the choice of test; had logistic regression not been applied, a total of six independent variables seem to have a significant one-to-one relationship with the dependent variable: the cycle, the group, gender, learner type 1 (activist/reflector), learner type 5 (pragmatist) and learner type 6(theorist) . It is in later comparison with the variables in the equation, i.e. once the regression test has been applied, that a difference in results is seen when the interaction of the various independent variables is taken into account.

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Cycle(1)	-1.454	.140	107.087	1	.000	.234
	Group(1)	-.106	.192	.302	1	.583	.900
	Age	.055	.078	.505	1	.477	1.057
	Gender(1)	.132	.115	1.319	1	.251	1.141
	Learn_Eng	.064	.026	5.970	1	.015	1.066
	Learn_Typ			15.366	6	.018	
	Activist/Reflector(1)	.603	.493	1.500	1	.221	1.828
	Reflector/Pragmatist(2)	1.190	.405	8.656	1	.003	3.288
	Reflector(3)	.779	.381	4.182	1	.041	2.178
	Reflector/Theorist(4)	1.251	.481	6.758	1	.009	3.494
	Pragmatist(5)	.704	.389	3.276	1	.070	2.022
	Theorist(6)	.962	.445	4.682	1	.030	2.617
	Constant	-3.416	1.867	3.349	1	.067	.033

Figure 4.2 *Factors affecting production of EFT(full data set).*

Figure 4.2 displays the actual results of the analysis. The Sig. value columns suggest that only the cycle and learner types two and four have an effect which points to significant at p.<.05. Whether one is more significant than the other is displayed by the Wald co-efficients. The higher the Wald value, the more significant that variable is likely to be on determining whether a t-unit is correct or not. Simply ranking them shows that by far the most significant

factor in determining whether or not a T-unit is correct is the cycle. This has a Wald value of 107.087.

The beta (B) coefficients for cycle (1) is negative at -1.454. In plain terms, this suggests that at the beginning of the cycle, or at the baseline, subjects were about one and a half times less likely to produce EFT, or, in other words, to write accurately, than at the final stage. The Exp(B) at the opposite side of the table concurs with this. In this case, the value .234 gives a prediction, suggesting that the odds of being accurate at the baseline point are cut by 23%. This seems to confirm what can be garnered from the initial overview of percentages. At the beginning of the course, the learners were less likely to be accurate than at the end of the course. At this point, this suggests nothing more than the overall learning effect of the course. It establishes that in both groups learning took place and accuracy improved.

What is interesting is that there is no indication that being in one group or the other had a significant effect on whether the T-units produced were error free, with a sig. value of 0.583. In effect, this suggests that a null-hypothesis stating that being in the EG will produce no difference in overall accuracy to that of being in the CG should be accepted.

Of the other variables suggesting significance, it appears that learner type is a predictor of whether or not accurate writing is produced, at p.018. In order of most to least significant, reflector/pragmatist, reflector/theorist, theorist and reflector all have a significance level of p.<.05.

One issue here is that the raw data shows quite a variation at the baseline level between the control group and the experimental group. This could be due to the sampling process and might be an effect of the purposive limitations discussed in the methodology section. This could mean it is difficult to identify any effect of being in a particular group because all the CG baseline scores fall around one area, and all the EG baseline scores fall

around another, meaning a comparison of how those scores improve dependent upon group is difficult. In order to try and validate the result presented above and test its reliability, I conducted a second test in which a reduced set of data was used, shown in figure 4.3. In this test, the number of subjects was reduced by pairing students from each group with similar scores in the baseline stage. This meant that the progression of similar students could be measured. The same statistical procedure was followed.

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	**Cycle(1)**	-1.432	.203	49.784	1	.000	.239
	Group(1)	.255	.276	.850	1	.357	1.290
	Age	.404	.141	8.243	1	.004	1.497
	Gender(1)	-.254	.192	1.753	1	.186	.776
	Learn_Eng	.014	.038	.140	1	.709	1.014
	Learn_Typ(1)	-.897	.235	14.621	1	.000	.408
	Constant	-9.140	3.353	7.430	1	.006	.000

Figure 4.3: *Factors affecting production of EFT (reduced data set).*

In this test, three variables showed a level of significance in determining whether or not a learner produced EFT. These were learner type, age and, again, the cycle. Once again, the Wald Coefficient for the effect of cycle compared to the others was far higher, at 49.784, compared with 14.621 and 8.243 for the other factors respectively. Beta coefficient values were very similar, at -1.432, suggesting a lowering of odds for being correct at the baseline

points, corroborated by the predictive Exp(B) value of .239, meaning the odds of being correct when at the baseline point are reduced by about a quarter.

Again, in terms of accuracy there seems to be little to suggest significance at this point of whether a subject was in the EG or the CG. Initial concerns of an imbalance in equality between the groups can be discounted, as the reduced data set of learners of a similar ability provided a result corroborating the main result of the full-data set test. The fact that age and being an activist/reflector appear to be significant at $p<.05$ is harder to interpret. Because the data set was reduced to fewer participants, it may be that the significance of these factors was inflated.

In summary, at this point there seems to be an overall learning effect produced by the course as accuracy seems to be statistically significantly higher at the final stage. It appears that being part of either the CG or the EG was not a factor in improvement in accuracy, although learner type might be.

4.3 Accuracy Over the Course: The Draft/Revision Cycle

Having established that the subjects involved on the course made an improvement in accuracy overall, but that there was no obvious significance in accuracy between members of different groups at the baseline and final stages, the following section offers a more detailed account of the iterations which the groups went through during the four stage draft/revision cycle. The graph below in figure 4.4 provides a summation of progress whilst the accompanying text describes the progress of the groups at specific stages.

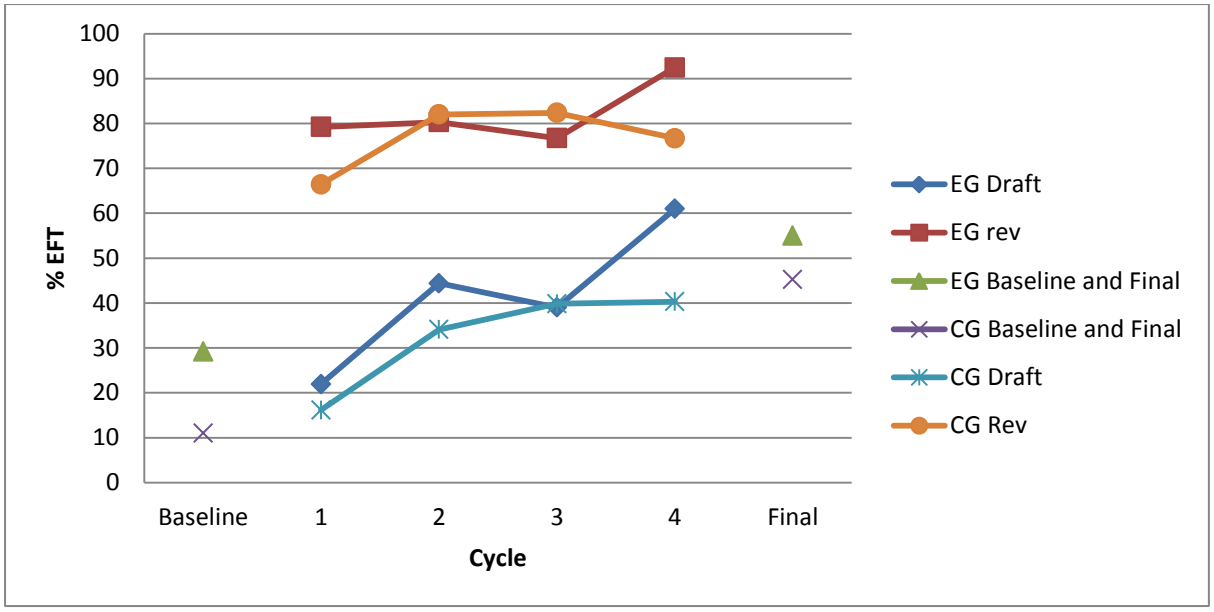


Figure 4.4: *Percentage of accurate (EFT) produced over 6 stages of the experiment*

The graph shows the improvement in accuracy in both groups from the baseline to final stages. It also indicates, as might have been expected, that the revised versions of texts had a higher degree of accuracy than the original drafts. However, it also shows that increase in accuracy was not a straightforward case of consistent improvement and that at points the CG seemed to be performing more accurately than the EG, both at draft and revisions stages. The exact details of these iterations are reported below.

4.3.1 Cycle 1

Draft 1 shows a slight increase in the accuracy of the CG and slightly less of a discrepancy between it and the EG than at the baseline phase. The average number of EFT for the CG is just over 16% with a range of between 0 and 50%. The EG had a mean EFT score of slightly less than 22% and a range of between 0% and 35.7%.

The revision at this stage shows a general improvement in accuracy across all subjects in both groups. In terms of the EFT count, the CG improved the mean accuracy by 50.3%,

rising to 66.45% of T-units being error free, with a range of between 23.10% and 90.9%. The EG improved by 57.59%, slightly more than the CG, with the mean rising to 79.52%. Individual subject scores ranged from 33.3% to 100% correct. In short, throughout this first cycle, the EG performed more accurately than the CG at both the draft and revision stages.

4.3.2 Cycle 2

For both groups, draft 2 represents an increase in overall accuracy. EFTs represent a mean 34.11% of TT in the CG, an increase in mean accuracy from draft 1 of slightly under 18%. The range between subjects is 46.7%, from 13.3% to 60%, representing a reduction in the range by just over 3% from draft and an upper accuracy limit which is 10% higher than draft 1. In the EG, initial mean accuracy doubled, from 21.93% to 44.4%. The spread of scores from 20% to 62.5%, giving a range of 42.5%, which is actually wider than that of the first draft.

At the revision stage, accuracy of T-units within the CG had a mean of 82%, an improvement of 47.89% compared to the draft. Ranges are from a low of 60% to a high of 100%. The revision 2 accuracy rate was 15.55% higher than the revision 1 rate, illustrating both an improvement in the draft itself and an improvement in the accuracy of revisions. The EG also showed an improvement on the draft, with mean number of EFTs increasing by 35.9%.

Both groups increased in accuracy in the drafts compared to cycle 1. Comparison of draft 2 to revision 2 also showed an improvement in accuracy. This improvement is slightly more marked in the EG than the CG, although the range between low and high scores in the EG is greater than that of the CG. The CG revision represents both an improvement on the draft, and also an improvement in the rate of accuracy compared to revision 1. The EG also improved on the draft, although whilst the overall accuracy of the EG was higher, the rate of

increase in accuracy compared to both the draft and the revision in cycle 1 was less than that of the CG.

4.3.3 Cycle 3

In the CG, mean EFT now stand at 39.83%, a rise from 34.11% in draft 2 of 5.72%. The range is from 0% to 66.7%. The EG figures, however, now show a drop in accuracy compared to draft 2. EFT dropped from 44.4% to 39%. This is slightly less than the mean EFT accuracy of the CG. The highest measurement of accuracy here is 52.9%, less than that of the highest measurement in the CG, though the range between subject high and low scores is 36.2%, approximately half that of the CG. .

In summary, whilst the CG showed some improvement in both measures of accuracy, particularly that of EFT, the EG displayed either a decrease in accuracy or virtually no change in accuracy at all compared to the former draft. The EG's draft accuracy was also lower than that of the CG in this instance.

Accuracy in EFT in the CG increased by 42.54% to 82.37% at the revision stage. Range between subjects also dropped to 53.3% . Accuracy in this revision was very slightly higher than that in revision 2, by just 0.37%.

In the EG, accuracy increased by 37.74% to 76.74%. There is also a range of 75% between members, with one subject improving by less than 9%. This level of accuracy was both less than the CG at this stage, and also less than in revisions 1 and 2. In draft 3, then, the CG proved to be the more accurate group. Additionally, the EG results do not show progress was made, with both a drop in accuracy compared to the CG and a drop in the accuracy rate of both the draft compared to draft 2 and the revisions in cycles 1 and 2.

4.3.4 Cycle 4

In the drafts, the CG accuracy increased very slightly to a mean of 40.31%, a less than 1% increase on draft 3. Range is now the highest between subjects, at 71.4%. The EG shows a higher mean EFT accuracy rate than the CG, at 60.98%. This also represents an improvement on all drafts, being over 16% higher than that of draft 2. Range between subjects in EFT accuracy was 69.8%, however.

At the revision stage, CG mean accuracy increased by 36.41% to 76.72% (almost identical to EG mean accuracy in revision 3). This is the lowest accuracy rate of revisions for this group. Range between subjects dropped to 50%. The EG mean EFT accuracy improved by just over 30% to 92.47%. This is the highest mean measure of accuracy, but does not represent the highest change in accuracy between draft and revision. The EG out-performed the CG and reached its highest accuracy level in all four cycles both in draft and revision stages in this cycle. The CG showed an improvement in draft accuracy, but a slightly less accurate revision rate.

4.3.5 Summary

This last section serves to describe the shape of the progression of subjects on the course and provides some background for the overall improvement from baseline to final stages of the experiment. The overview serves to show that, in general, improvements were being made between the draft and the revision at each stage. It also shows that the general trend seems to be that at each stage the drafts were improving in accuracy and the accuracy of the revisions was also improving. What implications these results have in terms of this experiment as a whole and the literature which has already been examined in chapter 2 will be discussed in the following chapter.

The next section will provide an overview of the types of error being generated by both groups, establishing the key problem areas, before going on to examine the methods of correction with which these errors were tackled, how the use of the corpus featured as one of the methods, and how effective the corpus was in dealing with particular errors and in comparison with other methods in order to answer research question RQ2.

4.4 Error Types

This section examines the types of errors being made during the course. Each cycle is examined in turn, with the tables illustrating the prevalence of error types at draft and revisions stages. In sub-section 4.4.6 I provide a summary of accuracy across the course and of the types of errors which were particularly problematic. This data serves as the foundation for section 4.5 which goes on to examine correction methods.

Figures 4.6 and 4.7 give an overview of the frequency of error types at each draft stage and the tables following show the types of errors made at each individual stage of the experiment, including the revision stages. In these tables, starting at figure 4.8, each error type is expressed as a percentage of all the errors at any given stage. The tables are arranged to show the highest ranking error on the left with decreasing rank signified by the position to the right. Figure 4.5 provides a summary of the error coding used in the analysis. As explained in the methodology chapter, section 3.4.1, the coding here is a reduced set as certain errors (spelling, for example) were not included in the analysis.

Symbol	Meaning
T	TENSE
WW	WRONG WORD
WO	WORD ORDER
WF	WORD FORM
G	GRAMMAR

F/I	APPROPRIACY / REGISTER
^	WORD MISSING
Agr	SUBJECT-VERB AGREEMENT
P/S	NUMBER
C/U	COUNTABLE/UNCOUNTABLE
?	SENSE/MEANING UNCLEAR
A	ARTICLE

Figure 4.5: *Summary of error codes*

Cycle	**Error Type**												
	WW	**^**	**G**	**F/I**	**P/S**	**Agr**	**WF**	**?**	**T**	**A**	**WO**	**C/U**	**TOTAL**
Baseline	58	26	26	78	4	9	18	2	3	9	1	8	242
1	84	47	38	57	10	19	0	5	8	16	3	6	293
2	58	30	47	15	11	13	0	2	11	11	1	2	201
3	54	50	36	8	10	7	0	1	7	8	7	5	193
4	27	29	28	7	5	12	0	3	3	2	3	5	124
Final	127	158	87	49	84	50	89	71	50	19	20	9	813
TOTAL	408	340	262	214	124	110	107	84	82	65	35	35	1866

Fig 4.6: *CG group error types*

Figure 4.6 shows the total number of errors produced by the Control Group at the baseline, draft and final cycles, 1866, and shows their distribution across the course. Most frequent errors are at the left and descending frequency is indicated by positioning on the right. Note that this is in terms of total frequency, and that the rank of an error type varies depending on the cycle; in the final cycle, for example, there were more errors of omission (^) than Wrong Word errors (WW).

Cycle	Error Type												
	WW	**^**	**F/I**	**G**	**WF**	**Agr**	**A**	**P/S**	**T**	**?**	**WO**	**C/U**	**TOTAL**
Baseline	38	17	44	3	16	3	3	3	5	4	5	2	143
1	68	15	27	26	0	17	19	8	0	6	0	7	193
2	35	20	16	17	0	3	11	10	3	2	2	2	121
3	30	29	15	16	0	6	10	9	6	1	3	0	125
4	18	13	0	11	0	5	1	2	1	3	1	1	56
Final	97	96	38	14	63	36	25	30	31	20	13	8	471
TOTAL	286	190	140	87	79	70	69	62	46	36	24	20	1109

Fig 4.7 : *EG Group error types*

Figure 4.7 shows the same information as figure 4.6 but for the Experimental Group, which produced 1109 errors in total. Again, rank is determined by the eventual total and there is some fluctuation in different cycles. The analysis below looks more closely at each cycle and the distribution of errors in both the CG and the EG as proportions of all errors produced in order to investigate their saliency.

Error types as percentages at the baseline stage are illustrated in figure 4.8.

CG	**F/I (%)**	**WW (%)**	**^ (%)**	**G** (%)	**WF** (%)	**A** (%)	**Agr** (%)	**C/U** (%)	**P/S** (%)	**T** (%)	**?** (%)	**WO** (%)	**Total** (%)
	32.23	23.97	10.74	10.74	7.44	3.72	3.72	3.31	1.65	1.24	0.83	0.41	100
EG	**F/I (%)**	**WW** (%)	**^** (%)	**WF** (%)	**T** (%)	**WO** (%)	**?** (%)	**A** (%)	**Agr** (%)	**G** (%)	**P/S** (%)	**C/U** (%)	**Total** (%)
	30.76	26.56	11.89	11.19	3.50	3.50	2.80	2.1	2.1	2.1	2.1	1.4	100

Figure 4.8: *Pre-course task error types (% of total errors)*

As can be seen from figure 4.8, the control and experimental groups show similarities in the types of error being produced at this initial stage. Not only do the two groups produce similar types of error, but errors are most prolific in the same categories. In each group, over half of the errors stem from just two categories. Nearly a third of errors in the CG (32.23%) are those of the F/I (Formal/Informal) type, indicating an error in academic register. At 30.76%, F/I errors account for almost the same proportion of the total in the EG. In the CG, WW(Wrong Word) errors represent nearly a quarter of errors, at 23.97%. In the EG, the number is again similar, although here representing slightly over a quarter of the errors produced at 26.56%. In short, the two groups display obvious commonalities in the types of error they are producing at the baseline stage, with lexical issues pre-dominating.

4.4.1 Cycle 1

CG	**WW**	**F/I**	**^**	**G**	**Agr**	**A**	**P/S**	**T**	**C/U**	**?**	**WO**	**Total**
	28.67	19.45	16.04	12.97	6.49	5.46	3.41	2.73	2.05	1.71	1.02	100
EG	WW	F/I	G	A	Agr	^	P/S	C/U	?	-	-	Total
	35.23	14.00	13.47	9.85	8.81	7.77	4.14	3.62	3.11	-	-	100

Figure 4.9: *Cycle 1 Draft error types (% of total errors).*

Figure 4.10 shows that WW (wrong word) and F/I (formal/informal) remained the most prolific errors for both groups, though as a percentage of total errors F/I errors dropped considerably from the pre-course task. Together, these two lexical errors combined again represented the majority of errors, at just under 50%. The six most frequent error types were the same for both groups, although with some difference in the rank order of frequency.

In reporting the revisions, the percentages reflect proportions within the individual type of errors as marked on the draft, rather than proportions of all error types as a whole. The ranking from left to right remains high to low.

Not Attempted (%)												
CG	**F/I**	**G**	**WW**	**^**	**A**	**P/S**	**Agr**	**T**	**WO**	**?**	**CU**	**-**
	17.54	10.53	4.76	4.26	0	0	0	0	0	0	0	-
EG	?	P/S	WW	G	A	F/I	^	Agr	T	WO	WF	CU
	33.33	12.5	7.35	3.85	0	0	0	0	0	0	0	0
Failed (%)												
CG	**WO**	**F/I**	**^**	**G**	**WW**	**?**	**A**	**CU**	**T**	**Agr**	**P/S**	**-**
	33.33	28.07	25.53	23.68	20.24	20	18.75	16.67	12.5	10.53	0	-
EG	F/I	^	?	WW	P/S	G	Agr	A	T	WO	WF	CU
	29.63	20	16.67	14.71	12.5	7.69	5.88	0	0	0	0	0
Successful (%)												
CG	**P/S**	**Agr**	**T**	**CU**	**A**	**?**	**WW**	**^**	**WO**	**G**	**F/I**	**-**
	100	89.47	87.5	83.33	81.25	80	75	70.21	66.67	65.79	54.39	-
EG	A	CU	Agr	G	^	WW	P/S	F/I	?	T	WO	WF
	100	100	94.12	88.46	80	77.94	75	70.37	50	0	0	0

Figure 4.10 *Cycle 1: Revision Outcomes (% individual error type)*

For example, in investigating F/I (Formal/Informal) errors in the CG, the table in figure 4.7 shows that of all the errors marked up as F/I on the draft, 17.54% were not attempted, 28.07%

of attempted corrections failed and 54.39% of corrections were successful. The ranking tells me that, of all types of error, F/I errors were the ones most often not attempted by the CG, they represented the second highest rank in terms of correction failures and, relative to the other error types, the lowest rank in terms of successful corrections. The conclusion which can be drawn from this is that whilst just over half the F/I revisions were successful, F/I errors were still amongst the most problematic relative to other error types. The other problematic error type from the draft, WW (Wrong Word), has a successful correction rate of 75%, and is ranked 5th in terms of failures. In the EG, a similar pattern is shown. Most of the errors were tackled successfully, although '?'(Sense) errors, or those which were responsible for a lack of clear meaning, had a correction success rate of just 50%. F/I was the next most problematic error type but again the success rate is 70.37%, and WW corrections have a success rate of 77.94%.

In summary, at this stage the two groups were fairly similar in the types of error they produce at the draft stage, both in ranked types and as percentage of errors. In the revisions, the same errors, F/I and WW, those which were most problematic in the draft, have similar ranks in terms of correction success, with F/I ranking at the low end and WW ranking somewhere in the middle. Nevertheless, there is still a success rate of over half in all cases for these type of errors.

These errors are important, however, when looking at the types of error which persist after the revisions were made. In the CG, F/I and WW errors represent 31.33 and 25.3% of all uncorrected errors respectively, or over 56% of uncorrected errors. In the EG, these errors account for over 65% of all unsuccessful revisions, or 42.86% WW and 22.86 % F/I.

4.4.2 Cycle 2

CG	WW (%)	G (%)	^ (%)	F/I (%)	Agr (%)	A (%)	P/S (%)	T (%)	C/U (%)	? (%)	WO (%)	Total (%)
	28.85	23.38	14.93	7.46	6.47	5.47	5.47	5.47	1	1	0.5	100
EG	WW (%)	^ (%)	G (%)	F/I (%)	A (%)	P/S (%)	Agr (%)	T (%)	C/U (%)	WO (%)	? (%)	Total (%)
	28.94	16.53	14.05	13.22	9.09	8.26	2.48	2.48	1.65	1.65	1.65	100

Figure 4.11: *Cycle 2 Draft error types (% of total errors)*

Figure 4.11 describes the distribution of errors at the draft stage of the second cycle. WW (Wrong Word) errors again occupied the highest position in rank for both groups, with almost identical rates between the two. G (Grammar) and ^ (Omission) errors still persisted as relatively frequent at the draft stage but F/I (Formal/Informal) errors reduced to just over 7% of errors for the CG and just under double that in the EG, although they still ranked as the fourth highest error type. All other types of error represented just over a quarter of the total errors for each group.

Revision cycle 2, as shown in figure 4.12, shows a similar pattern to revision 1 in both groups. In terms of least successful, or in other words most problematic or resistant errors, WW and F/I again appear fairly high in the rankings. The success rate, however, was still relatively favourable for both compared to the failure rate; F/I and WW in the CG both had correction rates of 73% and 65%, whilst in the EG the success rate was almost 75% for WW and 56.25% for F/I errors.

Not Attempted (%)												
CG	**F/I**	**T**	**WW**	**G**	**^**	**A**	**P/S**	**Agr**	**WO**	**?**	**CU**	**F/I**
	18.18	18.18	10.34	6.39	3.33	0	0	0	0	0	0	18.18
EG	**WW**	**A**	**F/I**	**^**	**P/S**	**G**	**Agr**	**T**	**WO**	**WF**	**?**	**CU**
	11.42	9.09	6.25	4.76	0	0	0	0	0	0	0	0
Failed (%)												
CG	**CU**	**WW**	**G**	**A**	**F/I**	**T**	**Agr**	**^**	**P/S**	**WO**	**WF**	**?**
	50	24.14	23.4	18.19	18.18	9.09	7.69	6.67	0	0	0	0
EG	**T**	**G**	**F/I**	**Agr**	**A**	**WW**	**P/S**	**^**	**WO**	**WF**	**?**	**CU**
	66.67	43.75	37.5	33.33	27.27	14.29	0	0	0	0	0	0
Successful (%)												
CG	**P/S**	**WO**	**?**	**Agr**	**^**	**A**	**F/I**	**T**	**G**	**WW**	**CU**	**P/S**
	100	100	100	92.31	90	81.81	73.34	72.73	70.21	65.52	50	100
EG	**CU**	**P/S**	**WO**	**^**	**?**	**WW**	**Agr**	**A**	**F/I**	**G**	**T**	**WF**
	100	100	100	95.24	100	74.29	66.67	63.64	56.25	56.25	33.33	0

Figure 4.12 *Cycle 2: Revision Outcomes (% individual error type)*

In terms of percentage of errors which remained after the revision process, particular types again seem to be prominent. In the CG, F/I errors represented only 8.33% of errors remaining. WW errors accounted for 41.67% of uncorrected errors whilst G errors represented 29.17% of those remaining. In the EG, WW and F/I errors again represented the largest proportion of those which had not been corrected successfully, with 22.58% being F/I (the same percentage as G errors) and 29.03% being WW errors.

In summary, whilst error correction attempts were generally successful in both groups, those errors which remained after correction were G and F/I but more persistently WW errors, which were also those that ranked highest at the draft stage.

4.4.3 Cycle 3

CG	WW (%)	^ (%)	G (%)	P/S (%)	A (%)	F/I (%)	Agr (%)	T (%)	WO (%)	C/U (%)	? (%)	-	Total (%)
	27.98	25.92	18.67	5.18	4.14	4.14	3.62	3.62	3.62	2.59	0.52	-	100
EG	**WW (%)**	**^ (%)**	**G (%)**	**F/I (%)**	**A (%)**	**P/S (%)**	**Agr (%)**	**T (%)**	**WO (%)**	**-**	**-**	**-**	**Total (%)**
	24	23.2	12.8	12	8	7.2	4.8	4.8	2.4	-	-	-	100

Figure 4.13: *Cycle 3 Draft error types (% of total errors)*

At the draft stage of cycle 3, figure 4.13 shows that WW (Wrong Word) errors still occupied the highest rank for both groups, though closely followed by ^ (missing word) errors. A comparison with the pre-course and previous drafts will show that, though there is a change in order, the same errors occupy the highest positions. Compared to the pre-course draft, F/I (Formal/Informal) errors have dropped noticeably in ranking, particularly amongst the control subjects.

The revision data (figure 4.14) shows that F/I and WW errors again occupied a similar place in the successful correction rankings, with F/I being among the least successful in terms of rank. In the CG, this success rate dropped to 50% whilst in the EG it dropped below this, with only 41.67% of F/I being corrected successfully (keeping in mind, however, that, in the draft during this cycle this error type had a lower ranking). WW occupied a mid-range in the ranking as in the previous revision sets, but again the majority of these errors were corrected successfully, at 70% in the CG and over 60% in the EG. It might be noted at this point the EG were regularly being more successful in the correction rate of WW and F/I errors than the CG. This will be discussed in greater detail below. In the CG errors remaining, 32% are WW, 26% are ^ errors, 16% are G errors, 8% F/I and 8% are T errors. In the EG, 27.5% are

WW errors whilst F/I errors occupy the next highest amount at 17.5% shared with ^ errors and G errors.

Not Attempted (%)												
CG	**?**	**T**	**F/I**	**Agr**	**A**	**WW**	**G**	**^**	**P/S**	**WO**	**CU**	**-**
	100	28.57	25	14.29	12.5	9.26	8.33	8	0	0	0	-
EG	**?**	**G**	**WW**	**^**	**F/I**	**A**	**P/S**	**Agr**	**T**	**WO**	**WF**	**CU**
	100	18.75	16.67	10.71	8.33	0	0	0	0	0	0	0
Failed (%)												
CG	**T**	**F/I**	**WW**	**CU**	**^**	**WO**	**G**	**A**	**P/S**	**Agr**	**?**	**-**
	28.57	25	20.37	20	18	14.29	13.89	0	0	0	0	-
EG	**F/I**	**T**	**G**	**WW**	**A**	**^**	**P/S**	**Agr**	**WO**	**WF**	**?**	**CU**
	50	50	25	20	18.18	14.29	11.11	0	0	0	0	0
Successful (%)												
CG	**P/S**	**A**	**Agr**	**WO**	**CU**	**G**	**^**	**WW**	**F/I**	**T**	**?**	**-**
	100	87.5	85.71	85.71	80	77.78	74	70.37	50	42.86	0	-
EG	**Agr**	**CU**	**WO**	**P/S**	**A**	**^**	**WW**	**G**	**T**	**F/I**	**WF**	**?**
	100	100	100	88.89	81.82	75	63.33	56.25	50	41.67	0	0

Figure 4.14 *Cycle 3: Revision Outcomes (% individual error type)*

In summary, again the majority of errors were dealt with successfully by both groups. F/I errors became less prominent in their ranking after corrections but again WW errors occupied the highest ranking in both drafts and errors remaining after the revision process.

4.4.4 Cycle 4

CG	^ (%)	G (%)	WW (%)	Agr (%)	F/I (%)	C/U (%)	P/S (%)	T (%)	WO (%)	? (%)	A (%)	Total (%)
	23.39	22.58	21.77	9.68	5.65	4.03	4.03	2.42	2.42	2.42	1.61	100
EG	**WW (%)**	**^ (%)**	**G (%)**	**Agr (%)**	**? (%)**	**P/S (%)**	**A (%)**	**C/U (%)**	**T (%)**	**W/0 (%)**	**-**	**Total (%)**
	32.14	23.21	19.64	8.93	5.35	3.57	1.79	1.79	1.79	1.79	-	100

Figure 4.15 : *Cycle 4 Draft error types (% of total errors)*

Figure 4.15 shows that at the final draft stage, , ^ (Omission) and G (Grammar) errors were the most prolific in the CG although there is a difference of only a couple of % between these errors and WW (Wrong Word) errors. WW in the EG still occupied the highest position, at slightly over 32%, almost 10% more than the next most problematic error. Notably, the CG drafts showed a marked drop in the percentage of errors which were F/I (Formal/Informal), at just over 5%, whilst in the EG group this error had been eradicated entirely by this stage.

In the CG, WW and A were the least successful correction areas, with WW at 51.85% successful. The EG corrections here were largely successful, with several 100% corrections and WW now at 94.44 % successful correction rate (figure 4.16) below.

Not Attempted (%)												
CG	**?**	**F/I**	**WW**	**G**	**^**	**A**	**P/S**	**Agr**	**T**	**WO**	**CU**	**-**
	100	85.71	25.93	7.14	6.9	0	0	0	0	0	0	-
EG	**WO**	**?**	**G**	**A**	**WW**	**F/I**	**P/S**	**^**	**Agr**	**T**	**WF**	**CU**
	100	33.33	9.09	0	0	0	0	0	0	0	0	0
Failed (%)												
CG	**A**	**G**	**T**	**WW**	**F/I**	**^**	**Agr**	**P/S**	**WO**	**?**	**CU**	**-**
	50	39.29	33.33	22.22	14.29	13.79	8.33	0	0	0	0	-
EG	**^**	**Agr**	**G**	**WW**	**A**	**F/I**	**P/S**	**T**	**WO**	**WF**	**?**	**CU**
	23.08	20	18.18	5.56	0	0	0	0	0	0	0	0
Successful (%)												
CG	**CU**	**P/S**	**WO**	**Agr**	**^**	**T**	**WW**	**A**	**F/I**	**?**	**-**	**-**
	100	100	100	91.67	79.31	66.67	51.85	50	0	0	-	-
EG	**A**	**CU**	**P/S**	**T**	**WW**	**Agr**	**^**	**G**	**?**	**F/I**	**WO**	**WF**
	100	100	100	100	94.44	80	76.92	72.73	66.67	0	0	0

Figure 4.16: *Cycle 4: Revision Outcomes (% individual error type)*

In terms of those errors which had the least successful correction rates amongst the CG subjects, 28.89% were WW and 28.89% are G errors, whilst F/I failures occupied 15.56% of persistent errors. In the EG, F/I errors were not present in the draft at all, and persistent errors are almost non-existent, with WW accounting for only 10% of errors, G 30% and ^ 30%.

The EG performed better here at the revision stage. In the CG, WW errors share their position as most resistant with A and only accounted for 24.44% of all corrections.

4.4.5 Final Stage

CG	**^ (%)**	**WW** (%)	**WF** (%)	**G** (%)	**P/S** (%)	**?** (%)	**Agr** (%)	**T** (%)	**F/I** (%)	**WO** (%)	**A** (%)	**C/U** (%)	**Total**
	19.43	15.62	10.95	10.70	10.33	8.73	6.15	6.15	6.03	2.46	2.34	1.11	100
EG	**WW (%)**	**^** (%)	**WF** (%)	**F/I** (%)	**Agr** (%)	**T** (%)	**P/S** (%)	**A** (%)	**?** (%)	**G** (%)	**WO** (%)	**C/U** (%)	**Total**
	20.60	20.38	13.38	8.07	7.64	6.58	6.37	5.31	4.25	2.97	2.75	1.70	100

Figure 4.17: *Final stage (extended writing) error types (% of total errors)*

The range of errors across both groups shown in figure 4.17 is similar to that of the baseline satge. F/I errors were less prominent than they were at the baseline stage and in the initial drafts. However, that error which was most persistent throughout the experiment, WW, remained prevalent, particularly in the EG. At the baseline phase, these errors accounted for 23.97% of errors in the CG. This is compared to 15.62% of errors at the final stage. The reduction was less pronounced in the EG, with WW errors constituting 26.57% errors in the baseline task. There was a reduction at the final stage, but this was to only 20.60%.

4.4.6. Summary of Accuracy and Error Types

This section has provided an overview of both the effects of the course in terms of accuracy on the groups and illustrated the types of error which, at the different stages of draft and revision, keep reoccurring.

In general, the EG group was more accurate than the CG at the draft stages of the cycles. Similarly, revisions amongst the EG were more accurate than the CG on the whole, though in cycle 3 this was not the case. In terms of an increase in accuracy from the baseline level through each iteration, the CG actually had a rate of improvement which was greater

than that of the EG. However, this was not borne out in the statistical tests and the factor which influenced greater general accuracy was not belonging to either the CG or the EG, but rather the cycle; the test statistics results point to final stage pieces at the end of the courses being more likely to be accurate than texts at the baseline. This at least seems to indicate a learning effect of the course, though the four cycle stages over the experiment show that progression was not a matter of smooth increase for either group. At this point, any claim for the EG's access to corpus data leading to an increase in accuracy is unfounded.

Just as accuracy improved, progression over the course resulted in a reduction of errors in drafts and an improvement in the accuracy of error revisions. The types of errors varied across both groups at different stages, particularly in the later cycles, but two patterns seemed evident. The first is that lexical errors initially appeared to present the most problems, particular in the case of WW and F/I errors. Not only were these most prevalent at the draft stage, but they proved generally the most resistant to accurate revision. They remained most prevalent at the final stage. As this was particularly true of the EG, lexical errors, and particularly those labelled as WW, provide the focus for the next stage of investigation.

The sections immediately following go on to focus in more detail on these problematic errors and methods of correction. The EG is looked at in closer detail, with usage of the corpus first being established. That is, how much the corpus was used by the EG throughout the draft / revision cycle and what was a predictor of its use, with a particular focus on the types of error which were tackled with the corpus. Having established when and how the corpus was used, its effectiveness as a correction tool is measured by examining the extent to which errors were treated successfully and how the use of the corpus compared in effectiveness to other correction methods within the EG group. I then return to a comparison of the CG and EG in terms of the effectiveness of error revision methods, this time establishing whether or not errors which seem to be more effectively tackled through corpus

consultation within the EG group are more or less effectively corrected by a group with no access to corpus data.

4.5 Correction Methods

The following section examines the correction methods which were applied to errors by both groups. It then goes on to look specifically at the EG and corpus use, examining when the corpus was used, in conjunction with which errors and how effective it was relative to other correction methods. The following tables record correction methods used at the different revision stages, as marked on the revision sheets by the learners. Knowledge corrections ('K') are those made without recourse to any reference materials. Corrections made with a dictionary are marked 'D', with online sources 'O' and the BYU-BNC corpus 'C'. Errors which were not attempted are recorded as 'N/A'.

Correction Method and Outcome						
Cycle	**K**		**D**		**N/A**	**TOTAL**
	Correct	**Incorrect**	**Correct**	**Incorrect**		
Essay 1	182	55	27	8	21	**293**
Essay 2	140	24	17	4	16	**201**
Essay 3	129	27	16	3	18	**193**
Essay 4	74	24	4	1	21	**124**
TOTAL	525	130	64	16	76	**811**

Fig 4.18:*CG correction methods and outcomes.*

	Correction Method and Outcome									
Cycle	**C**		**K**		**D**		**N/A**	**O**		
	Correct	**Incorrect**	**Correct**	**Incorrect**	**Correct**	**Incorrect**		**Correct**	**Incorrect**	**TOTAL**
1	38	8	110	17	6	0	12	2	0	**193**
2	17	4	62	19	8	1	7	3	0	**121**
3	8	1	76	20	3	0	15	2	0	**125**
4	7	0	39	6	1	0	3	0	0	**56**
TOTAL	70	13	287	62	18	1	37	7	0	**495**

Fig 4. 19: *EG correction methods and outcomes.*

Figures 4.18 and 4.19 show an overview of the number of errors at each draft cycle in the total column on the right. To the left, the different correction methods employed are given in addition to a count of whether or not the corrections were successful. Below, the tables present the proportions of errors corrected by particular correction methods and then regression analysis is employed to investigate the factors behind a correction method being chosen and its effectiveness.

Figures 4.20and 4.21 express the methods used as a percentage of the whole number of corrections at each stage. They do not at this stage relate to the effectiveness of any of the methods.

	K (%)	**D (%)**	**N/A (%)**	**Total (%)**
Revision 1	80.89	11.94	7.17	100
Revision 2	81.59	10.45	7.96	100
Revision 3	80.83	9.84	9.33	100
Revision 4	79.03	4.04	16.93	100

Figure 4.20: *CG correction methods (% of all corrections)*

	C (%)	K (%)	D (%)	O (%)	N/A (%)	Total (%)
Revision 1	23.83	65.8	3.11	1.04	6.22	100
Revision 2	17.36	66.94	7.44	2.48	5.78	100
Revision 3	7.2	76.8	2.4	1.6	12	100
Revision 4	12.5	80.35	1.79	0	5.36	100

Figure 4.21: *EG correction methods (% of all corrections)*

In both groups, the favoured method of correction was consistently that of knowledge. The CG had a higher usage of the dictionary than the EG and, overall, a higher rate of errors that are not addressed in the revision process.

The use of the BYU-BNC amongst the experimental subjects never rose above 25%, though it was used consistently more frequently than both dictionaries and other online sources. Corpus usage will be addressed in more detail later in this section, as usage was restricted to particular types of error.

The fact that knowledge corrections were the most prolific of the methods might be explained by the nature of many of the errors. In some cases, even the meta-linguistic coding was almost explicitly providing the answer for the students. This was particularly true in errors such as plural and singular, countable and uncountable and subject-verb agreement. The dichotomous nature of these errors - they are essentially one thing (plural) or the other (singular) - meant that highlighting the error possibly prompted the learners to merely choose the opposite form. Additionally, and this is supported by some of the interview and questionnaire evidence, even more complex or at least less obvious errors could be corrected by students once they had merely noticed the error. Due to this, the fact that the numbers of errors dropped in drafts is unlikely to be solely attributable to a new found accuracy and almost certainly involves some element of an improved discipline in checking work for errors

before submission. It is also highly likely that some of the items marked as errors were actually just mistakes. Indeed, as will be discussed in more detail later, the students of both groups cited the revision and correction cycle as effective in helping them to recognise their individual errors, not necessarily to correct them.

4.5.1 Accuracy of Corrections Made with Reference Sources

Figures 4.22 and 4.23 show the relative effectiveness of the correction methods at each revision stage for the CG and EG respectively.

	K corrections accurate (%)	**D corrections accurate (%)**
Revision 1	76.79	77.14
Revision 2	85.37	80.95
Revision 3	82.69	84.21
Revision 4	75.51	80

Figure 4.22: *CG reference corrections (% correct by method).*

	C corrections accurate (%)	**K corrections accurate (%)**	**D corrections accurate (%)**	**O corrections accurate (%)**
Revision 1	82.61	86.61	100	100
Revision 2	80.95	76.54	88.89	100
Revision 3	88.89	79.17	100	-
Revision 4	100	86.67	100	100

Figure 4.23: *EG reference corrections (% correct by method).*

D and O corrections, in the EG, had a higher rate of accuracy than any of the other correction methods but they accounted for far fewer of the corrections in general (it will be seen that in

the EG these corrections were not as effective on particular types of error). Total combined errors corrected by these means was just 5.25%.

In the CG, the mean success rates for different types of error are very similar, with 80.09% of knowledge corrections being successful as opposed to the slightly higher 80.58% rate for dictionaries. The quota of knowledge corrections in the CG was slightly higher overall than in the EG, at 82.25%. Taken as a mean figure, corpus corrections had a success rate of 88.11%. This was higher than any other correction method in either of the groups (again, however, bearing in mind the fact that this method was not as frequently used as the knowledge method).

In summary, knowledge corrections were the most prolific form of correction in both groups and account for the greatest proportion of corrections. Success rates for knowledge corrections were broadly similar between groups, although the EG had greater success and made less use of the dictionary. Though the corpus was employed far less frequently to correct work than knowledge corrections, it was the second most employed method by the experimental group and had the highest mean success rate. In fact, only in the first revision phase did the corpus prove less successful overall than knowledge corrections.

4.5.2 Application of the Corpus Within the EG

In the section above, it was seen that the corpus was used far less than knowledge corrections, but more than other types. This section examines in more detail the pattern of corpus use. Use is first examined in terms of how often it was employed during different cycles. Subsequently, the type of errors being addressed with the corpus and effectiveness of those corrections is looked at. The tables in the previous section seemed to show corpus corrections as being marginally more effective than knowledge corrections on the whole, but effectiveness of corpus corrections has to be measured against effectiveness of other

corrections for different types of error in order to get a better picture of any potential advantage of using the BYU-BNC over other correction methods.

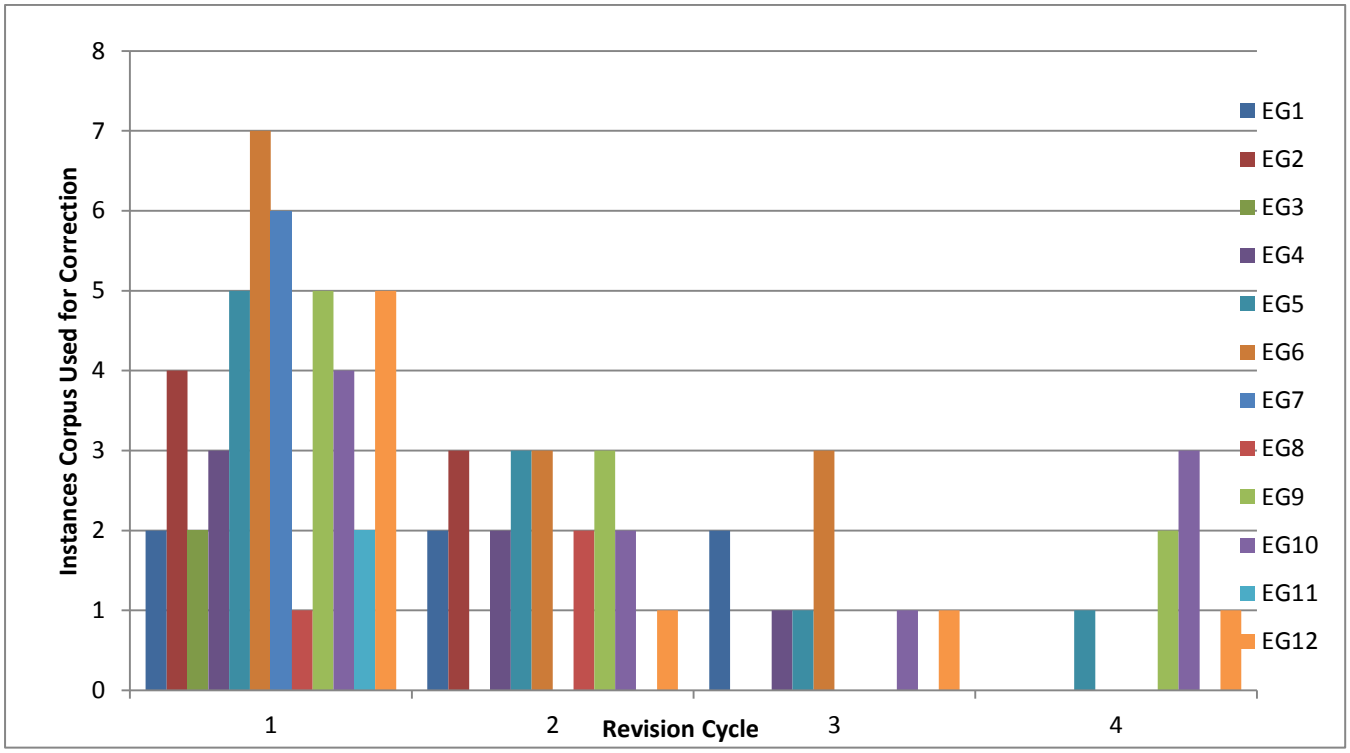


Figure 4.24: *EG Corpus use by cycle and individual learner*

Figure 4.24 illustrates that frequency of corpus use varied quite widely across the EG. The maximum number of times a learner used the corpus was seven instances in any one correction session, and the maximum number of times it was used per subject across the entire course was just 13 times. One immediately observable trend is that, for the vast majority of subjects in the EG, instances of corpus use declined over the term of the experiment. Whilst in the first revision cycle all members made use of the corpus at least once, with a mean average for this stage of 3.83 instances of use, members using the BYU-BNC dropped to 75% in the second cycle, 50% in the third cycle and just over 30% in the fourth.

In addition to the number of subjects using the corpus declining, the instances of corpus use amongst these subjects also decreased. Following the initial revision process, where the corpus was used between one and seven times, the following three stages saw the maximum number of times fall to three per subject. In fact, only four of the experimental subjects employed the corpus in any capacity across the whole course, representing only one third of the group.

The results of the Honey and Mumford (1992) questionnaire are of interest here because they may also suggest an indicator of corpus use. Figure 4.25 summarises learner types and instances of corpus use within the EG.

Learner	**Learner Type**	**Uses of Corpus**	**Number of Cycles Corpus Consulted**
EG1	Reflector	6	3
EG2	Reflector	7	2
EG3	Reflector	2	1
EG4	Pragmatist	6	3
EG5	Pragmatist	10	4
EG6	Theorist	13	3
EG7	Activist/Theorist	6	1
EG8	Reflector	3	2
EG9	Reflector/Pragmatist	9	3
EG10	Reflector/Pragmatist	10`	4
EG11	Reflector	2	1
EG12	Pragmatist	8	4

Figure 4.25: *EG learner types and instances of corpus use*

This data should be examined with reference to the results of the statistical test in figure 4.21 below, which show that predictors of corpus use include being a reflector/pragmatist or a theorist. These learner types represent only 25% of the group as a whole, yet they account for 40% of all corrections attempts using the corpus. The reflector learner type is the most representative of the group, with just under half the learners of this type, but accounts for less than a quarter of corpus corrections. It also seems that reflectors were unlikely to continue with the use of the corpus throughout the course, with no learners of this type using it at every stage.

To a certain extent, the declining trend of corpus use follows that suggested by the general accuracy and correction results examined in section 4.3; in general, accuracy rates in the draft increased as the course progressed and, correspondingly, the number of errors available to correct dropped. One must note, however, that as corrections with the corpus decreased, corrections using knowledge increased. This could be for a number of reasons, including error types more immediately suited to knowledge corrections, increased confidence in use of knowledge, increased aptitude based on classroom knowledge or even increased knowledge as a result of corpus searches made during previous revision cycles. These issues will be discussed in more detail in chapter 5.

Thus far, I have established that the overall trend was a decline in the use of the corpus across the cycles of the course and that knowledge corrections predominated as a form of correction. However, there is also some indication in figure 4.21 above that corpus corrections were, when made, effective. The next section identifies exactly which types of error the corpus was used to correct.

Figure 4.26 shows a statistical test of usage. The question being asked is 'What is a factor in corpus consultation?', so the dependent variable is 'Corpus used – 'Yes' (1) / 'No'

(0). Against this, all the independent variables for the EG were examined: cycle, age, gender, learner type and error type.

		B	S.E	Wald	df	Sig	Exp (B)
	Cycle	.033	.054	.371	1	.542	1.033
	Age	-.546	.281	3.776	1	.052	.579
	Gender(1)	-.398	.408	.948	1	.330	.672
	Lrn_Typ			9.953	4	.041	
	Activist/theorist	1.051	.962	1.193	1	.275	2.861
	Reflector	1.327	.923	2.067	1	.151	3.771
	Reflector/pragmatist	1.929	.940	4.211	1	.040	6.881
	Theorist	2.245	.995	5.094	1	.024	9.441
	Error			34.970	11	.000	
	A	.369	1.169	.099	1	.753	1.446
	WW	1.893	1.073	3.114	1	.078	6.642
	F/I	1.498	1.100	1.854	1	.173	4.471
	P/S	-18.833	7327.784	.000	1	.998	.000
	G	-.076	1.159	.004	1	.948	.927
	^	-1.162	1.286	.817	1	.366	.313
	Agr	-18.934	6993.966	.000	1	.998	.000
	T	-18.787	12494.083	.000	1	.999	.000
	WO	-18.504	17885.136	.000	1	.999	.000
	WF	-17.931	40192.970	.000	1	1.000	.000
	?	-18.946	14012.789	.000	1	.999	.000
	Constant	9.172	6.251	2.153	1	.142	9623.538

Figure 4.26: *Factors influencing corpus use*

The variables in the equation table above suggests that predictors of whether the corpus was employed or not seem to include, at a level of statistical significance which is p<.05, whether the learner had a learning type which reflected that of the reflector/pragmatist (p.040) or theorist (p.024). In terms of error types being a predictor, the results do not immediately suggest that any individual error is a significant predictor at the p<.05 level. However, in the Wald coefficient reading, WW (p.078) ranks as the highest predictor in terms of error type (3.114), with F/I errors (p.173) following at the second highest ranking, (1.854). Perhaps more interestingly, it is only WW and F/I and A (p.753, Wald value .099) errors which have positive beta scores. This suggests that it is only these errors which were likely to have been corrected by the corpus. All the other error types have a negative value, which suggests the odds of them being corrected using the corpus as a method is less likely.

The fact that WW and F/I errors are identified here as potential predictors of corpus use is of some interest. It supports findings, analysed in more detail below, that show the corpus was only used by learners to try and eliminate certain types of error, mostly of the WW and F/I type. The following section will examine how errors of this nature were tackled and, having established when the corpus was used, investigate how effective it was in the correction of errors compared to other types of method.

4.5.3 Efficacy of Corpus Corrections on Certain Errors

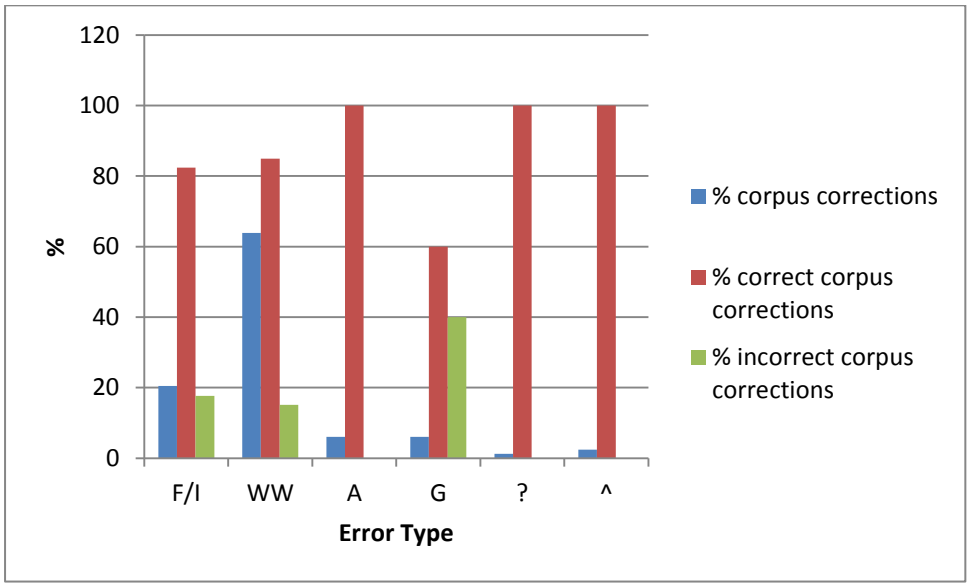


Figure 4.27. *EG corpus corrections*

Figure 4.27 shows that the corpus was used to correct only five of the error types across the whole course. 80% of instances of corpus use were on WW and F/I errors, both types which were mentioned earlier as some of the most persistent errors in the experiment for both groups of subjects and identified by the logistical regression test described in section 4.2 (above) as areas which might be worthy of investigation. Other areas of correction were articles, grammar, incomprehensible areas and missing words, though these accounted for less than 16% of correction attempts using the corpus data.

Accuracy of corpus corrections was consistently over 80% apart from in the revision of errors marked as grammatical errors. 40% of revisions to grammatical errors using the corpus were judged to be incorrect.

The most important area to look at, then, is revisions which were made to errors labelled as wrong word (WW) and secondly, those of a formal and informal nature (F/I). The

following data illustrates the relative effectiveness of corpus and knowledge corrections on WW and F/I errors.

4.5.3.1 WW (Wrong Word) Errors

Of WW errors, the use of corpus and knowledge corrections was proportionately quite close, with 52.68% of errors being tackled with knowledge and the remaining 47.32% being addressed with corpus data. Across the course, WW corpus corrections and knowledge corrections can be split into the following instances (figure 4.28) based on total errors of that type for that cycle.

WW revisions	Knowledge Accurate (%)	Knowledge Inaccurate (%)	Corpus Accurate (%)	Corpus Inaccurate (%)
Revision 1	32.14	3.71	52.83	11.32
Revision 2	52.17	13.04	30.44	4.35
Revision 3	47.37	21.05	26.32	5.26
Revision 4	64.71	5.88	29.41	0

Figure 4.28: *Effectiveness of WW corrections using the corpus or knowledge (% of total corrections)*

Figure 4.28 shows that in the first revision stage, corpus corrections to WW errors actually accounted for more than 60% of the total. In the following three cycles, however, this had fallen to approximately 35%, 31% and 29% respectively, mirroring the overall decrease in corpus usage.

In addition to the effectiveness of each method in terms of corrections as a whole, however, one has to look at the success rate of the corrections within the methods themselves

in order to compensate for the fact that frequency of use differed between the two methods, as in figure 4.29.

WW Revisions	Knowledge Accurate (%)	Corpus Accurate (%)
Revision 1	94.74	82.35
Revision 2	80	87.5
Revision 3	69.23	83.33
Revision 4	91.67	100

Figure 4.29: *Accuracy of EG correction methods (%) on WW errors when measured independently.*

Figure 4.29 highlights that, whilst in revision one, corrections with knowledge had a higher success rate than those with the corpus, the following revision cycles show that, whilst the numbers fluctuated, the corpus corrections were more accurate than those which were made with knowledge. Wrong Word revisions made with knowledge had a mean accuracy of 83.91%. For the same error type, 88.3% of corpus corrections were effective.

In brief, whilst corpus usage declined across the course both in numbers of subjects who utilized the corpus and frequency of use in each revision cycle, corrections of wrong word errors had a mean accuracy rate which was greater when the corpus was used and, in three out of the four cycles, was also greater. The implications of this are discussed in chapter 5, section 5.2.3.

4.5.3.2 F/I (Formal / Informal) Errors

F/I revisions	Knowledge Accurate (%)	Knowledge Inaccurate (%)	Corpus Accurate (%)	Corpus Inaccurate (%)
Revision 1	36.36	22.73	31.82	9.09
Revision 2	35.71	21.43	35.71	7.15
Revision 3	30	50	20	0
Revision 4	100%	0	0	0

Figure 4.30: *Effectiveness of EG F/I correction methods (% of total corrections)*

Use of the corpus was far less than use of knowledge when attempting F/I corrections (figure 4.30). Knowledge corrections always made up the greater share and the use of the corpus dropped markedly in the final two stages. It might be noted that at these stages, F/I errors in the EG had reduced quite considerably in the drafts, meaning there were fewer revisions to be made than, for example, with wrong word errors. Possible causes of this are discussed in chapter 5.

.

F/I revisions	Knowledge Accurate (%)	Corpus Accurate (%)
Revision 1	61.54	77.78
Revision 2	62.5	83.33
Revision 3	37.5	100
Revision 4	100	-

Figure 4.31: *Accuracy of EG correction methods (%) on F/I errors when measured independently.*

The relative success rate of corpus and knowledge corrections on F/I errors can be seen in figure 4.31. Success rates fluctuated (note again that the number of errors of this type are decreasing with each cycle), though a general pattern was increasing accuracy over the first three revision phases in terms of the corpus corrections and the inverse for knowledge corrections.

A summary of correction method use and effectiveness for both types of error is that as the course progressed, corpus use declined on F/I and WW errors. At the same time, accuracy was generally maintained at a rate which was higher than knowledge corrections and, in the case of F/I corrections, which increased over time.

A statistical assessment of the effectiveness of corpus corrections follows in figure 4.32. In this case, only instances of successful error corrections have been input. The question is whether or not the independent variables included have a bearing on whether the successful correction method was the corpus or not. The dependent variable is then defined as 'Correction Method' (Corpus, 1, Other 0).

		B	S.E.	Wald	df	Sig.	Exp(B)
	Cycle	-.213	.153	1.956	1	.162	.808
	Age	-.131	.083	2.481	1	.115	.877
	Gender(1)	-.054	.458	.014	1	.907	.948
	Y_lrn	-.061	.085	.503	1	.478	.941
	Lrn_Typ			5.185	4	.269	
	Activist / reflector	.911	1.063	.734	1	.391	2.486
	Reflector/pragmatist	.962	1.022	.886	1	.347	2.616
	Reflector	1.390	1.000	1.930	1	.165	4.014
	Reflectore / theorist	1.821	1.085	2.818	1	.093	6.180
	Error			31.250	11	.001	
	A	1.004	1.334	.567	1	.451	2.730
	WW	2.312	1.256	3.390	1	.066	10.091
	F/I	2.399	1.288	3.472	1	.062	11.016
	P/S	-18.354	7668.343	.000	1	.998	.000
	G	.185	1.378	.018	1	.893	1.203
	^	-.506	1.438	.124	1	.725	.603
	Agr	-18.451	7463.565	.000	1	.998	.000
	T	-18.283	17877.671	.000	1	.999	.000
	WO	-18.360	17829.330	.000	1	.999	.000
	WF	-18.028	40192.970	.000	1	1.000	.000
	?	-18.477	17803.242	.000	1	.999	.000

Figure 4.32: *Factors influencing success of corpus corrections*

The table above gives a detailed view of whether or not a successful correction of a particular error type was a corpus correction or of another method. In this instance, 'error' rather than any other type of variable is the only significant predictor, but the individual errors are not significant. WW and F/I errors have sig. values of p.066 and 062, so no significance can be claimed at the p.<.05 level. They do, however, have the highest Wald rankings (3.390 and 3.472), indicating that these errors may be more successfully tackled by the corpus than other correction methods. Interestingly, analysis of the Beta value column shows that the majority of the other errors, excepting article errors and grammar errors have a negative value. Though none of the sig. values represent any kind of significant result, the negative direction suggests that these corrections belong to the 'other' methods rather than the corpus i.e. correction using the corpus could be less likely to result in success than by using other methods.

4.5.4 Summary

The use of the corpus as a correction tool declined as the course progressed, both with respect to the number of users employing it and in the frequency with which it was employed by individual members. As an initial answer to research question RQ1, statistical data suggests that significant predictors of corpus use include being either of a reflector or reflector / theorist learner type. No particular error type proved to be statistically significant at the p<.05 level. However, the data shows that the corpus was used to correct only certain types of error, mainly WW and FI, and the statistical test also suggests, through the direction of the beta co-efficient, that most of the other errors encountered were unlikely to be addressed in revision with the BYU-BNC.

In terms of research question RQ2, the success of corrections for these two most prominent error types was higher when the corpus was used than when knowledge

corrections were used. However, when these two correction methods are measured as proportions of corrections overall, the corpus accounted for fewer instances of successful revisions. Statistically, there is nothing which immediately points to successful WW or F/I corrections being significantly more likely when the corpus is employed.

4.6 EG Corrections Compared with CG Corrections

One comparison that can be made is with the CG and the method of correction being used on WW and F/I errors and the effectiveness at different stages of the course. Based on the evidence examined earlier in sections 4.2 and 4.3, which suggests that whilst both groups improved, the control group actually seemed to improve at a rate which was greater than the EG. This section compares the effectiveness of knowledge and dictionary corrections within the CG and shows that the use of the corpus as a reference source in the EG seemed to be marginally more effective than the use of the dictionary as a reference source in the CG.

F/I	**Knowledge Correct (%)**	**Knowledge Incorrect (%)**	**Dictionary Correct (%)**	**Dictionary Incorrect (%)**
Revision 1	55.12	44.88	75	25
Revision 2	90	10	50	50
Revision 3	66.67	33.33	-	-
Revision 4	0	100	-	-

Figure 4.33: *Accuracy of CG correction methods (%) on F/I errors when measured independently.*

Figure 4.33 shows that just over two-thirds of F/I errors were corrected by knowledge. This compares to a similar 80% of knowledge corrections in the EG. As with the EG, the drop in

F/I errors and the lack of a correct knowledge correction at revision 4 stage might be explained by the lower frequency of errors being made at this stage. As with the EG, the success rate fluctuated but in the CG, dictionary corrections actually decreased as contributions to overall accuracy. In the EG, correction with the corpus, despite seeing a drop in usage, saw a rise in accuracy. This, however, is only observable between two cycles and so isn't necessarily a reliable observation.

WW	Knowledge Correct (%)	Knowledge Incorrect (%)	Dictionary Correct (%)	Dictionary Incorrect (%)
Revision 1	75.93	24.07	71.43	28.57
Revision 2	75	25	73.33	26.67
Revision 3	76.47	23.53	80	20
Revision 4	66.67	33.33	80	20

Figure 4.34 *Accuracy of CG correction methods (%) on WW errors when measured independently.*

For wrong word corrections, as shown in Figure 4.34, over 71% of knowledge corrections proved successful. A majority of the corrections made with student knowledge proved effective, though there was a slight drop in accuracy in stage 4, with a mean average of 73.52%. This is about 10% lower than the successful correction average within the EG. The mean success rate in term of dictionary corrections here was 76.20%. This is interesting for two reasons. The first is that this is about 15% lower than the effectiveness of corrections with the corpus in the EG, and is actually lower than the mean correction average of the EG knowledge corrections. Secondly, the gap in effectiveness between the corrections made

with the dictionaries and with knowledge was narrower than in the EG, at less than 3%. In the EG, the gap between corpus corrections and knowledge corrections was just under 5%.

There does not really seem to be a great deal of difference here in using the corpus or in using the dictionaries compared to making knowledge corrections. Both the corpus and the dictionary were between 2 and 5 per cent more effective than relying on knowledge alone.

4.6.1 Summary

For wrong word errors, both groups showed a preference for making knowledge corrections. Whilst these were on the whole accurate, they were not actually as accurate as when these corrections were made with reference materials either in the form of dictionaries in the CG or the corpus in the EG. However, the gap between the accuracy of knowledge corrections and those made with reference materials was slight in both groups, accounting for less than a 5% difference in effectiveness. Again, the EG achieved an overall better rate of success than the CG, but only marginally. Additionally, the data does not suggest that the use of the corpus was very much more effective in error correction than the use of dictionaries. The implication of this comparatively low increase in yield of accuracy is discussed in chapter 5, section 5.2.4.

What follows is a more detailed look at exactly the types of wrong word error the corpus was used to address in an attempt to shed some light onto why or why not corpus corrections were successful and if there is a particular type of wrong word or formal / informal error - synonyms or prepositions, for example – which was targeted more effectively than others.

4. 7 Breakdown of F/I Errors and WW Errors

As wrong word errors and formal informal errors seemed to be the only categories which were corrected using the BYU-BNC corpus on a regular basis, and were error types where the percentage tables showed that the corpus might be an effective tool for revision, I decided to examine these areas in more detail and divided the broad categories WW and F/I in several sub-sets. This was done for two reasons, both already mentioned in the methodology chapter, section 3.4.5, which will be reiterated here. The first reason was that, although statistical data from the inter-rater reliability tests showed no significant difference in the classification of these areas, subsequent discussion and reading highlighted the fact that there is a propensity for ambiguity when trying to classify errors into one of these two broad categories.

The second reason for more detailed analysis was that, given the results noted corpus use in conjunction with these errors, I felt a more precise analysis would yield further information as to exactly where the corpus was applied to errors and how effective it might be.

The issue of classification of errors remains a potentially problematic area, in that it is open to the subjectivism of the classifier, but below I have outlined as clearly as possible the new categories and a justification for them.

- Register errors: Lexical errors which were register related -the use of statements including subjective pronouns and statements which contained explicitly inappropriate constructions such as the use of idiomatic words.

  Example error: 'As far as I'm concerned…'

  Example correction: ' It could be argued that…'

- Emotive language: Lexical errors using language which, whilst correct in meaning, was too emotive or subjective in nature. Typical of this category was a lack of academic caution as evinced by a lack of hedging and the use of extreme adjectives or intensifiers.

  Example error: 'It is known to all that…'

  Example correction: 'It is particularly well known that…'

- Prepositions: Whilst this is part of speech and might be seen as distinct from the other categories, which, it might be argued, are largely conceptual, I included this as a category because research by O'Sullivan and Chambers (2006) suggested the corpus might be an effective tool in examining this area.

  Example error: '…can be seen *by* the following ways.'

  Example correction: '…can be seen *in* the following ways.'

- Word Form: Words which on reanalysis of the co-text were judged to be errors of POS (part of speech).

  Example error: '…being social unstable.'

  Example correction: '…being socially unstable.'

- Wrong Words: These were error produced a breakdown in sense due to being entirely inappropriate.

  Example error: '…fertilize their human resource.'

  Example correction: '…begin to fully utilize their human resource.'

- Contrary Meaning: Words used incorrectly which produced the effect of expressing an opposite or contrary relationship with the co-text from that intended by the learner.

  Example error: '…either the rich or poor would be satisfied.'

  Example correction: '…neither the rich or poor would be satisfied.'

- Synonyms: lexical items which, whilst conveying the meaning intended, did not fit with the context or co-text of the structure.

  Example error: ' …setting subsidiaries…'

  Example correction: '…establishing subsidiaries…'

Below, in figure 4.35, is a total breakdown of the elements as whole, regardless of group or original error designation, across the four draft cycles.

WRONG	SYN	CONTRA	REGISTER	EMOTIVE	PREP	FORM
11.15%	49.23%	5%	9.82%	4.42%	12.69%	7.69%

Figure 4.35: *Breakdown of lexical errors across the course – sub-divisions.*

Incorrect synonyms were by far the most prevalent type of error, accounting for about half of the total. The table in figure 4.36 gives a more detailed breakdown by group and original, broader error designation. Clearly, synonyms again predominate as the most frequent type of error, with only register in the CG F/I errors showing a higher number, and by less than 2% at that.

	WRONG (%)	**SYN (%)**	**CONTRA** (%)	**REGISTER** (%)	**EMOTIVE** (%)	**PREP** (%)	**FORM** (%)
F/I EG	6.78%	58.85%	1.69%	23.73%	16.95%	-	-
WW EG	14.67%	58%	6%	-	-	18%	9.33%
F/I CG		40.91%	2.27%	42.05%	14.77%	-	-
WW CG	14.34%	46.19%	6.28%	-	-	21.53%	11.66%

Figure 4.36: *Sub-errors by group and original category (%)*

It seems the use of a word which was near in meaning but incorrect in context was responsible for the majority of wrong word or formal / informal errors. Errors in prepositions provided the second largest group with complete errors in meaning closely following. The

greatest number of errors, therefore, are wrong word errors rather than merely being identified as register problems indicated by either pronouns or emotive language. This is interesting in the light of some of the information given in the interviews, which suggest that students were primarily utilising the synonym function in the BYU-BNC interface. This is discussed in greater depth in section 5.4 of chapter 5.

4.7.1 Correction Methods and Effectiveness on WW and F/I Error Elements.

The question of most interest is the extent to which the EG applied the corpus to these most persistent types of error and to what extent those corrections were effective. Figure 4.37 gives a breakdown of the types of error made and the methods of correction which were adopted.

	Wrong (%)	**Syn (%)**	**Contra (%)**	**Register (%)**	**Emotive (%)**	**Prep (%)**	**Form (%)**
Corpus corrections	15.39	43.59	10	0	20	44.44	35.71
Knowledge Corrections	61.54	37.61	50	85.70	60	50	28.57
Dictionary Corrections	7.59	5.98	30	0	10	0	28.57
Online Sources	0	3.42	0	7.15	0	0	0
Unattempted	15.48	9.40	10	7.15	10	5.56	7.15
TOTAL	100	100	100	100	100	100	100

Figure 4.37: *EG error types and correction methods (% of total corrections).*

The use of the corpus is most common in the correction of synonyms, being for that error type the highest ranking correction method. In errors related to form, the corpus was also

used more than other correction formats. For the rest of the error types, however, knowledge corrections were the most common type of correction. Again, the interviews and surveys seem to support this in that students identified synonyms as an area which the corpus helped with and specifically that the synonym function of the BYU-BNC interface was employed during corrections.

The table below (figure 4.33) shows the relative accuracy rates of the various correction methods when measured independently of each other.

	Wrong (%)	**Syn (%)**	**Contra (%)**	**Register (%)**	**Emotive (%)**	**Prep (%)**	**Form (%)**
Corpus corrections	25	92.16	0	N/A	50	75	100
Knowledge Corrections	81.25	79.54	100	50	66.67	77.78	50
Dictionary Corrections	100	100	66.67	N/A	0	N/A	100
Online Sources	0	75	N/A	100	N/A	N/A	N/A

Figure 4.38: *Independent accuracy rates of correction methods EG (%)*

Whilst dictionary corrections have a high rate of success, it should be remembered (see figure 4.21 in section 4.5 and table 4.38 above) that, within the EG, they accounted for proportionately few correction attempts and that the key correction methods still stand as corpus and knowledge corrections. Taken independently, knowledge corrections are consistently more effective than corpus corrections. However, in the most prolific error group, synonyms, corpus use achieves an effectiveness of slightly over 90% whilst in the same category knowledge corrections are slightly less than 80% effective. For prepositional

errors, knowledge corrections are slightly more effective than corpus corrections, by a margin of just less than 3%.

	Wrong (%)	**Syn (%)**	**Contra (%)**	**Register (%)**	**Emotive (%)**	**Prep (%)**	**Form (%)**
Corpus	6.25	51.09	N/A	N/A	20	46.15	45.46
Knowledge	81.25	38.04	71.43	85.71	80	53.85	18.18
Dictionary	12.5	7.61	28.57	N/A	N/A	N/A	36.36
Online	N/A	3.26	N/A	14.29	N/A	N/A	N/A

Figure 4.39: *Accuracy of correction methods EG (% errors as a whole).*

The figures above accuracy of correction methods as proportions of accurate corrections as a whole. The results follow a similar pattern to that of the usage table, with synonyms and form being the types of error where success is attributable mostly to the use of the corpus. In the case of synonyms, just over half, at 51.09%, of successes were made using the corpus, with knowledge corrections accounting for 38.04% of accurate revisions. In terms of form errors, the corpus accounts for over 45% of all successes. In all other error areas, knowledge corrections predominate in both frequency of usage and relative accuracy.

The consistent results for corpus use with synonym errors – most frequent method of correction, highest ratio of successes and proportionately accounting for more correction success – suggests that it is in the area of synonyms, and it seems synonyms only, that the corpus might be valued as a correction method over others.

4.7.2 CG Comparison

The EG results above, however, might be compared with those of the CG. Figure 4.40 illustrates usage of different correction methods.

	Wrong (%)	**Syn (%)**	**Contra (%)**	**Register (%)**	**Emotive (%)**	**Prep (%)**	**Form (%)**
Knowledge Corrections	59.38	30.94	62.5	62.16	76.92	72.92	73.08
Dictionary Corrections	18.75	59.71	25	2.7	0	18.75	23.08
Online Sources	0	0	0	0	0	0	0
Unattempted	21.87	9.35	12.5	35.14	23.08	8.33	3.84
Total	100	100	100	100	100	100	100

Figure 4.40: *CG error correction methods by error type (%)*

Note the key similarity between the EG and CG is in the area of synonyms and that, for these errors, a reference tool has been applied to corrections more frequently than recourse to knowledge. In the case of the EG, the corpus is used and in the case of the CG, the dictionary, it accounting for almost double the number of corrections compared to knowledge.

Figure 4.41 is a breakdown of the accuracy of each method taken independently of each other.

	Wrong (%)	Syn (%)	Contra (%)	Register (%)	Emotive (%)	Prep (%)	Form (%)
Knowledge Corrections	78.95	75.9	80	69.57	80	65.71	89.47
Dictionary Corrections	83.3	65.12	50	100	N/A	77.78	100
Online Sources	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Figure 4.41: *CG accuracy of correction types (%, independent)*

There is a different pattern of effectiveness to that produced by the EG when employing reference materials in the form of the dictionary. Wrong Words, form and register errors seemed to be tackled more effectively by the dictionary, but not synonyms. This again is interesting, because in the EG it was identified that synonyms were the problem area and it was synonyms which were tackled more frequently and more effectively by the corpus. The following is an illustration of overall successful corrections and to which correction method they can be attributed to within the control group.

	Wrong (%)	Syn (%)	Contra (%)	Register (%)	Emotive (%)	Prep (%)	Form (%)
Knowledge	75	69.23	80	100	94.12	76.67	73.91
Dictionary	25	30.77	20	N/A	5.88	23.33	26.09
Online	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Figure 4.42: *Accuracy of correction methods, CG (% of whole)*

Figure 4.42 shows a distinct difference between the relative effectiveness of dictionary usage and knowledge corrections. For every error type, knowledge corrections accounted for a greater proportion of the correction successes than any other method. Although the dictionary was shown in figure 4.40 to have accounted for the majority of synonym corrections, this usage did not translate into success, knowledge corrections accounting for almost 40% more of the successes than dictionary attempts.

In summary, whilst both groups tended to access a resource when encountering synonym errors, the comparative success rates differ greatly. In the EG, recourse to the corpus led to a higher success rate for this error type whereas, in the CG, corrections were more effectively applied to synonyms when knowledge was used.

4.7.3 Statistical Testing of Synonym Corrections

The following test (figure 4.43) looks for any statistical evidence to support the suggestion that the corpus was a more effective method of making a correction to a synonym than other correction methods, measured within the EG. The dependent variable is whether or not a synonym revisions was correct (1) or not (0). Independent variables used were gender, learner type, and the correction method. At this point, however, none of the results were statistically significant. This might be due to the reductive nature of the test; the number of cases to examine was reduced quite considerably and it is possible that the effectiveness of the test has been compromised by a lack of data. This will be discussed further in chapter 5 with reference to the limitations of the experiment and suggestions for improvements.

		B	S.E.	Wald	df	Sig.
Step 1[a]	Gender(1)	1.129	.841	1.801	1	.180
	Lrn_Typ			2.811	4	.590
	Activist / reflector	.027	1.575	.000	1	.987
	Reflector/pragmatist	1.570	1.356	1.340	1	.247
	Reflector	1.394	1.298	1.153	1	.283
	Reflector / theorist	1.641	1.694	.939	1	.333
	Correct_Mthd			4.313	3	.230
	Corpus	-19.708	14987.988	.000	1	.999
	Knowledge	-18.502	14987.988	.000	1	.999
	Dictionary	-20.632	14987.988	.000	1	.999
	Constant	15.321	14987.988	.000	1	.999

Figure 4.43: *Factors affecting successful synonym corrections (EG)*

Exactly the same test was applied to the CG in order to act as a comparison, (figure 4.44) but the results lacked any statistical significance. Again, this could be due to the large reduction in the amount of samples being examined when the item in question is only synonyms.

		B	S.E.	Wald	df	Sig.
Step 1[a]	Gender(1)	-1.162	.673	2.985	1	.084
	Lrn_Typ			4.396	3	.222
	Activist / reflector(1)	19.244	17966.580	.000	1	.999
	Reflector/ Pragmatist(2)	-1.113	1.118	.990	1	.320
	Reflector(3)	.564	.825	.468	1	.494
	Correct_Mthd(1)	.183	.451	.165	1	.684
	Constant	2.012	1.162	3.000	1	.083

Figure 4.44: *Factors affecting successful synonym corrections (CG)*

4.7.4 Summary

Lexical errors in both groups were largely a result specifically of synonym choice. Within the EG, it was the corpus which was used most frequently as the correction method for these type of errors. Additionally, it was with synonyms that the corpus was both most accurate compared with other errors corrected using it and, in comparison with the other correction methods, the use of the corpus was more effective. When the CG applied dictionaries to correct synonym based errors, a lower rate of success was found than in the comparative area within the EG. However, statistical analysis using binomial logistic regression does not, with this sample, reveal anything statistically significant about this finding.

The next section will examine the errors made at the baseline stage and the final stage in more detail. The aim here is to investigate whether or not the frequency of wrong word or informal errors, and their subcomponent parts, particularly synonyms, show any significant

reduction between the two phases. If there is a notable reduction in the instance of wrong word errors / formal / synonym based errors, this may simply indicate the effect of general learning. However, if this effect is more pronounced in the experimental group, given that these areas were targeted by the corpus more often than any other method, it may be a basis for suggesting that the use of the corpus had an influence on the language learnt over the course.

4.8 Longer Term Improvement

Research question RQ2 asked whether or not the learners who used corpora showed any quantitative sign of improvement in accuracy over the course. In section 4.2 of this chapter, I established that both the CG and EG made improvements in overall accuracy, but that there was no statistical evidence to support a greater improvement in the EG.

Research question RQ2.2 investigates what aspects of student writing improved and, in the case of the EG, if that might be attributable to corpus use. To investigate this, I compared the error features which were prevalent at the baseline and final stages. Previous sections have shown that whilst there was an improvement in accuracy across errors, some were more persistent than others and only certain types were revised with the aid of corpus data. The two tables below, figures 4.45 and 4.46 are statistical tests of the errors in each group, with the dependent variable being the cycle (baseline or final) . A significant result ($p.<.05$) equates to the relevant independent variable being a factor in error production at the final stage.

	B	S.E.	Wald	df	Sig.	Exp(B)
Learn_typ			2.214	3	.529	
Activist/reflector(1)	.191	.451	.180	1	.672	1.211
Reflector/pragmatist(2)	.652	.479	1.848	1	.174	1.919
Reflector(3)	.327	.335	.954	1	.329	1.387
Error			136.053	11	.000	
A(1)	.654	.630	1.077	1	.299	1.923
WW(2)	.802	.501	2.566	1	.109	2.231
F/I(3)	-.441	.510	.746	1	.388	.644
P/S(4)	3.065	.698	19.279	1	.000	21.436
G(5)	1.204	.527	5.224	1	.022	3.333
^(6)	1.823	.520	12.295	1	.000	6.190
Agr(7)	1.754	.598	8.610	1	.003	5.776
T(8)	2.806	.761	13.583	1	.000	16.542
WO(9)	2.958	1.132	6.829	1	.009	19.265
WF(10)	1.628	.541	9.064	1	.003	5.093
?(11)	3.557	.872	16.650	1	.000	35.057
Constant	-.114	.594	.037	1	.847	.892

Figure 4.45. *Factors affecting error production at baseline/final (CG)*

The results from the CG show several factors which are significant predictors of an error being made at the final stage, and these include several error types. In fact, apart from article errors (A), wrong (WW) and register errors (F/I) all errors have a statistical significance which suggests they were more likely to be a factor at the final stage. This is interesting,

because whilst the accuracy in terms of T-units was shown to have improved at the final stage, the information here suggests that students were more likely to make an error at the end point. This seems to indicate a contradiction in the results and this will be examined further in the discussion chapter under section 5.3.2.

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	**Learn_typ**			21.948	4	.000	
	Activist/reflector(1)	.417	.447	.872	1	.350	1.518
	Reflector/pragmatist(2)	1.803	.427	17.850	1	.000	6.066
	Reflector(3)	1.180	.408	8.366	1	.004	3.253
	Reflector/Theorist(4)	1.295	.570	5.167	1	.023	3.650
	Error			53.604	11	.000	
	A(1)	1.522	1.044	2.125	1	.145	4.583
	WW(2)	.338	.854	.156	1	.693	1.402
	F/I(3)	-.849	.860	.975	1	.324	.428
	P/S(4)	1.541	1.030	2.239	1	.135	4.668
	G(5)	1.297	1.005	1.666	1	.197	3.659
	^(6)	1.088	.874	1.550	1	.213	2.969
	Agr(7)	1.745	1.027	2.888	1	.089	5.724
	T(8)	1.100	.962	1.307	1	.253	3.005
	WO(9)	.165	.992	.027	1	.868	1.179
	WF(10)	.674	.881	.586	1	.444	1.963
	?(11)	.904	.996	.824	1	.364	2.470
	Constant	-1.042	1.039	1.006	1	.316	.353

Figure 4.46. *Factors affecting error production at baseline/final (EG).*

Figure 4.46 shows the same test applied to the data from the EG. This time, none of the error types appear to be significant predictors of whether an error occurs at the final stage. This result, neither positive or negative, might denote a lack of effect of the course, and the implications will be discussed under chapter 5, section 5.3.2.

What may be gleaned from the results above is that, when the groups were taken independently and all the results were examined, the CG seemed more likely to produce errors at the final stage and there seemed to be no significant effect with the EG . However, when the percentage data in figure 4.19 in section 4.4 is examined, particular errors appeared to be problematic. This data showed that at the baseline, both groups produced similar errors at similar rates. Most noticeably, Formal / Informal errors, wrong word errors and missing word errors occupied the top three ranks in both groups. F/I and WW errors were between two and three times more frequent than the next highest error, and in both groups accounted for more than half of the total errors.

At the final stage, the pattern of errors had changed. In the CG, missing word errors accounted for the greatest proportion of errors followed by wrong word errors. In the EG, this pattern was reversed, but the frequency of these types of error was virtually the same in the EG, with slightly under 21% WW errors and slightly over 20% ^ errors. F/I errors had dropped to between 6 and 8% of total errors for both groups, with the CG producing these less often than the EG. It seems then that whilst the proportion of WW errors reduced in both groups, this error category was still particularly prevalent. More importantly in terms of the analysis, it was this category of error to which corpus revisions had been applied most frequently.

The following test examines what was a predictor of a wrong word error being made at the baseline or final stage. The dependent variable, therefore, is cycle. The test was set up so that the baseline stage was labelled as ‘0’ and the final stage was labelled as ‘1’. A

significant result then indicates the independent variable being a predictor of a wrong word at the final stage, the independent variables being group and learner type.

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Learn_typ			8.303	6	.217	
	Activist/reflector(1)	.563	1.019	.305	1	.581	1.756
	Reflector/pragmatist(2)	.255	.779	.108	1	.743	1.291
	Reflector(3)	1.148	.697	2.713	1	.100	3.150
	Reflector / theorist (4)	.809	.954	.718	1	.397	2.245
	Pragmatist(5)	.505	.715	.499	1	.480	1.657
	Theorist(6)	-.205	.876	.055	1	.815	.815
	Group(1)	-1.169	.595	3.867	1	.049	.311
	Constant	1.207	1.101	1.203	1	.273	3.344

Figure 4.47: *Factors affecting WW error production at final stage*

There is only one independent variable which produces a significant effect on whether a wrong word error was produced at the final stage rather than the baseline stage, and this is 'group (1)', or the EG at p.049. However, the Beta figure is -1.169. This minus sign indicates that being in the EG meant a wrong word error was significantly less likely at the extended assignment stage. This is not conclusive evidence of corpus use aiding linguistic acquisition, but it does suggest some beneficial effect in that the error type which had been targeted most by learners in the EG with the corpus was less likely to occur at the final stage compared to the group which had no access to the corpus.

As in the previous section, WW errors were re-examined in their sub-types. Figure 4.48 shows the changes in the production of these errors at the baseline and final stages. As noted in the previous section, synonyms in particular created a problem during the draft/revision/feedback cycles and here they are interesting for the fact that the final stage of the experiment shows synonym errors accounting for a greater proportion of errors than at the baseline stage.

Group	**Cycle (%)**	**Syn (%)**	**Prep (%)**	**Wrong (%)**	**Contra (%)**	**WF (%)**	**Pron (%)**	**Caution (%)**	**? (%)**
CG	Baseline	32.61	7.97	5.8	1.45	0.73	47.82	3.62	
	Final	46.29	16	13.71	1.14	1.71	16	1.14	4.01
EG	Baseline	27	8.64	3.7	2.47	4.94	45.68	6.17	1.4
	Final	40.46	17.56	10.69	3.82	2.29	10.69	7.63	6.86

Figure 4.48: *Error subtypes baseline/final stage (%)*

The statistical analysis in figure 4.49 examines looks at the predictors in creating a synonym error at the baseline or final stages and includes all instances of synonym errors from all users in both groups. The dependent variable is the cycle, either baseline or final, with the independent variables being learner type and group.

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Group(1)	-2.580	.799	10.427	1	.001	.076
	Learn_typ			7.432	6	.283	
	Activist/reflector(1)	-.897	1.711	.275	1	.600	.408
	Reflector /pragmatist(2)	-.669	.916	.533	1	.465	.512
	Reflector(3)	.777	.815	.909	1	.340	2.176
	Reflector / theorist(4)	.197	1.036	.036	1	.849	1.218
	Pragmatist(5)	.611	.857	.509	1	.476	1.842
	Theorist(6)	-22.104	23205.422	.000	1	.999	.000
	Constant	3.613	1.373	6.921	1	.009	37.075

Figure 4.49: *Factors affecting synonym error production at final stage*

Again, there is only one statistically significant result. This is similar to the previous result. Group is a predictor of whether a synonym error is statistically likely to occur at the final stage, at p.001. Again, the minus Beta value, -2.580, means that being in the EG is less likely to result in an error of this type being made at the final stage.

4.8.1 Summary

The percentage data shows that between the baseline and final stages, the proportion of errors which were WW errors fell slightly, although this error type still represented one of the most problematic. This type of error, as established in section 4.5, was the one which was corrected by the corpus most frequently and accurately, although within the EG there was no

statistical evidence to suggest it was more effective than other methods. However, when the CG and EG production of WW errors at the baseline and final stages was examined, the results suggested that the EG, or the group which had used the corpus, was statistically less likely to produce this type of error than the CG. The implications of this result in terms of research question 2.2 will be discussed in the following chapter.

Section 4.4 showed that, within WW errors, synonym errors caused the greatest problem during the draft and revision stages. These were most often targeted with the corpus and it seemed successfully so. In this section, statistically again it seemed that the creation of synonym errors at the final stage was less likely within the group that had used the corpus throughout the course. However, when the percentages are looked at, the proportion of synonym errors had increased compared to the baseline stage. Again, this will be discussed at greater length during the next chapter, but the result implies that, whilst the EG may have been comparatively less likely to produce this type of error than the CG, the improvements in WW correction overall were due to another subset of that error.

4.9 Learner Perception and Evaluation

The statistical tables of answers to the Likert scales can be found in appendices 11.1 and 11.2. Below is a summary of the main points, followed by a summary of the answers given in the interviews, which can themselves be found in full in appendix 12. The results of the survey are reported in four sections. The first details the habits of the students in terms of error correction before and after the course. General attitudes of the learners towards error correction are then described. Following this, the use of the different reference methods are examined with the learners responding to questions on how easy or difficult it was to use either the dictionary or the corpus data, and why, Finally, I report on the learners' self-evaluations of the different correction methods.

4.9.1 Learning Habits of the Students

Responses to questions which dealt with the learning habits of the students were similar across both groups, though whilst just under 70% of the CG claimed they were used to checking their written work before they submitted anything for marking, only about 42% of the EG group were in the habit of doing this. However, upon receiving feedback 77% of the CG and 91% of the EG said that they regularly read the feedback that was given to them. 69% of the CG and 75% of the EG reported that they always then attempted to make revisions based on the feedback they had been given. This suggests that the mechanism employed in the experiment of draft/feedback/revision was something which the learners would have been familiar with in some format.

Responses about the same habits but after attending the course indicate that there was either a positive change or no real difference. 76% of the control group said that they always checked their work before submitting it during the course, all of them said they read the feedback given to them and 84% said they always then attempted to make revisions. In the EG, over 90% said they checked their work before submission, all of them claimed to read the feedback which was supplied and just under 92% said they went on to make revisions based on the feedback.

These results seem favourable in the sense that the vast majority of learners seemed to be engaged with the process during the course and were happy to take control of their own corrections. However, some caution must be exercised for two reasons. The first is that the students were explicitly being asked to engage in the feedback and revision cycle and to provide evidence of having done so in the form of revisions and the method of correction coding. Whilst it might be tempting, therefore, to see any improvements in these habits as pointing towards a more developed level of autonomy, they can only be realistically judged to have been a result of the mechanism of the experiment. Similarly, these positive results

may well have been biased by the effect of the researcher both being involved in the teaching process and the administration of the experiment and questionnaire. It is likely that, whilst these questionnaires were taken at the exit of the course and further contact with the researcher was not to be assumed, some learners wanted to provide an answer which they deemed as being the correct one in the eyes of the researcher, or, in other words, the most pleasing answer.

4.9.2 Learner Thoughts on Correction

On the whole, students of both groups evaluated the usefulness of making corrections positively. Time spent on making corrections to their errors was valued by the students. In the CG none of the students thought that corrections took up too much time. The EG group was slightly less certain, but still nearly 60% felt correcting work did not take too much time.

Most learners seemed to think that the process of feedback and the activity of correcting their individual errors helped to raise their awareness of the types of errors they were producing in general. In the EG, the response was 100% positive whilst the EG was in a majority of over 75%. With this in mind, the response to whether or not the correction process had helped the learners reduce their amount of errors is interesting. Whilst, as noted, the vast majority of responses were positive in terms of being able to recognise errors in the EG only 50% of learners answered in the affirmative when asked if the process had actually helped them correct their errors. Respondents were also far less confident in their personal perception of being able to correct errors in the CG, with the level of confidence dropping to just over 60%.

In both groups, over half the respondents thought that grammar had been helped by correcting work but the response was particularly favourable where vocabulary was concerned. Approximately 70% of the CG and 83% of the EG thought that correcting their

work had helped them learn vocabulary. This is perhaps not surprising as the majority of corrections which the students eventually engaged in were of a lexical nature.

In brief, whilst the learners of both groups seem to have valued the correction process and had the impression that it helped them recognise their errors, and even thought it had helped them learn aspects of the language, they were far less confident about the actual effect of the process on any subsequent elimination of errors in their compositions.

4.9.3 Use of Correction Methods: CG

Firstly, the results correspond to what was ascertained from observation of the corrections and through questioning the students, namely that the students were using bilingual electronic dictionaries rather than monolingual paper dictionaries. 92% of students answered affirmatively that they preferred to use an electronic dictionary, with the other 8% providing a neutral answer. This was corroborated by the answers to the companion question about the use of paper dictionaries, with nobody giving a positive answer about the use of paper dictionaries. Evidence from the interviews provides further evidence that the preferred tool, when knowledge corrections were not being made, was an electronic dictionary.

Learners in the CG did not, on the whole, feel they experienced any difficulty using electronic dictionaries to correct their errors in terms of time and effort. However, whilst all the students used electronic dictionaries at draft and correction phases, usage of the dictionaries was not seen as wholly straightforward process, despite all students having invested in their own dictionaries and having used them before; nearly half of learners claimed a degree of difficulty due to dictionaries taking too much time and effort to use. Over 90% of students also claimed that dictionaries could be both too difficult to understand and roughly the same number experienced difficulties in making the actual searches.

Nearly half the respondents said they had experienced difficulties in analysing the dictionary entries once they had been found. Whilst none of the learners experienced difficulties due to limited information being supplied by dictionaries, the group was split when asked if there was too much information. It is unclear in exactly what way this manifested itself, though about 40% of responses suggested there was some difficulty encountered with the language used in the examples and definitions given in the dictionaries. An overload of available information is also mentioned by one of the interview participants and the problem seems to result from the display of lots of alternative words being listed after electronic dictionary searches.

A majority of students used dictionaries as an aid when problems were encountered and also claimed that when searching for information in dictionaries they found the answer that they needed. Over 60% of respondents claimed that they used dictionaries when they needed help with language.

One of the effects of using the dictionaries was that students seemed to be moving between the target language and their mother tongue when trying to find the correct vocabulary. 70% of learners said that when they used a dictionary to find an answer, they were translating directly in and out of their own language.

4.9.4 Use of Correction Methods: EG

75% of the group said the training course helped to learn the technique and a similar amount noted that they felt relatively confident in using the corpus. This is interesting in light of some of the contradictory responses below, which suggest an often negative attitude towards ease of use.

In terms of time and effort spent to make corrections with the corpus, the response was negative. That is, two thirds of learners felt using the corpus took too much time and

effort. This is interesting, because although the group has responded positively in terms of the correction process as a whole not being too taxing on time, this answer suggests that use of the corpus within this process is too taxing.

A minority of students, or 25%, cited the concordance lines being difficult to use due to the vocabulary contained within them. This does not mean, however, that the remaining 75% found concordance lines easy to understand. It might equally suggest that students were not experiencing difficulty because they were not using the concordance lines at all. Indeed, interview evidence suggests that other facets of the interface were more popular with the students. In fact, subsequent evidence form the survey suggests that concordance lines were problematic. 50% of the students said that they found the concordance lines themselves difficult to analyse and a similar number cited that concordances were a problem because there were too many of them.

Apart from the concordances themselves, less than 17% of learners were able to affirm that the search technique did not provide them with any difficulties. 50% of the students, on the other hand, answered in the affirmative to this, suggesting that on the whole, the search technique did cause problems. Unlike the CG, who reported that they were largely able to find the information they were looking for when making searches, half the EG recorded that they did not find the information that they needed when using the corpus. In light of the other answers and the interviews, this could be an issue of time and effort. Certainly, when asked if they did use the corpus independently of the correction process, 60% replied with a definite negative.

In summary the CG acknowledged that there were problems searching the dictionaries and in understanding entries and definitions. However, most of the learners thought they were able to find the information they were looking for. The EG, cited several difficulties in the use of corpus, including understanding the information given within the concordances,

difficulties performing the search the technique itself , corpus use just taking too much time and effort to use and problems finding answers to their queries meaning that it was not employed in the draft or independent writing process.

4.9.5 Learner Evaluation of Correction Methods

When it comes to evaluating the use of dictionaries, the control group has varied answers as to their usefulness. The majority felt that dictionaries helped in understanding the meaning of vocabulary, at 61%. However, the opposite was true when it came to actual usage of vocabulary, with under 39% responding positively. Again, this might be attributed to the use of electronic dictionaries, which bring up a limited amount of relevant examples of contexts for language use. The same reason might explain the fact that only about 30% of the group found dictionaries useful in the learning of phrases. Similarly, only 23% found dictionaries useful for learning grammar. This might be expected given the general nature of dictionary use and the fact that most of the errors encountered during the study were those of a lexical nature. 76.92% did find that dictionaries helped them correct their work (this is borne out in the corrections data) and half the group cited dictionary use helpful for writing in general.

In terms of learning the meaning and usage of vocabulary, the EG gave a response which was the inverse of the control group, in that only 42% could affirm that the corpus was helpful for learning the meaning of vocabulary, whereas the response was overwhelmingly positive when asked if the corpus was helpful for ascertaining the use of vocabulary, with over 91% responding positively. Again, the corpus was not really felt to aid in learning grammar and this could largely be due to the nature of the corrections and searches made by the students, which were, on the whole, lexical. Neither the use of dictionaries or the corpus was judged to have had a positive effect on reading.

Learners did not judge the corpus to have had a positive effect upon their independent learning, with only 33% of the students giving a positive response to this question. 46% of the CG felt that using the dictionaries had helped in developing their autonomy.

Interestingly, some of the answers given by the students seem to contradict their actual usage of the corpus and even their earlier answers. It might be that a wish to give a positive response affected the students here, as they knew some of the basic aims of the experiment. Thus, while the data from the corrections illustrates both a relatively low use of the corpus as a whole and definite decline in usage across the course, just over half of the respondents said that the corpus had a positive effect upon their writing. This is also interesting in light of the fact that other survey questions suggest that the students found the corpus difficult to use, particular in the interpretation of concordances, that it took too much time and effort to use and that they did not use it independently of the correction process.

A further question put to the experimental group was interesting. Only half the group suggested that the corpus was more helpful than a dictionary for English writing. It is possible that the lack of dictionary usage within the experimental group corrections is accounted for to some extent by dictionary use in the draft phase. Indeed, as will be seen in the interview summaries, students appeared to favour dictionary usage and seem to have used the corpus as a type of last resort.

4.9.6 Summary of Learner Evaluation

Evaluation of dictionary use by the CG was largely positive, with some evidence that it was valuable in both the drafting and correction phases. Students felt that dictionaries helped learn word meaning rather than word usage and dictionaries were employed independently of the revision process in the draft stages. On the other hand, the EG, whilst finding the corpus useful for learning the usage, rather than the meaning of vocabulary, the response was much

more varied and at times negative. Generally, learners did not perceive the corpus as being helpful to their development as autonomous writers and were divided as to whether usage actually helped with their writing. Subjects in this group suggested that they might actually still value the dictionary as a more effective and helpful tool in terms of their writing performance.

4.10 Summaries of Interviews with the EG and the CG

This section reports on the interviews conducted with volunteers from both the CG and the EG once the course had finished. Five learners from the EG were involved whilst only two learners from the CG could participate.

Subject EG4 felt that her writing had improved over the course, particularly in terms of knowledge of academic style and knowing what was formal and informal, but felt this had improved as a result of classes, not just the correction process.

In terms of correction methods, she said she preferred to use the dictionary, although evidence of dictionary corrections is not outstanding in her work or the EG as a whole. This was because the dictionary was familiar, and correction with the corpus meant opening up the program and was something of an inconvenience compared to other methods.

Wrong words were identified as errors which were particularly difficult to correct, in general terms, not just with the corpus. It was in this area that the corpus seems to be mentioned as most useful, with prepositions, collocates and synonyms being the features of language it was most useful for dealing with.

In terms of autonomy, she felt the corpus had had little effect. Firstly, this was because use of the corpus did not leave the subject with a feeling of confidence that the correction was the right one, just that it was an alternative, and the subject wanted to have this error corrected by the tutor or looked at by other students. This might suggest that the

concordances were not being used and only the word list was being produced as a tool for reference. In the production of drafts and the extended assignment, the corpus was not used at all; use only occurred when directed to revise drafts. Though the student thought she would use the corpus in the future, this was with the caveat of having enough time to use it and enough time to become more familiar with it.

Subject EG5 felt that her progress in grammar across the course had been most noticeable. This she attributed to the teaching and the influence of her peers and corrections made on homework and in class. There is the suggestion that this learner saw the usefulness in the corpus over the dictionary and used it as an alternative, and again the ability to find synonyms and, in this case, observe correct usage in concordance lines was one of the features which was a benefit compared to dictionary use. This is largely due to the subject's technique of translating a words directly out of her L1 into the target language, resulting in a wide range of options in dictionary output which were not contextualised. The student expressed a wish to use the corpus in the future, particularly in terms of getting register correct and being more exact. However, when asked if she had used the corpus in the composition of her work, the answer was no.

Having felt that her grammar improved particularly, it is perhaps telling that EG3 explains that the corpus was seldom used for grammatical reference. Rather, it was used for changing words from formal to informal and for looking for synonyms of wrong words. A problem with usage was found, and this, as in studies examined in the literature review, was the amount of information available to the learner. The learner implied, in fact, that the act of having to click on words and identify how they are used by looking at the concordances was a weakness in itself, suggesting that an improvement in the future would be a reduction in the amount of data presented. This student thought that a more precise list

of relevant tokens ready for use, without having to examine context and co-text, would be desirable. This suggests interaction with the corpus was not really fully achieved.

Again, learner EG3 did not find the corpus a useful tool when generating texts, and an electronic dictionary was preferred which translated from the L1. The corpus was cited as being too complicated to use, particularly during the creative process when ideas needed to be taken down, the student thought, at speed. The corpus, she claimed, would only be useful when the errors had already been identified for the learner.

EG2 identified the corpus as being useful for checking the common usage of the word in particular genres, using the chart function to examine words for formality and register, suggesting it was being used for FI corrections. However, in terms of word usage the corpus was not deemed to be useful, despite specific reference to examining the concordance lines. Only the first one or two concordance lines were looked at because the number of the rest of them seemed to be overwhelming. It was felt that working through concordance lines just took too long and the student preferred the use of electronic dictionaries, again, particularly when composing drafts. In the future, the student cited the potential to be able to utilise the corpus to get a variety of words as a possible reason for continued usage.

EG11's computer broke towards the end of the course, and this is one particular example of independent corpus use being vulnerable to failures in technology; after the computer had broken, the learner was restricted to using it in the public labs of the university. Another interesting situation was in the fact that some instances of language use were marked as wrong by the tutor, but student investigation into their use resulted in a positive finding in the corpus. This led to the dilemma of whether to use it or not, or whether it was actually correct. The student cited the use of the corpus as being useful for the correction of register and wrong words and mentioned prepositions specifically. The student did not use the corpus

in the composition of work, though predicted that they would use it in the future and that it might be more helpful in accurate understanding when translating texts, but did not go on to suggest the mechanism by which this would be done.

Subject CG6, from the control group, cited vocabulary as the key problem in writing and correcting. This was mainly due to wrong words, or rather not being able to find the right words for the right contexts, and register, or not knowing what was academic or not. Dictionaries were helpful up to a point but occasionally too complex and she felt that electronic dictionaries, though faster, did not offer the accuracy of the paper variety. The dictionary, she thought, is not as helpful with wrong words as it could be, and she still struggled with these.

Learner CG7 was ambivalent as a whole about the reasons for her progress. She felt that the electronic dictionary was generally helpful in finding wrong words, and that the examples were sufficient to learn from. She did say that paper dictionaries were better but that she continued to use only an electronic dictionary.

4.11 Conclusion

The data shows a number of findings. The first is that there was a learning outcome, in terms of accuracy measured in EFT, over the length of the course. This, however, does not seem to be attributable to whether a learner was in the CG or EG. The second finding is that certain types of error, particularly wrong word (WW) errors, were both more persistent across drafts and also more resistant to effective correction.

In terms of corpus use within the EG, factors in its consultation appear to have been learner type, the cycle stage (with use progressively decreasing) and the type of error being targeted, with wrong word and register error attempts being most prolific. The evidence suggests that corpus use was effective in these areas, but that it was virtually limited to just

these areas and accounted for a small portion of contributions towards accurate revisions on the whole.

Corpus use may have been marginally more effective with the error types mentioned above than dictionary use in the CG, but the evidence is equivocal as to whether or not corrections during the revision process contributed to any significant ability to avoid those types of error subsequently.

Learner attitude towards the use of the corpus was ambivalent. Whilst members of the EG seemed to recognise some theoretical benefits of having access to a corpus, their response to its practical effectiveness is less than favourable.

In the following chapter, the points which have been raised through reporting the results will be discussed in relation to the research already examined in the literature review. This reveals implications for the use of the BYU-BNC within the pre-sessional classroom and the chapter identifies a number of areas for continued research.

5

Discussion

5.1 Introduction

In this chapter, I discuss the results presented in chapter 4 with reference to the research questions and examine the implications of these findings, with a particular focus on the context of pre-sessional courses. I follow this with an examination of the limitations of the experiment and suggestions for continuing areas of research.

The following three sections address the research questions directly. Section 5.2 focuses on the contribution both of the course and of the corpus to error corrections as outlined in RQ1, arguing that there was an observable benefit to using the corpus to correct some error types, but that this effectiveness was undermined by a number of factors which limited the use of the corpus as a reference tool.

Section 5.3 discusses RQ2 and the contribution of the corpus towards general accuracy. I note that the draft/feedback/revision process seemed to have a benefit on the accuracy of writing during the course and that there was a measurable learning effect between the baseline and final stages. However, I also note that it is unclear as to whether or not corpus use contributed towards this rise in accuracy or as to whether or not corpus use encouraged long term language learning.

Section 5.4 section examines RQ3. I find that learners were aware that the use of the corpus as a reference tool seemed more applicable to lexical areas and this conforms to the patterns of use suggested by the quantitative data results discussed in sections 5.2. I also find that learner evaluation of the corpus corresponds to some of the more negative attitudes recorded in the studies already mentioned in chapter 2.

In section 5.5, I examine the results of the experiment in the wider contextual framework of autonomy within EAP. Section 5.6 assesses the experiment itself, discussing the limitations and their effects on the results and findings. I go on in section 5.7 to propose avenues of further research which might benefit the understanding of the effects of corpus use within EAP.

5.2 Treatment of errors

RQ1.1 is concerned with whether or not learners are able to correct their errors using corpus data. This over-arching question is broken down into two more specific queries. RQ1.2 asks if particular circumstances dictate the use of corpus data over other methods such as dictionary use. RQ1.3 expands on RQ1.2 by examining whether corpus data enables a more effective treatment of errors than other methods.

5.2.1 Error correction

The first point to make here is that learners in both groups showed that they were capable of revising their errors accurately. The tables in sections 4.3 and 4.4 of chapter 4 show that at each revision stage, the groups performed positively on the majority of corrections. At stage 1, the CG (Control Group) mean rate of successful corrections, taking into account all error types, was over 71% and the EG (Experimental Group) revised over 81% of errors correctly. In cycle 2, the figures were 78.11% and 74.38% respectively and again in cycle 3 75.13% and 71.2%. At the final cycle, learners in the CG produced 62.91% of revisions accurately and the EG 87.5%.

There are a number of interpretations of the results in sections 4.3.and 4.4. The first is that a majority of the errors made could be corrected successfully by the learners, regardless of the group they were in. The proportion of errors which resulted in no attempt being made

at correction was below 17% in both groups at all stages. This suggests that the draft/revision/feedback cycle was having a beneficial effect and justifies its adoption as the method used in my research. That learners were responding to the feedback suggestions over 80% of the time supports the views of Yates and Kennel (2002) and Chandler (2003) that producing error feedback for learners is an important part of writing instruction. It also corroborates the assertions that, when learners have their errors pointed out to them, they can make successful corrections (Gass, 1983 and Makino, 1993).

The type of feedback employed in the experiment is also validated to an extent. I cannot show that employing other types of feedback, such as the direct method (Ferris, 2002 and Sheen, 2007) or indirect feedback, (Ferris and Roberts, 2001) would have produced more or less successful rates of correction or subsequent awareness of the language, but the accuracy of correction figures achieved across the course using the meta-linguistic correction code are in keeping with the positive results found by Ferris (2006); over the course of her experiment a mean of just under 80% of corrections were judged to be successful. Taken as a whole, the mean successful correction rate in my experiment was slightly over 75%. Additionally, this method of correction did not seem to suffer from the potential negative effects which Bitchener (2008, p. 108) warned of. In cataloguing a wide range of errors, he suggested that there was the potential to confuse and discourage the student. The vast majority of corrections were attempted by learners, and fruitfully, it does not seem that pointing out numerous error types constituted a negative affective factor at all. In fact, Bruton (2009, p. 139) claimed that learners must be allowed to make unlimited errors naturally within their writing. Having these personal error types pointed out to them would mean that learners are being addressed on an individual basis. He saw this as an essential factor in ensuring that error correction exercises resulted in a contribution towards the individual learners' writing as a whole as opposed to becoming merely a vehicle for

addressing a discrete language point identified by the practitioner. Certainly, had I focused on only one or two particular error types, there would have been areas of the learners' writing which would not have been addressed as the type and frequency of errors fluctuated throughout the experiment.

I noted in chapter 2 that EAP courses have a tendency to emphasis the macro-levels of writing, such as rhetorical function, over linguistic elements (Archibald, 2001; Hartshorn et al 2010). Of course, language input is often present in the form of the structures which contribute towards the construction of genre and rhetorical function, but I have previously pointed to evidence from Basturkmen and Lewis (2002) and Banarjee and Wall (2006) which implies that language related elements may go unaddressed on courses because an improvement in this area is either thought of as unrealistic or not necessarily a priority, as found in Cho (2003). However, a further interpretation which might be drawn from my experiment is that there may be a place within the short, six-week pre-sessional course for the incorporation of linguistic input. Providing an arena for linguistic focus through the draft/feedback/revision cycle meant that specific aspects of the language related to a particular learner could be addressed without infringing on the time needed for other types of input during the lessons. Employing this method as part of a pre-sessional course seems to have some potential as a way of addressing the lack of emphasis on accuracy and written proficiency which is bemoaned by Turner (2004) as a possible cause of problems regarding content and expression later in academic life. If so, then error correction has a place in contributing towards the goals of EAP courses in general in equipping learners with the necessary tools for further academic study (Alexander et al, 2008).

It can be seen then that, in the case of this pre-sessional course, correcting errors proved to be a practical exercise which produced positive results in the form of accurate revisions of drafts. However, whether the EG group was more successful than the CG is a

more problematic area, as is whether or not the learners were able to correct their work using the corpus.

If one looks even at the overall data for correction methods supplied in chapter 4, section 4.5 (figure 4.16), it becomes immediately obvious that the corpus as a correction tool had a limited effect on successful error correction. Taken as a whole, corpus corrections accounted for less than 17% of all corrections in the EG. This is compared to over 70% of corrections made with learner knowledge. Corrections made with the corpus account for only slightly over 18% of accurate corrections, compared again to 75% made with learner knowledge. The reasons for this limited use will be discussed in greater detail below. The acknowledgement of a lack of use must be balanced with the evidence that when it was employed by learners to correct errors, the corpus was a successful instrument. Corpus corrections, when considered independently of corrections as whole, were never less than 80% accurate. The mean rate of accuracy attained through corpus corrections was, in fact, over 88%. Knowledge corrections, on the whole the preferred method, had a lower mean accuracy, of approximately 82%.

In its basest form then, the answer to whether or not learners were able to correct errors using the corpus is yes. That is, when errors were corrected with the corpus, nearly nine times out of ten they were accurate. As in previous research (Watson-Todd, 2001; Chambers and O'Sullivan, 2004; Gaskell and Cobb, 2004; O'Sullivan and Chambers, 2006; Gilmore, 2009;), all of which reported accuracy rates of revisions between 60% and 80%, a tangible benefit of using a corpus as a reference tool has been identified. However, unlike these studies, my thesis is equally concerned with the figures of 17% - the proportion of all errors treated with the corpus - and 18% - the proportion of all successful error corrections attributable to corpus reference. In short, though the corpus may have been a useful tool when it was used for revisions, the evidence I have collected shows that it was used for less

than a fifth of all the errors examined. Unlike these previous studies, this thesis shows where the corpus was being used. It should be remembered from the literature review that not all previous studies had employed a mechanism which attempted to differentiate between corrections which were really made with the corpus as opposed to those which were made by a user who had access to a corpus, but may have made the correction without its use, and what the effectiveness of differing methods was. It is this information that allows RQ1.2 and RQ1.3 to be examined below.

5.2.2 Corpus Use

RQ1.2 examines whether the corpus was chosen over other correction formats and whether there was a reason for selecting it over another correction method. The results section suggests there was a reason for using the corpus at particular points. To see corpus use in the context of the course as a whole, a review of the types of error which were being made at the draft stage and, more importantly, which were proving a problem at the revision stage for the EG, is needed.

WW (wrong word) errors were the most persistent errors in the draft stages for the EG. Apart from the initial, pre-course task, where F/I (formal / informal) errors ranked highest, WW errors composed the greatest proportion of error types at every draft stage. They accounted for between 20% and 35% of errors across the different stages, or slightly under 30% of errors overall. This error type accounted for over 60% of all instances of the corpus being used. Whilst the error type did not prove to be a statistically significant predictor of corpus use, the logistic regression test that was used (figure 4.21 in section 4.5.2), shows in the beta column that, of all error types which were potentially a predictor of corpus use, only WW and F/I errors were more likely to be corrected using the corpus than another method (in the case of the EG, due to the rarity of instances of dictionary and online sources for

corrections, this other method was almost certainly knowledge). Other areas where the students attempted to use the corpus as an aid to correction were with article errors, grammar errors, missing words and a loss of sense. However, these accounted for only slightly over 15% of instances of corpus use.

In the more detailed analysis of WW and F/I errors in section 4.7 of chapter 4, the most common type of sub-error was found to be 'synonym', or words which were close in intended meaning but judged to be incorrect in context. It was this type of error and errors which were prepositional in nature which were targeted with the corpus more often than any other. Prepositional errors are also recorded by Chambers and O'Sullivan (2004) as areas which were dealt with effectively in their study of L2 learners of French, although they noted that corpus use was far greater for grammatical errors of gender and agreement.

There were two other indicators of corpus use. The first was learner type, and this was found to be a significant indicator of whether or not the corpus was employed. Theorist and reflector/ pragmatists used the corpus data more often than other learner types within the EG and this will be discussed further in section 5.4 with regard to research question RQ3.

The other indicator seemed to be the cycle, though again this was not borne out statistically. However, it can clearly be seen from figure 4.19 in chapter 4, section 4.5.2 that use of the corpus dropped as the course continued. Members of the EG using the corpus dropped from 100% in the first cycle to 30% by the end of the course. Instances of corpus use also dropped from up to seven queries per learner per text at the beginning of the course to a maximum of three at the end.

There are several possible reasons for this lack of corpus use. The first, with regard to the type of error, might first be explained with reference to 'treatable' and untreatable' errors, as defined by Ferris (2006, p. 96). Treatable errors consist of those which involve language which is rule governed and can be corrected by application of those rules in the

form of grammatical knowledge. Untreatable errors are those to which a set of obvious rules cannot be applied and are largely of a lexical nature. The wrong word errors identified in the learner texts belong to this category. Whilst the majority of other errors were addressed with knowledge, wrong word errors, and to a certain extent errors of register, were tackled with corpus data. 60% of corpus corrections targeted wrong word errors and 20% targeted formal/informal errors. It is possible that the corpus allowed for the treatment of these types of error because it may have aided learners in top-down processing. Flowerdew (2009) has suggested that the examination of concordances, because they offer examples of lexical features within a context, may help learners to develop knowledge of language use which cannot necessarily be taught as rule governed and which they may not have had access to before. However, as will be shown below in section 5.7, she stresses that the success of this may largely be dependent upon the approaches the learners take towards the information provided; reading the information in a vertical fashion, without examining the concordances horizontally, may not provide as much benefit because the learners may remain without an appreciation of the context and co-text of items being examined.

A second reason for the lack of corpus use seems to be that, for a majority of the errors, a reference source was not always deemed necessary by the learner at the point when they were making the revisions. This type of effect has been identified by Lee (1997, p. 471) who suggests that it is not the correction of errors which actually causes difficulties. Rather, it is the recognition of the error that is the key issue. Once an error has been brought to the learner's attention, it may be that, as with most of the errors encountered in this study, the error can be corrected using knowledge the learner already has. In these cases, as far as the learner is concerned, no corpus consultation is necessary.

Different types of error and perhaps awareness raising seem to have been factors which influenced whether or not the corpus was employed. However, initially, it was

employed by all users. This may have been due to the enthusiasm for a potentially useful new learning tool. Jarvis' (2004) research noted that nearly all foreign EAP students expected to be utilising computers as part of independent study and it may well have been that the introduction of the corpus through the training session at the start of the term proved initially motivating. This is quite likely given that the use of the corpus was directly linked in with not only academic work in general but had been presented as something which may help learners address aspects of their own individual work. Jarvis (2004), Charles (2012) and Gaskell and Cobb (2004) suggest that when the tool or methodology being introduced is tied directly to the work of the individual students, motivation for its use is likely to be high. The key question is therefore why there was a subsequent decline in use, particularly when, as seen below, the accuracy rates of corrections actually improved over time.

There is firstly the possibility that as the course progressed, the users became more proficient in the language they were using because they had learnt from the various iterations and therefor did not need to use the corpus. The evidence from the essays, however, seems to be counter to this and leads me to reject this interpretation; the types of error which were being addressed with the corpus, albeit successfully at the revision stage, remained issues in subsequent drafts and eventually at the final cycle of the experiment.

A second explanation would be that, as described above, the simple fact of being made aware of their errors meant that they felt they were able to correct their errors without the need for the corpus. This is more plausible as the beneficial effects of just noticing the incorrect language have been noted. However, instances of corpus use were, as noted above, linked with wrong word errors. These remained prevalent in drafts and were marked as such in the feedback, yet corpus use still fell. It might have been that learners saw the persistence of these error types in recurring drafts and revisions as evidence that, even using the corpus, they would remain largely untreatable.

I would suggest that a third possibility is more likely, and that is difficulty of use. Section 5.4 will go on to talk in more detail about the implication of learner perceptions about the corpus, but it is worth here reiterating the problems of using corpus data which other studies have encountered. Thurstun and Candlin (1998), Yoon and Hirvela (2004) Frankenberg-Garcia (2005), Lee and Swales (2006), Dudeney and Hockly (2007) and Sun (2007), have all pointed to the harmful effects of providing the students with a large amount of linguistic data to investigate; it is interpreted by the students as too much information and the increased cognitive load involved in analysing corpus data is a reason for avoiding its use altogether. Given the responses supplied by the student questionnaire, I am inclined to think that this is largely responsible for the drop in use.

Having noted in what contexts the corpus was used and having suggested why it was used in these contexts, the effectiveness of the corpus will next be examined. This is in relation to different error types and relative to other methods.

5.2.3 Corpus Effectiveness

This section addresses RQ1.3, which examines in what areas the corpus was effective as a correction tool, and whether or not it was more effective than either corrections made with learner knowledge or with other reference sources, of which the main type was the dictionary.

As already noted, only WW and F/I errors were corrected by the EG with any frequency using the corpus. At the first draft stage, corpus correction accounted for over 52% of accurate WW revisions, compared with 32% of correct knowledge revisions. The proportion of successful corrections being made with the corpus fell over the next three cycles, from 30%, to 26% and then back up slightly to 29%. At the same time, wrong word corrections made accurately with knowledge grew to 52%, dropping slightly in cycle 3 to 47% and then reaching 64% by the end of the revisions cycles. The reduction in the

proportion of successful corrections accounted for by the corpus follows the general trend of deteriorating corpus use towards the end of the course.

However, when measured independently of each other, the results provide a slightly different story. Corpus corrections were accurate 82% to 100% of the time. Corrections made with knowledge had a wider range, with 69% to just under 95% accurate. Mean accuracy of corpus corrections was about 5% higher than knowledge corrections, suggesting that on a like for like basis, the corpus was a slightly more effective method of correction. F/I corrections suggested a similar pattern, with corpus corrections gradually representing fewer of the accurate corrections overall, but achieving a mean accuracy rate of 87% compared to 65% for the knowledge corrections. Again, the percentages suggest that, independent of other methods, the corpus is a more effective reference tool for corrections. Nevertheless, when considered as part of the correction process as whole, F/I corrections accounted for only a small proportion of accurate revisions.

As mentioned in section 4.4 of Chapter 4, F/I errors had a slightly different pattern of occurrence to other errors, and WW errors in particular, in that they decreased quite sharply after the initial cycle. This is perhaps the one area that may be attributable to class and materials input from the course. Learners are shown quite early on the salient features of academic register (in terms of the pre-sessional; this is quite prescriptive and, when considered in the light of Biber's 2006 corpus based study of university registers, rather over-simplified), and how to avoid making basic errors. These largely consist of pronoun use (as in 'In my opinion…') and the misuse of extreme adjectives and intensifiers ('of huge concern'). These aspects of language are largely frowned upon on pre-sessional courses as they are deemed to indicate a lack of objectivity. Input in the form of Bailey (2011) and the course text (Corballis and Jennings, 2009) in the first week may have served to cut down this type of error rather quickly, rather than the error/feedback/revision cycle.

The statistical testing of corpus effectiveness in section 4.5.3.1, figure 4.27, did not provide a result of any significance which could confirm whether or not particular types of accurate error revision could be attributed to the use of the corpus. However, the beta column results suggested that only article errors, wrong word errors and formal/informal errors had any likelihood, significant or not, of being corrected successfully through the use of the corpus. All the other errors had a minus sign next to them in the beta column, indicating that corpus use was less likely to be a factor in their successful correction.

Once the errors were broken down into sub-sections, it seemed that the corpus might be a more effective tool in correcting synonym errors than simply relying on knowledge, with accuracy levels being approximately 92% and 79% respectively, when the methods were judged independently of each other. In terms of the proportion of accurate synonym revisions as a whole, the corpus derived corrections accounted for over half the amount, with knowledge corrections accounting for 38% of correct revisions. This suggests that in terms of synonyms, corpus corrections may have been the most successful correction method. The statistical test, however, did not provide a significant measure of whether any one particular variable was a factor in accurately correcting a synonym within the EG.

In one sense, then, corrections made using the corpus to both WW and F/I errors, and more specifically to synonym type errors, were more effective than using other methods of correction. When measured independently of other correction methods, the corpus consistently produced more accurate results. In the case of these particular error types, there is the additional data about the effectiveness of corpus corrections measured as a percentage of accurate revisions overall; a higher percentage of correct revisions could be attributed to corpus use than any other revision type. However, this has to be balanced with the fact that the area of effectiveness was restricted to a very low number of error types, that the corpus

was only up to about 5% more effective than knowledge corrections and there was no statistical evidence to suggest that the corpus was a more effective tool.

I examined the same type of information from the CG in order to see if a group which did not have access to the corpus had accuracy results which were any higher or lower. F/I corrections by knowledge had a mean accuracy rate of 52.95% whilst dictionary corrections had an accuracy of 62.5% when measured independently. In terms of accurate WW corrections, knowledge corrections were 71% accurate whilst dictionary corrections were 76% accurate. When looking at synonym type errors, again there was no significant statistical result to suggest that the dictionary was more effective than knowledge corrections. The CG had an overall lower degree of accuracy for these error types, and the dictionary seemed to deliver a lower level of accurate corrections than the corpus on WW errors in particular.

5.2.4 Summary

Research question RQ1 asked if learners were able to treat errors using the corpus. The answer to this is yes, at least through the mechanism which was used in the experiment. The draft/feedback/revision cycle allowed all learners involved in the experiment to target their individual errors, regardless of the particular aims of the pre-sessional syllabus. Within the limited amount of time available, both groups were able to show improvements in accuracy at the revisions stages. This did not necessarily translate into an improvement in accuracy on the next draft, and this will be discussed in the next section.

RQ1.2 asked if the corpus was used rather than other types of correction method. The answer again is yes, and this was dependent upon learner type, cycle and error type. The corpus was used on an extremely limited proportion of the total number of errors and to a large extent was confined to the correction first of wrong word errors and then formal informal errors. The former error type in particular could be regarded as 'untreatable' errors

and the comparative success of the corpus with wrong words compared to both knowledge corrections in the EG and corrections made in the CG suggest that the corpus is a more effective tool in dealing with these types of error at the revision stage.

However, this success is partially undermined by a number of factors. The first is that despite the percentage data from the samples, there was no statistically significant result to suggest the corpus was a predictor of accurate revision. The second, which is perhaps more important, is that correct revisions using the corpus accounted for less than one fifth of all the correct revisions made by the EG group. Partly, this was a result of the learners only deciding to use the corpus to correct particular errors. Whilst these particular error types were the most problematic, they only ever accounted for about a third of all the errors at any one point. From the start, then, it seemed the use of the corpus was only conceived by the learners as useful for a particular error type. Secondly, the use of the corpus regardless of the type of error diminished as the course progressed. By the final cycle, the majority of learners in the EG had stopped employing it as a reference tool. I have suggested that this is likely to be a result of the problems encountered when using it which are essentially the extra time, thought and energy taken to both use the interface and to analyse the concordances.

This means that the corpus can be judged as an effective reference tool when used for correction, but the number of caveats to this produced by the results beg the question as to whether or not it is actually worth employing the corpus as a correction tool at all. This will be clarified to a greater degree in the next section when overall accuracy and the potential benefits of the corpus to acquisition are looked at. However, when Cobb (1997, p. 314) suggested that the time taken to learn how to use a corpus, and I would then add the time taken to then go on to make effective use of the corpus, might be better spent elsewhere, he may have had results like these in mind. Certainly, responses from the learners in the EG seem ambivalent, if not to say sceptical, about the effects of the corpus corrections; few

seemed sure of any tangible benefits and comments on use were often negative, as will be shown in more detail in section 5.4..

The next section discusses the improvements in accuracy of both groups during the course and examines the extent to which the draft/revision/feedback cycle and the use of the corpus may or may not have encouraged language acquisition and a more autonomous approach towards language learning.

5.3 Accuracy and Long Term Improvements

The previous section showed that the revision process had a positive effect upon the errors being produced; revisions were consistently better than the drafts. However, from draft stage to draft stage, the same type of error kept reoccurring. Research question 2.1 asks if the learners who used the corpus were able to show a measurable increase in accuracy. RQ2.2 investigates the extent to which the corpus may have influenced any improvement in accuracy and if this is an indication that the corpus had longer term benefits for language acquisition.

5.3.1 Improvements in Accuracy

In answer to RQ2.1, the group which used the corpus did demonstrate an improvement in accuracy across the course, as measured in EFT (Error Free T-units). The line of progression fluctuated slightly, meaning progression was not linear over the course, but the general trend was that of improvement. Each revision represented an improvement in accuracy on the drafts of up to 58%, with a mean improvement of 41%. This follows the general trend, of course, of errors being successfully corrected at the revisions stages. Additionally, the drafts at each cycle show an overall improvement in accuracy over the period of the course. Drafts at the fourth stage were approximately 39% more accurate than drafts from the first cycle.

This suggests that the course was having a positive effect on the accuracy of the EG written work.

However, these results should be compared to those of the CG, which had no access to the corpus. Accuracy also improved in this group. Revisions represented an improvement of up to 50% on the drafts. This is lower than the EG, but the mean average improvement across the course in the CG was 44%, which implies a more consistent improvement in revisions that the EG. Drafts at the fourth cycle were just under 24% more accurate than drafts in the first cycle.

The pattern is repeated in the baseline and final stages; both groups showed an improvement in accuracy, but this is greater in the CG, at nearly 35% whereas the EG shows an improvement in accuracy of just over 25%. A conclusion might be that belonging to either group meant an improvement would be made, but that belonging to the CG was more likely to be a factor in improvement. However, the statistical results support only the first of these interpretations. That is, being a member of either the CG or EG group was not a significant factor in improvement. Rather, the main factor was cycle. This in itself is at least confirmation that the course had some positive effect on the learners, though it doesn't indicate exactly how this improvement can be accounted for. It merely shows that there was a learning effect and this could be measured in terms of linguistic accuracy.

This supports the comment made in section 5.2 that there is a place for linguistic improvement on EAP pre-sessional courses. Whilst examination of improvements in error production was able to show a benefit of the revision process at individual stages, the effect of an improvement between the baseline stage and final stage of the course, independent of any revision phases, implies that the course may have had longer term effects on the linguistic proficiency of participants. Again, this is an important effect to note because, as stated above and in the literature review, practitioners of EAP courses have seen linguistic

improvement in the short space of time available in a pre-sessional context as either of secondary importance to text organisation or as something unachievable (Basturkmen and Lewis, 2002; Banarjee and Wall, 2006). Another variable which was a factor in improvement is learner type, several of which were shown to be of statistical significance. This will be examined in detail below in section 5.4.

It seems that the learners in the group which had access to the corpus use were able to produce more accurate drafts by the end of the course, that these were more accurate than the CG, and that the improvement in accuracy was greater than that in the CG, but that statistically it was largely the cycle which contributed to accuracy. I will now go on to examine the extent to which corpus corrections might have contributed to these improvements, but first there are a number of other factors to take into account which are likely to have influenced a higher degree of accuracy.

The first is, predictably, input from classes. Whilst the course text, (Corballis and Jennings, 2009) did not place a specific focus on linguistic accuracy, and was used mainly to present the macro-features of texts, there was ample opportunity for the learners to gain knowledge of linguistic features. The texts incorporated into the book for reading practice, for example, covered a range of ESAP lexical items which were relevant for users. Classes also included direct input on the type of academic register appropriate to the needs of the course, and, as pointed out with reference to the reduction of formal and informal errors early in the draft cycles, this may account for an improvement in accuracy. Further, as noted in the methodology chapter, learners were involved in classes which aimed to develop summarising, paraphrasing and synthesising skills, all of which involve a significant amount of language manipulation. Practice in these areas may have contributed towards both a greater depth of knowledge about the language and a heightened awareness of the need for linguistic accuracy.

The individual interlingual stages of the students also need to be taken into account; in order to provide measurable data, I have looked at the learners as a group. This entails losing a sense of the learners as individuals and their improvements, or indeed fluctuations, in accuracy may well have been due to interlingual phases and any number of personal factors which were either not controllable or not possible to measure. As Wolfe-Quinetro et al (1998) have described, accuracy is just one facet of written proficiency as a whole. In measuring only accuracy within this experiment, I am of course only likely to obtain part of the picture of the learners' linguistic development as a whole. Accuracy, for example, may have improved at the expense of fluency. Alternatively, learners may have been producing far more complex sentences at the end of the cycles and this may have caused a decrease accuracy. The experiment didn't measure any of these potential variables or the interaction between them. Examining the relationship between fluency, accuracy and complexity would be an interesting point for further research as it would give a more holistic view of learners' ability to write proficiently. It may also be possible to investigate the effects of corpus consultation on areas of fluency and complexity. Perhaps, for example, corpus consultation in a research rather than reference vein would encourage students to develop more complex written structures. Research by Cresswell (2007), where learners examined and theorised about language use or Lee and Swales (2006), where learners investigated the differences between their own writing and expert writing, indicates that through exploration of concordances, students might be able to develop both a better knowledge of language rules and also begin to adopt patterns of usage.

5.3.2 Improved Longer Term Accuracy and the Corpus

I have already made the case that the corpus was an aid to error correction and that it may be a tool which is particularly useful for investigating lexical errors, especially those which

involve wrong word and, specifically, synonym errors. However, I have also noted that the use of the corpus accounted for very few of the accurate revisions performed by the EG as a whole.

What is more difficult to judge, due in large part to those less tangible factors mentioned above, is whether or not the corpus had a beneficial effect not just upon the accuracy of the language at revision stages, but on the accuracy of learner language outside the revision process. That is, if the corpus can be judged to have some longer-term impact on learner accuracy and language knowledge.

This is a problematic area within the literature. Truscott (1996) has asserted that there is no tangible long term benefit on long term learner development that can come as a result of participation in error correction cycles. This is because language acquisition, she maintains, is a process which occurs over a long period, replete with interlingual phases, so that any measurement of the efficacy of interventions made during a short course are unlikely to provide reliable data. Similarly, whilst Polio, Fleck and Leder (1998) suggested that interlanguage can be influenced by the awareness raising nature of error revisions, they were unable to provide any evidence for this having a long term effect on language development. Gaskell and Cobb (2004, p. 305) also state that only through studies with a longer time frame can researchers hope to identify elements of development which are distinct from the " 'U'-shaped pattern" typical of non-linear acquisition of language. Their experiment ran for a period nearly three times the length of mine and was still judged to be too short

The statistical tests to ascertain which independent variables produced errors at the baseline and final stage in both groups certainly imply that measuring the effects of the course are challenging. The test on the CG (Figure 4.40, section 4.8, chapter 4) described an increase in the prevalence of errors at the final stage. As noted, this seems to contradict the results which report on accuracy because these show a statistical improvement in the

production of error free t-units at the final stage. What this seems to indicate is that a rise in overall accuracy does not necessarily correlate with a fall in the capacity for error production. This experiment has not measured elements of written complexity or fluency, but it is possible , according to Wolfe-Quintero et al (1998) that writing could improve in only one or two of these areas at any one point without being more accurate. A possible scenario, therefore, is that learners improved in the complexity of their language. This might have produced more complex sentences with a greater number of accurate t-units but, at the same time, have been responsible for the production of more errors. This serves as an example of the complexities of achieving a satisfactory measurement of improvement in writing. The tests on the EG serve as another example; this group appeared to have improved to a greater extent than the CG but, when the baseline and final errors were compared, there was no statistical evidence to suggest that on that errors less likely to be produced at the final stage. Again, this points to a need for further research into both an understanding of the relationships between different elements of writing proficiency and the way in which these can be measured.

I examined the presence of wrong word errors at the baseline and final stages within the EG to ascertain the effect of the course on the production of a particular error type. This was the most prevalent form of error produced by the EG at the draft stages and the type of error which the corpus was most often used to correct in revisions.

Data from examination of the baseline and final stages shows that there was some effect in that the percentage of WW errors was lower in the final stage. In the EG at the baseline stage, WW errors accounted for 26% of errors but at the final stage this had dropped to 20.59% of errors. For the CG, WW errors at the same stages accounted for 23.97% and subsequently 15.62% of errors. In both groups, then, there was a positive result in a reduced percentage of wrong word errors. This is off-set of course by the fact that a reduced

percentage of wrong word errors indicates the likelihood of either an increased percentage in another error category, or a greater variety of errors being made. The problem is compounded by the fact that when these errors were examined in more detail and synonym type errors were looked at, the proportion of synonym errors actually increased. In both groups, by the extended assignment synonym errors accounted for 13% more of the total errors than in the baseline. So, whilst the percentage of WW errors fell, the percentage of synonym errors within this category seemed to rise. This means that the reduction in the percentage of WW errors was accounted for not by a decrease in synonym errors but by a decrease in one of the other sub-categories. Synonym type errors not only remained problematic for both groups, but were actually shown to be more of a problem at the final stage. The implication is that there was no beneficial effect on the production of synonym type errors in writing independent of the draft/feedback/revision cycle.

The statistical test in figure 4.42 (Chapter 4, section 4.8) compared both the CG and EG and examined whether or not at the final stage one group was more likely to create wrong word errors than the other. The test showed that learners in the EG were statistically less likely to produce wrong word errors than the CG. When the data was reduced to examine only synonym errors, the results again suggested that the EG was less likely to produce synonym errors than the CG. This evidence implies that learners who had used the corpus to make revisions to wrong word errors and synonym errors throughout the course were subsequently less likely to produce that type of error in work independent of the draft/feedback/revision cycle than learners who had not used a corpus. However, this positive effect is confounded by the evidence which shows that, within the EG itself, the error type which had overall been the most problematic, wrong word, did not seem to be reduced. In fact, proportionately, it seemed to be more of a problem than at the baseline.

5.3.3 Summary

The course itself can be judged to have been beneficial in terms of overall accuracy for both of the groups. There was a significant, positive effect when the baseline was compared to the final stage. Attributing better performance at the final stage to involvement with the corpus is problematic due to the difficulty of making an accurate measurement. Given the results that I have I cannot reliably claim that the use of the corpus during the course had any subsequent positive effect on error production. Nor can I confidently claim that the use of the corpus may have been something which contributed to longer term acquisition. Firstly, both groups showed improvement and this was regardless of whether or not they had used the corpus. Secondly, the learners that had used the corpus did not show a reduced production of errors which had been treated in the revision phases. Although it seemed that when compared to the CG there was some statistical evidence to the effect that the EG was less likely to produce a wrong word or synonym error, when the errors within the group from the baseline and final stages were compared there seemed to have been no appreciable improvement. Again, this interpretation of the results suggests that an area for research could be into the development of a more reliable and precise measure of linguistic proficiency, and more will be said about the interference factors in section 5.6 which discusses the limitations and problems encountered during the experiment.

5.4 Student Reaction to Use

RQ3.3 examines learner attitudes and looks at whether these help to corroborate findings in previous research and also if they help to explain the quantitative data collected during this experiment.

Firstly, data which was collected through the Mumford and Honey (1992) questionnaire placed the majority of learners who took part in the experiment in the category

of reflector, at just over 61%. Of the remaining learners, half were placed in categories which included a blend with the reflector learning type. It is perhaps not surprising then that reflector was one of the significant factors in predicting an increase of accuracy by the final stage (figure 4.1, section 4.2.1.2, chapter 4). The fact that the reflector / pragmatist is the most likely predictor of accuracy is interesting. This learner type only accounted for about 11% of learners, a relatively small amount compared to other learner types. This might mean that this reflector/pragmatist learners were disproportionately more likely to react positively to the course. That is, whilst accounting for proportionately few of the learners, this learning type represented a significant predictor of increased accuracy. In the EG, statistically learners of a reflector/pragmatist blend or theorist were more likely to employ the corpus in the correction process (figures 4.19, 4.20 and 4.21 section 4.5.2,, chapter 4). Again, learners of this type represented a relatively small proportion of the EG, at 25% of the group. However, cross referencing the learner type data with that of instances of corpus use (figure 4.19, section 4.5.2, chapter 4) showed that learners of this type were those who either made more searches with the corpus, or who continued to use it over the term of the course.

To summarise Mumford and Honey's (1992) assessment of these types of learner, theorists are those learners who, upon making observations attempt to incorporate them into a logical system through rational and objective analysis. Reflectors are those who prefer to gather as much information about data as possible before reaching any conclusion as to exactly what that data means. Pragmatists are those who like to experiment with new ideas and methodologies in a commitment to problem solving activities. Some of these features suggest that theorists and reflector/pragmatists are suited to the type of learning which the use of the corpus promotes; inductive learning requires learners to adopt an investigative approach to the information which is presented to them, as in the form of the data presented

by the BYU-BNC interface, and encourages individual analysis. This may account for these learner types employing the corpus.

What is perhaps more important, however, is that reflector/pragmatist blends and theorist learner types accounted for a small proportion of the group. That is, three quarters of learners in the EG did not fall into these categories. In fact, nearly half were clearly reflectors. One of the elements of a reflector's learning style, to refer to Mumford and Honey (1992) again, is that that their *modus operandi* is one characterised by caution and an unwillingness to reach a conclusion. Considering the amount of information supplied by a corpus search, and the potential for some of this to be in conflict with the learners' expectations of the language, or other sources, including the teacher, (Breyer, 2009, p. 167) one conclusion is that this type of learner could face considerable difficulties in making satisfactory use of the corpus. The real issue within this experiment, then, is that 75% of the group which had access to the corpus fell within a learner category which was not a statistical predictor of corpus use, and so might be unlikely to benefit from the corpus at all. Some of the conclusions Poole (2012) drew from her experiment are relevant here. Particular learning activities may not be suited to all types of learner. If that is the case there are implications both for learner training, in trying to help learners who have difficulties with particular learning scenarios, and in for how much a particular application, for example a corpus, is employed on a course. Of course, there is also need for further research with a much larger sample.

Overall reaction to the use of the corpus was mixed. A majority of the learners within the EG viewed the process of reviewing and correcting their errors to be have a positive effect in the sense that it had made them more aware of the problems they had. However, from the point of view of the learners, this did not translate into a feeling that they had actually been able to correct the errors successfully or that the process had helped them

reduce their capacity for making errors. This is actually at odds with the data which was produced: students were both able to reduce their errors at the time of revision, and improved subsequently during the drafts. The implication of this is important, because it suggests that although accuracy was improving, the student perception was largely that there was not a tangible benefit to the process. This in turn might imply that the particular aspects of the language that students felt to be important or problematic were not being addressed by the feedback/revision cycle.

Only about half the EG thought that the corpus had had a positive effect upon their writing. As already discussed, whilst there is quantitative data to suggest that the use of the corpus helped with error correction, this did not translate convincingly into an overall linguistic improvement. Nevertheless, information from the interviews corroborated the data found from examining the corrections in that learners reported using it to address wrong word errors, to help with register and to find collocates; the use of subjective language like "I" and "In my opinion" was singled out as something, for example, which an interviewee had assumed was academic because it was acceptable for responses to IELTS tests, and also "strong" language, which made writing seem "too extreme". A majority of learners reported that the corpus had not helped learn the meaning of words but had helped with learning the use and this implies that the corpus might be useful in developing pragmatic or stylistic knowledge. The fact that the data shows synonyms were being examined with the corpus supports this; learners were not having problems necessarily with the semantic aspect of the language, but its use within particular contexts. An example was provided in the interviews by learner EG3. She described her typical reason for searching concordances with an example:

"The most difficult, I think, is to find the more exact words. For example, I want to say something 'has an effect on', but I think 'has' is too usual, so I don't want to use 'has' again, so I want to find a more…how to say it…a more exact word and I think if I just use the dictionary I can't find how to change 'has' but with the corpus I have the [synonym] function and I can also use 'exert' and so on."

If it is the case that the corpus is an aid to pragmatic understanding, then it offers some support for Flowerdew's (2009, p. 402) assertion that the corpus may help in aiding a top-down approach to language. The corpus seems to have been valued by learners for the information it provided about language which was not rule governed, or at least the rules had not been encountered in previous learning or reference sources, and which needed some form of cultural or social knowledge to be manipulated. In this sense, then, the use of the BYU-BNC may contribute to the creation of linguistic capacity (Widdowson, 1983) in that it could help learners to operate in a range of contexts which are not necessarily addressed in the immediate context of the classroom or in the learning materials provided on a pre-sessional course. If this is the case, then the use of the corpus might be a tool which can be used to contribute to the wide-angle approach to language knowledge that EGAP courses seek to encourage (Bruce, 2011).

This examination of language for the furtherance of pragmatic knowledge by the learners adds to the weight of the arguments which have placed an emphasis on the benefits of the authentic language displayed in corpora over other reference materials. Thurston and Candlin (1998), Romer (2004) and Yoon (2011) have all noted that the use of authentic language might be of more benefit than that found within traditional teaching materials both because of its variety and because it can highlight assumptions made about the language which are not necessarily correct in practice. There was, in fact, an instance of this during the

experiment, when I had annotated a construction made by learner EG11 as incorrect. He had produced 'keeping the balance' which I had immediately marked as wrong, my instinct being that this should be 'keeping a balance' within the context in which it was written. However, on returning the draft to me, the learner had noted that, according to the concordances he had looked at, there was a precedent for his usage and he was in fact correct. This was an illuminating incident, showing a real instance of the language data correcting the assumptions of the instructor and also illustrating that an inductive, critical and analytical approach towards the language was being encouraged through the use of the corpus.

These potentially positive aspects of corpus use aside, the learners did note a number of problematic issues. A suitable place to start is with the issue of authentic language and the example of the learner who noted the discrepancy between my information and that provided by the corpus. Learner EG11 found that there was a problem in that when he found language in the corpus that was authentic but infrequent, he was unsure as to whether or not this was grounds for using it. This meant that his next course of action was to try and "think of something better" and only use the data from the corpus as a last resort. This uncertainty about the language information supplied by searches of the corpus and analysis of the concordances was prevalent with other learners. Learner EG4 noted that, after consultation with the corpus, she wanted confirmation that she had chosen the right language and so checked with peers or a dictionary and EG2 explained that when a search had been completed and the concordances produced a number of options which she would need to consider, she would turn to an electronic dictionary for help.

There were a number of other problems associated with corpus use, all of which have a precedent in the literature. The majority of the learners felt that corpus consultation took too much time. This may in part explain the reason for its lack of use throughout the course in general but could also help to account for the decrease in use after cycle one. As already

noted (Yoon and Hirvela, 2004; Frankenberg-Garcia, 2005), the time and effort involved in making searches was a nuisance factor in its use. Coupled to this is the fact that when learners did employ the corpus, half of them were unable to find the information they thought they needed. The same number of students claimed that it was difficult to make searches using the corpus, that there were too many concordances and that the concordances themselves were difficult to understand. The interviews corroborate this; the corpus was described as "inconvenient", "difficult" and "not that effective". One of the main reasons for this seemed to be the amount of information that, upon having made a search, the students then had to work their way through, and so it was judged to be too time consuming.

The learners also seemed reluctant to use the corpus in the composition of work. 60% of survey respondents provided a negative answer when asked if they had used the corpus independently of the draft/revision cycle. The interviews highlighted a number of reasons why this was the case. The first, again, was convenience. Electronic dictionaries were judged to be a better tool due to familiarity and convenience; interviewee EG4 actually implied that merely having to go onto the internet and then open the program was a reason she did not use the BYU-BNC. It seemed that the corpus was also judged by some to be a barrier to effective composition. Learner EG5 stated that the reason she did not use it when actually writing was because:

"You can see many things in the [dictionary] page, but if you use a corpus you have to…, it's complicated tool for us to use, so when we write we need to catch the ideas sparkling in our minds, so we don't stop to find the word."

She seems to be suggesting that in taking the time to use the corpus to examine particular elements of the language, the compositional flow would be jeopardised.

The lack of use of the corpus at a compositional stage is almost certain to have some grounding in the way the BYU-BNC was presented to the learners. My experiment was concerned with the use of the corpus as a reference tool and it was only ever presented to the students in this guise. The functions were examined only in conjunction with correcting errors which were already in front of the student and interviewees noted that it was only helpful when the error had already been pointed out to them. None of the learners, therefore, seemed to recognise the potential of the BYU-BNC as a research tool for exploring the language independently, as in the studies by Turnbull and Burston (1998) and Cresswell (2007). This is hardly surprising given the lack of training in the corpus as a language research tool and also given the already established problems of time and complexity. Claims that corpus investigations of language may help decentre the language (Lee and Swales, 2006) cannot be reliably supported by this study, again because the students did not use the corpus as a tool to research their own language but remained reliant on my input in the form of the correction code to guide them to the language they should be investigating. Chambers and O'Sullivan (2004) reported a similar finding. Whilst they used indirect feedback, learners in their experiment underwent a similar process of using concordances to correct the errors which were underlined. The researchers instructed students that, as well as attending to the areas of the text which were marked, they should feel free to use concordances to look at any aspect of language within their draft independently of teacher intervention. None of the learners did so, with one reporting that without having the areas highlighted for them, they would find it hard to improve (p. 163).

A year after the completion of the course, several of the learners who had originally employed the corpus during the pre-sessional replied to a follow-up survey. Three of the five learners had not employed the corpus since entering their course. The reasons given were similar to those for not using the corpus originally. One learner reported that she hadn't felt

fully proficient in using it at the time of the pre-sessional and that, as it was akin to learning an entirely new process, the use of the corpus was too time consuming once her master's programme had started. Another learner found that they did not often encounter the type of linguistic problems with which the BYU-BNC might be helpful and that, when they did, alternative sources of information, such as language learning blogs and forums were more effective in meeting their needs.

5.4.1 Summary

The evidence provided by the questionnaires and surveys adds to that provided by the quantitative data in that it helps to explain the patterns of use which were encountered, serving to answer RQ3.1. The application of corpus derived data to revisions was limited to a very particular area of language, as was its effectiveness. It seems the learners quickly became aware that the most useful application for them would be for the correction of errors which were lexical in nature and which were generally derived from some gap in pragmatic knowledge. For this reason, the learners singled out the synonym and collocate functions as ones which were most often employed and also identified the use of the corpus as something which helped with understanding the usage of words rather than the meaning, which had almost always been prevalent in the draft anyway.

Limited use, however, was not just a result of learners recognising that the corpus might be a benefit with a particular type of error. Use was also limited by learner type and it certainly seems that the investigation of corpus data was suitable for only a small proportion of learners. Employment of the corpus was also limited by what learners seemed to judge as inherent disadvantages: the time taken to make searches; the volume of information; the difficulty of reading corpora concordances and the uncertainty as to whether the information supplied by the corpus data was either right or was subsequently being applied in the right

manner by the learner. Learners also only ever seemed to view the corpus as a tool for the correction of errors, rather than for the generation of writing and as such did not make regular use of it outside of the revision process, giving a negative answer to RQ3.2.

These conclusions are slightly at odds with those drawn by participants in Yoon and Hirvela's study (2004). Although they note that learners perceived the corpus to be potentially beneficial for understanding the meaning of words (p. 277), they note a generally positive attitude towards the corpus, particularly in terms of contribution to improvement in writing and increased confidence (p. 278). My experiment indicated a less enthusiastic attitude towards the use of a corpus within the classroom and an ambivalence about whether or not it had had a beneficial effect overall. Like Yoon and Hirvela's study (2004) mine suffers from a small sample size made from a very specific group of students, and so further studies into student attitude on a greater scale are needed to provide a more accurate picture.

5.5 Autonomy

Because the students did not use the corpus as a research tool and because they did not employ it outside the revision cycle in compositional work, one might conclude that the corpus had little, or even no effect, on the development of autonomy during the course. I would argue that the opposite could also be true. Certainly, the limited use of the corpus and what seems to be a lack of concordance investigation on the part of the students might be evidence that inductive learning was not occurring to the extent that practitioners such as Nunan (1998) envisaged, with learners making enthusiastic explorations of the language. In fact, it might be argued that providing students with the language they were to correct may have undermined their potential to investigate the language on their own terms. Equally, the data used to answer the question as to whether or not the corpus users were able to demonstrate that they had actually 'learned' language as a result is by no means conclusive.

Finally, if one of the requirements of an EAP course is to equip learners with the tools necessary for future, independent improvement (Alexander et al, 2008), then the mechanism involved in the experiment might not be deemed the most suitable as learners are unlikely to be given the same level of detailed feedback, particularly linguistic feedback, when completing written work within their department. However, when one reviews some of the concepts of autonomy within EAP, aspects of the quantitative data and the responses of the students show that an independent attitude was being developed by the learners and that the implementation of the corpus helped to achieve this.

Firstly, research into learner autonomy has already attested to the importance of the teacher being part of the process (Little, 1990) and that developing autonomy is not simply a case of the learner being left to 'get on with it'. Although Hedge (2000) and Holec (1991) describe autonomy as a series of stages which the learner is ultimately responsible for, the achievement of these stages is dependent on the right kind of environment being in place. The draft/feedback/revision process employed in this experiment might seem to be prescriptive, particularly when I signalled to the learners what the errors were. However, when one looks at some of the criteria which have been established for developing autonomy, it can be seen that the process provided the opportunity for learners to do so. Learners were able to think objectively about the errors they had made. More importantly, they were able to make critical decisions as to what mode of correction they employed during the process. In being given a task in which they were acting on their own to correct only the errors they had produced, the learners were also being encouraged to act independently.

If part of being an autonomous learner involves being critically disposed towards the suppositions of the instructor, the methodology and tools being employed and the language itself, (Jordan, 2002; Fletcher, 2004; Perez-Parades and Cantos-Gomez, 2004), then the choices made by the learners as to what correction method they chose for particular contexts

are as much evidence of independent learning as they are evidence of problems with corpus use. The decision not to use the corpus based on either its unsuitability for a grammar error or the opinion that concordances provide too much information compared to other correction methods and that time has to be managed is evidence, it could be argued, of the learner taking responsibility for their own language learning. Yoon (2008) has said that even if corpus isn't felt to be useful in itself, the awareness raising it prompts might be.

The quantitative data shows that learners were capable of identifying that a particular error type might be better served by a particular correction method. The surveys and questionnaires show that the learners were aware of the usefulness of the corpus in particular areas and were making choices in its use based on criteria they had developed through assessment of the corpus: it was useful for synonyms and wrong words, register and linguistic problems of a pragmatic nature. At the same time, the learners were able to make critical evaluations of the BYU-BNC based on its complexity and the added effort and time they would have to make in order to use it. The fact that the learners chose not to use the corpus as a tool for composition reflects their assessment of its benefits and drawbacks. A rejection of the corpus may also indicate a healthy awareness of the different approaches available and, as Hunston (2002) has said, if learners feel more confident in treating English as a set of grammatical rules rather than a process of lexical discovery through concordances, "they should be allowed to do so" (p. 196).

It must also not be forgotten that the course was only for the period of six weeks. Just as identifying improvements in language knowledge is complicated by different learner types, attitudes and interlingual phases, so autonomy is difficult to assess accurately over a short length of time. This suggests that an area for further research could be in trying to get a more detailed understanding of how learners felt their learning independence developed over a pre-

sessional, perhaps by eliciting a series of responses at different stages, rather than just at the culmination of the course.

5.6 Limitations

As has been noted at various points in this chapter, the findings of the experiment are subject to a number of limitations which need to be discussed before any recommendations or conclusions can be made.

The first limitation is the population from which the sample was drawn. All of the findings relate only to a particular type of learner: mainland Chinese students going on to study management degrees at masters level. They were also of a comparable level, which was between 6 and 6.5 on the IELTS scale. The results can only reliably apply to a population comprised of this description of learner. The same experiment undertaken by Arabic undergraduates might supply an entirely different set of results, not only because of the language background but because of the learning environment which students had been used to. Similarly, learners of a higher or lower level may affect any of the outcomes of the experiment – overall accuracy, the use of the corpus, the effectiveness of the corpus and the perceived success of the corpus. Future research could test the results of my experiment against those of different learner populations. It would also be of interest to discover if learners who might have a vested interest in corpus use, such as linguistics students, produce entirely different results to those of students whose degree major is not connected with language.

The sample size is an issue. A large amount of data in the form of errors and revisions was generated but this came from a sample size of less than thirty participants at N=26. Garner (2005, p. 113) states that it is at N=30 that any sampling distribution fits closely to the population mean. When statistical testing is taken into account, an

improvement to this experiment would be a greater sample size in order to produce a more confident result. This might be achieved by co-ordinating a number of practitioners to collect data. In my experiment, I was limited to the number of students I had direct access to. However, assuming that corpus training could be given to a larger body of students, the draft/feedback/revision mechanism could be performed by any number of instructors as part of their normal teaching practice, with copies of student work merely passed on to a researcher for analysis. Having a greater sample would also allow for a randomisation of the subjects involved, encouraging more confident interpretation of statistical tests.

The context of the course is also a factor which needs to be taken into account. This experiment was made in accordance with the time constraints of a six week course and it seems there is some observable benefit in the use of the corpus. However, whether this would be true on a shorter course or over a cycle of fewer iterations is unknown. A longer course might also provide differing results. Certainly, there is the argument that the longer the course and the more iterations, the more likelihood there would be in distinguishing between results which suggest development as a result of the course and development derived from changes in interlingual stages. How both a reliable sample could be found and adequate control of conditions could be assured in a longitudinal experiment is a factor which does not yet seem to have been addressed in the literature.

Another inevitable constraint with a study of this sort is the correct labelling of linguistic items. This was discussed in the methodology but it is worth reiterating here that error coding on my part is almost certain to be affected by some form of subjectivity, particularly as most of the errors which were examined in detail were lexical in nature. Here, I encountered the same problem as the learners in terms of treatable and untreatable errors. Those treatable errors, the ones which conform to a rule, are possibly those which are less likely to suffer from subjective assessment. Those items of language, largely lexical, which

are not subject to a prescriptive grammatical framework are much more susceptible to inaccuracy in their labelling. Whilst I made sure that an inter-rater reliability test was made of the initial errors, this was not repeated for the sub-error categories of wrong word and formal/informal errors. An improvement on the experiment, if only for reliability, would be an additional inter-rater reliability test on these errors.

A further issue regarding the type of errors examined involves the distinction between an individual error and error type. This is foregrounded by Bruton (2009). He examines a previous study by Truscott and Hsu (2008) which itself was trying to identify if error corrections of drafts could contribute to any observable difference in the accuracy of the same language in subsequent writing. The problem Bruton identifies is one in which a student had made a number of indefinite article errors in the first piece of writing. These had been corrected after error feedback and then the second text was produced. In this text, indefinite article errors were again a problem, leading the original researchers to identify no significant observable improvement in subsequent writing i.e. the same errors was being made. However, Bruton (2009) shows that the indefinite article errors produced in the first text were not the same type of indefinite article errors which were produced in the second text. He concludes that "None of the other errors in this text could be attributable to a lack of learning from previous corrections as none of them correlate" (p. 139). This has obvious implications for any study which uses error categories, in that the degree of accuracy which can be achieved is restricted. Even though a range of error types were recorded in this experiment through the use of the error correction code, in Bruton's terms this is a relatively blunt instrument. For the lexical errors which I recorded, the issue is particularly acute. In theory, and in fact in all likelihood, a learner could create ten wrong word errors, revise them correctly, and in the subsequent draft create ten more wrong word errors using entirely different words in entirely different contexts. As the number of potential semantic and

pragmatic errors which could be created are innumerable, it would be virtually impossible to track the recurrence of them across a number of cycles. A way of establishing more control would be to limit the amount of errors examined or perhaps to change the conditions under which the texts are produced. However, there would be theoretical, practical and ethical problems with this. Theoretically, the problem would be that learners would not be free to create a range of errors, and so establishing where the corpus might be useful would be more of a challenge. Practically, tighter controls on the types of error being made assumes that a particular type of error, and one which conformed to a grammatical rule enough for it to be trackable through a number of iterations, would definitely be produced. Obtaining this information for a pre-sessional course would be difficult, given the time constraints that usually apply, but possible. Learners could produce a text before the start of the course, perhaps submitted online, which could be analysed for typical errors. Alternatively, there are guides as to the type of errors which learners from different cultures and language backgrounds are likely to make, such as Swan (2001). Ethically, examining a very limited amount of language, particular in a real teaching context, is unlikely to meet a range of learner needs. For the purposes of this experiment, and others like it, the error coding may be imperfect but it is at least practicable.

As noted in the methodology, I was aware of a number of controls which were of a lower level due to the context of the pre-sessional course: learners from both groups were free to converse with each other, I could not be solely responsible for the language input the learners received and learners were free to make the corrections at home, for example. As the experiment was made in order to discover what effects corpus use might have in a particular learning context, these conditions are realistic and results obtained through a more controlled environment would not be indicative of the effect of corpus use within the same context. However, in terms of the accuracy of the results it would almost certainly have been better to

have more control of when the learners corrected their drafts in order to ensure that they were using the methods they stated on their revision sheets.

5.7 Implications for Future Research and Practice

In this section, I present a number of ideas for future areas of research and for research design. I also outline considerations which need to be made when implementing corpus use on a pre-sessional course. I have addressed a gap in identifying which particular errors were treated with the corpus as opposed to other methods, how effective this treatment was compared to other methods and, to a certain extent, why learners utilised the corpus in particular contexts. However, a further level of detail which is missing is what exactly the learners were doing when they used the BYU-BNC.

Perez-Paredes et al. (2011) looked at instances of actual corpus use. Through the use of a piece of software called 'Fiddler'(http://www.fiddler2.com/fiddler2/) they were able to track student-computer interaction to examine how often they were consulting the corpus and how many queries they were making. Though they concentrated on analysing only the frequency of searches, Fiddler can record "all the actions carried out on each web-page and all the information typed in" (p. 240). With the numerical information, they were able to examine how often users consulted the BYU-BNC, how many searches they made and how the frequency of these searches then compared to other search tools. By doing so, they were able to achieve a good impression of the actual use of the corpus by students, rather than use which was reported by the students. Gaskell and Cobb (2004, p. 306) were also able to record concordance search events through their system of supplying learners with links to the relevant corpus data (see Chapter 2, section 2.4.1).

The use of software like Fiddler in an experiment like mine would supply an extra level of detail and, assuming the software was stable, would probably provide a more

accurate report of how frequently users were making use of the BYU-BNC. What would be more interesting, however, is information about the nature of the searches themselves. For example, when a particular error was encountered, it would be of use to know how many searches were made before a decision was taken on how the correction would be made. Additionally, exactly what type of search made would further illuminate the cognitive process of the learners using the corpus. That is, whether the user was merely using the frequency lists of words or if they were actually scrolling through pages of concordances.

Following this method would also allow indirect feedback methods to be examined. In this experiment, the meta-linguistic correction code was used not only because it directed the learners towards their errors, but because it was a way of recording exactly which types of error were being targeted. With a system that automatically logged the corpus searches that were being made, there would not necessarily be a need for the code to be used. The researcher could simply highlight either an error with no code, or merely signal the line to which the error belonged, and the learner would then be free to analyse exactly what the error was and go on to formulate appropriate corpus searches if necessary. This may be of benefit to more advanced learners if the claims made by Ferris and Roberts (2001), that this method is likely to be more helpful in fostering accurate long-term language use, are ever shown to be reliable. In fact, Chambers and O'Sullivan (2004) and O'Sullivan and Chambers (2006) employed indirect feedback in a feedback and revision cycle because they had access to advanced learners. As already noted the results were suggestive of a positive effect although unfortunately there was no system employed for the accurate tracking of corpus consultations. This lead to the admission in the 2004 study that it was impossible to ascertain whether the corpus had actually been used in producing an accurate correction or if it was merely a product of awareness raising (2004, p. 164).

The interview responses suggested that the learners who took part in my experiment were making searches based on the synonym and collocate functions and that they were examining the genre charts to see if the word they wanted to use was of an appropriate register. What I do not know from this experiment is the extent to which the learners were using the concordances themselves. Flowerdew (2009) has reported that there may be an issue in the way learners interact with corpus interfaces which implies they are only reading vertically and are using corpus interfaces rather superficially in a "phrasebook fashion" (p. 399). That is, looking at the lists of words for the most frequent item, for instance, and employing that word based on its frequency. She notes that this is "in line with the slot and filler notion espoused by substitution tables" (p. 393). Not only does this mean they are potentially unaware of the saliency of words regardless of their frequency, but it suggests that they are not using the concordances and that therefore are not necessarily engaging with the language in the beneficial, inductive way that practitioners hope for. This is an effect which was also observed by Sun (2007). He noted that frequency of an item seemed to be the main criteria in selecting it for use, bringing into question what other criteria there might be and how much real processing of the language was actually going on. Shedding light on precisely how users interacted with corpus data and how this interaction in turn affected the accuracy of their responses to feedback or indeed affected their production of written work would help to ascertain the levels to which corpus consultation can be useful and to what extent it should be employed on EAP courses.

Preparing an experiment that measures all this data would almost certainly involve a much wider, well resourced, approach to data collection. The amount of data likely to be generated would mean an efficient way of processing the it would have to be established in order for any analysis to begin. Software which tracks student computer use might also be a logistical and ethical problem; I am not sure how willing learners would be to download

tracking software onto their computers and to have all the information relayed to a researcher. This could mean that the experiment was limited to a university computer lab, which could then eat into class time and, for the purposes of a pre-sessional course at least, have negative consequences in other areas of learning.

A further issue which this study did not address was that of EGAP and ESAP. All the learners who participated in the experiment were going into similar departments for post-graduate study, all of which had a focus on finance or business. There was the potential, therefore, to address the relevant ESAP vocabulary. My thesis made no attempt to dissect the actual language the students were examining in order to identify whether or not it might be relevant in the future. The problem noted by Bruton (2009) could perhaps be addressed if the feedback and revisions had been limited to a specific lexical range. The AWL (Coxhead, 2000), for example, could have been used to limit the vocabulary items being corrected by the learners. This may have provided the potential not just for looking at error types, but for tracking the accuracy of the use of actual words. This in turn might result in a set of conditions under which the actual acquisition of language on an EAP course could be measured, rather than just the effectiveness of the corpus in revisions. Similarly, a corpus relevant to learners, perhaps constructed before hand from the reading texts employed on the course, could be used.

A range of other factors are likely to influence whether or not the corpus is to be made use of on EAP pre-sessional course, or courses of any other type. The importance of training has been stated several times, but with only about 10% of British Universities employing the use of corpora on their courses, according to Jarvis (2004), it is likely that many practitioners are both unfamiliar with their operation and, more importantly, their practical application within the class. Frankenberg-Garcia (2012b) has observed that many practitioners are unaware of how to conduct corpus searches or of how to make use of the results which are

presented. Harmer (2001) has noted a general reluctance amongst instructors to use corpora due to having to cope with a new technology and approaches to language which are unfamiliar. If, as in this study, the usefulness of the corpus is limited to one area of language, it is arguable whether the amount of time invested in learning how to use the information in a corpus, learning the software and subsequently passing this information onto the learners will be balanced by the concrete benefits which are returned in improved accuracy.

A further consideration is the framework and goals of pre-sessional courses. Whilst I have demonstrated here that the accuracy of language can be addressed on a pre-sessional course, even a short one, and that corpus use may actually aid in this, how much weighting syllabuses give to linguistic improvement is an issue. It is likely that courses which emphasise text structure and function will dominate pre-sessional classrooms in the future. There are valid arguments as to why this should be so and because of time constraints many practitioners are quite naturally sceptical of the long term linguistic effects of courses. Convincing institutions that corpus use has a part to play in pre-sessional courses is unlikely to be straightforward.

Student reaction is also an important factor to take into account when considering the value of implementing corpus learning on a course. I have previously noted the fact that even though the learners might eventually reject the use of the corpus, this choice is perhaps evidence that they are taking responsibility for the way in which they choose to learn. However, this is surely the view from the researcher or practitioners side of the fence. Whether the learner is willing to evaluate the choice in the same way is doubtful. The BYU-BNC could simply be evaluated as an extra, unnecessary application which results in long-winded, confusing, chore-like exercises. I have already noted the fact that the students who responded to the follow up survey had largely abandoned the use of the BYU-BNC post-

course and this was due to their evaluation that the time taken to use it did not justify the results.

Another improvement which lies beyond just the use of a corpus which is currently available would be to produce a corpus interface which is designed not for researchers or those interested in linguistics for its own sake, but which would be accessible to learners in general. This is advocated by Krishnamurthy and Kosem (2007), who suggest that negative learner attitude and unwillingness to use corpus data might be due to the interfaces which are currently available. By creating interfaces which are more user friendly in terms of the classroom and the needs of EGAP learners, perhaps more use would be made of corpus data and more evidence for the benefits of its use would appear. The number of concordances generated by the BYU-BNC is one aspect of corpus consultation which learners found to be discouraging. One area of research already suggested by Frankenberg-Garcia (2012a) is that of access to a reduced number of concordance lines. Her study suggested that even access to a single concordance line provided more beneficial results in correcting errors than dictionary definitions, and access to three concordance lines produced even better results (2012a, p. 287). The compilation of a corpus which at once mediates the relevant data provided to the learner and is able to answer a broad range of queries has obvious challenges. It would have to involve numerous tokens for single items used in different contexts, for example. However, depending on the type of queries being made, it might be possible to produce a useful, simplified interface. The increased availability and sophistication of smart-phones and the already popular and widely available personal electronic dictionaries could provide a platform for making corpora of this type more accessible to learners if specific, user friendly corpus applications were produced.

5.8 Conclusion

The use of the corpus on an EAP pre-sessional course can be an effective way of helping learners to treat their written errors. However, this effectiveness seems to be restricted to a limited number of error types, specifically wrong word errors and errors of register. It appears that reference to the corpus can help with untreatable errors and that it may be of use to learners struggling with pragmatic and stylistic issues of the language. However, the extent to which the corpus is employed on a course as a reference tool should include a consideration of how important these error types are for the development of the learners. The increased cognitive strain placed on learners using the corpus for potentially limited rewards means that time might be better spent on other methods of correction or instruction.

There is not enough evidence from this study to make an unequivocal statement about whether or not long term acquisition or understanding of particular error types was aided by corpus use. Inductive learning may not have been encouraged in all learners who had access to the corpus, although the choices they made in methods of error correction might be suggestive of an independent, evaluative approach towards materials and methodology.

Use of the corpus was dictated by learner type and this is something which practitioners should be aware of. Any adoption of the corpus on an EAP course needs first to take into account the different learners within the class and the methods of interacting with corpus data need to be adjusted accordingly.

Areas for future research include the utilisation of the corpus with other population samples and the adoption of different methodologies which employ a corpus in the classroom. Using a corpus as a research rather than reference tool might encourage different patterns of use. Tracking and analysing patterns of use in the form of the different search queries which learners make in order to make a correction, for example, would encourage a better understanding of the cognitive processes employed by learners using corpora. This

information would be valuable for developing both methodologies of use for the classroom but could also inform materials created with corpora and might influence the design of future corpus interfaces.

6

Conclusion

6.1 Introduction

In this chapter, I provide a summary of the aims and findings of this thesis. I begin in section 6.2. by restating the general aims and context for my study and the research questions I investigated. I go on in section 6.3 to review the findings from the experiment before reiterating the implications that these have for the use of corpora on EAP pre-sessional courses in section 6.4. Finally, in section 6.5 I reflect on possible improvements of the experiment and summarise the areas identified as possible avenues for future research.

6.2 Summary of Experiment Context and Aims

In order to begin to address a gap between research into the beneficial applications of corpus use and its actual uses in a learning environment (McCarthy, 2007; Breyer, 2009), my experiment was designed to collect data on corpus use and effectiveness in the hands of real learners on an EAP pre-sessional course. The selection of this type of course is justified by the increasing importance of EAP preparation courses within the British university system and the emphasis in this system on written work as a measure of achievement. (Archibald, 2001; Alexander et al, 2008; Hyland, 2009a). Because the duration of these courses is usually quite limited, previous research has identified the prioritisation of developing competence in the organisational aspects of writing at the expense of improved linguistic awareness and ability (Archibald, 2001; Basturkmen and Lewis, 2002; Cho, 2003; Turner, 2004; Hartshorn et al., 2010). I observed that this was an area which might be addressed through the application of corpus data to revisions of linguistic errors made by learners. I also identified the positive effect that interaction with a corpus might have on inductive

learning and that this would in turn contribute towards increased levels of learner autonomy (Johns, 1991; Nunan, 1998; Benson, 2001; Watson Todd, 2003). In order to investigate the benefits of corpus use in these areas, I devised the following research questions:

RQ1.1 Are learners on a pre-sessional course able to correct errors in their writing by using corpus data and concordancing as reference tools?

RQ1.2. Is corpus use chosen as a method of error correction over other methods in particular circumstances?

RQ1.3. Are particular error types more or less effectively dealt with using corpora as reference tools than with other correction methods?

RQ2.1 Do learners who use corpora as reference tools show any measurable improvement in their written proficiency?

RQ2.2 What aspects, if any, of student writing improve and to what extent might this improvement be judged to be as a result of having accessed corpus data when making corrections?

RQ3.1 What are learner attitudes towards the use of a corpus as a reference tool on a pre-sessional course?

RQ3.2 Do learners begin to use the corpus independent of teacher instruction?

RQ3.3 Do learners' perceptions of corpus use corroborate the quantitative data and to what extent does this conflict with the existing literature?

To address these questions, I designed an experiment which took place over the course of a summer pre-sessional course at the University of Sheffield. 26 Chinese learners going on to

study master's degrees in management and finance related subjects took part and were split into a control group (CG) and an experimental group (EG). Over the course of six weeks, the learners produced six pieces of written work. These were a baseline and final text, with 4 texts produced between these stages. These texts were subject to a draft/feedback/revision cycle. Learners corrected their errors and noted down their method of correction. The EG was given training in and access to the BYU-BNC (Brigham Young University – British National Corpus) interface to aid correction. This data was supplemented with information about learner type and responses to surveys and interviews. The entire experiment was embedded within the pre-sessional course, entailing a number of limitations but also providing an observation of corpus use within a real learning context.

6.3 Summary of Findings

Here, I summarise the observations from the experiment which are detailed in chapters 4 and 5. Section 6.3.1 addresses research questions RQ1.1, 1.2. and 1.3. and notes that there was some benefit to using corpus data as a reference tool for the correction of a limited range of errors. Section 6.3.2 summarises the findings related to RQ2.1 and 2.2. The conclusion here is that whilst there was a measurable effect of the course on accuracy, to what extent this can be attributed to the use of the corpus remains uncertain. Finally, section 6.3.3 reviews the information about learner type and the responses of the students to questions about their use and perception of the corpus. Corpus use might be suited to some learner types more than others and, in general, the attitude towards the effectiveness of the corpus was ambivalent.

6.3.1 Error Correction and Corpus Use

The results provide evidence to the effect that there might be benefits to using corpus data for error correction in written work. A number of errors could be corrected with the data from

the BYU-BNC, but in particular wrong word errors were targeted both most often and most effectively. Specifically those errors which involved a correction of a synonym were addressed through reference to the BYU-BNC. There is some evidence, though not statistically significant, to suggest that these types of error might be more effectively corrected with the BYU-BNC than with other correction forms that are commonly used by the learners. Students had produced constructions where the sense was correct but where the use of a particular word was incorrect within the context. Examination of corpus data enabled them to choose a suitable alternative.

These language items are generally what has been termed 'untreatable' in that they do not follow a particular grammatical rule (Ferris et al, 2000). This may be an indication that the corpora can be useful in the correction of errors which are pragmatic in nature. Sinclair's COBUILD project, resulting in the Collins range of dictionaries (Hands, 2009) used corpus to develop a dictionary style which provided single authentic examples to address this issue. Frankenberg-Garcia (2012a) has produced research which suggests that corpora themselves may be even more effective in this area. Corpus data may be a valuable resource for learners who do not have a fully developed sense of certain elements of the language. These might be questions of register, of a particular subject area, or of cultural knowledge. The use of corpora might, therefore, be able to address aspects of the language which would otherwise be inaccessible due to a lack of cultural knowledge, and as such might be an aid for students trying to engage in top-down processing of the language. However, the use of the corpus was extremely limited, or non-existent, in other areas of error correction. It accounted for a very limited number of accurate error revisions, at under 20% across the whole pre-sessional course. This brings into question whether or not the benefits of using the corpus justified the time spent by learners in training and corpus consultation.

6.3.2 Accuracy and Corpus Use

The linguistic accuracy of learners increased across the course, in both the EG, which had access to the BYU-BNC and the CG, which did not. These increases in accuracy were observable both between drafts and revisions, between the drafts themselves, and between the different revisions. Both groups also showed an increase in accuracy from their baseline point at the start of the course and their final, extended writing point. This indicates that the pre-sessional course had a beneficial effect on the linguistic aspect of students' written work, as opposed to just the organisational aspects. This is a useful finding because there has been some conjecture as to whether or not pre-sessional courses, typically short in length, can help learners in a linguistic sense (Basturkmen and Lewis, 2002; Cho, 2003). The evidence from my study is that they can.

However, overall increases in accuracy were not matched by significant decreases in problematic error types. In the EG, the effective corrections of wrong word errors in the draft/feedback/revision process did not eliminate this error as a problematic type. Although there was some statistical evidence to show that this type of error was less likely to be produced at the final stage rather than the baseline stage in the EG, it was not possible to confidently state that this was as a result of using the corpus over the course. Because of the fluctuations in accuracy and the short length of the course, it was also not possible to judge whether these improvements were anything other than symptoms of interlingual stages.

6.3.3 Learner Evaluation of Corpus Use

As well as the corpus being limited to only particular types of error, its use was limited to particular learners. Only those learners who fell into the reflector/pragmatist and theorist categories made any significant use of the corpus. These learners accounted for only a quarter of the EG. In other words, the great majority of the group chose not to use the corpus

as a reference tool when correcting their errors. This indicates that the usefulness of the BYU-BNC might be restricted to a particular type of learner. If this is the case, then caution needs to be exercised when prompting students to make use of corpora because it may be that only a limited number of learners are likely to gain any benefit.

Even amongst learners who utilised the corpus, frequency of use dropped considerably as the experiment cycles went on. This could partly be due to its limited areas of application, as noted above, but can also be explained by the attitudes expressed by the students when asked about their use of the BYU-BNC. Whilst some recognised that the BYU-BNC was useful for correcting wrong word and synonym error types, the majority felt that it was too time consuming. Searching the corpus was found to be difficult, analysing the data caused problems and learners reported a lack of success in being able to find the information they were searching for. This may be a result of the interface rather than a linguistic issue. As will be detailed below, alternative and more attractive interfaces with corpus data may help to overcome these problems. Although corpus revisions had, on the whole, been successful, the learners did not feel that the use of the corpus had had a beneficial effect on their language in particular. Learners also failed to recognise the possibility of the corpus being a tool which could aid composition. In fact, some commented that it would actually hinder the writing process, and so it was largely just used within the draft/feedback/revision cycles. A year after the course, most of the learners who responded to an online survey were not using the corpus as a writing tool and felt it did not fit in with their studies as a whole. These largely negative views repeat those which have been made in other research which has examined student attitude towards corpus use (Lee and Swales, 2006; Yoon and Hirvela, 2004; Frankenberg-Garcia, 2005; Sun, 2007).

6.4 Teaching Implications

Perhaps the most important finding is that, on a short pre-sessional course of only six weeks, linguistic improvement can be made even when the syllabus has emphasised, as is typical of these courses, the macro-organisational elements of written texts. There has been doubt as to whether learners can actually improve their language in a short amount of time and, as such, the area can be overlooked in the classroom. This has given rise to the criticism that pre-sessional courses do not necessarily either provide learners with the type of linguistic input which they expect to have from EAP courses and that they are, at the culmination of the course, left unable to operate as effectively as they need to at the micro linguistic level. This can then have ramifications for the future if learners are unable to express themselves in future academic contexts (Turner, 2004).

However, how much corpus use contributed to this improvement is ambiguous. Firstly, improvement in revisions was over many different types of error. Corpus corrections accounted for both a very small range of error types and a small number of the total corrections made. The majority of corrections made, in both groups, were made without a reference source and with previous knowledge. The increase in accuracy in the revisions, and this is supported in the interview and survey evidence, is likely to be as a result of students developing an awareness of their errors through noticing (O'Sullivan and Chambers, 2004; Yoon, 2008). The increase in accuracy between the baseline and final stages may also be a result of this rather than a learning effect from any other input; learners valued the awareness raising during the corrections and this may have contributed to an overall awareness of the types of errors they were likely to make, thus enabling them to reduce these in the final stages.

The corpus was useful for correcting wrong word errors, and the way it was used is suggestive that it might be useful for the correction, and possibly raising awareness, of pragmatic elements of vocabulary, rather than syntactic elements. Increased knowledge of

the context in which language items can be used is obviously a benefit, but, given the results, whether or not the low proportion of errors corrected with the corpus justifies its use as a reference tool is still a matter of conjecture. Firstly, whilst corrections made using the BYU-BNC were more accurate than other methods, these other methods, knowledge and, in the CG, knowledge and dictionaries, were also largely accurate. Secondly, if the BYU-BNC is helpful in correcting errors which might be pragmatic in nature, this has only been shown to be so in a pre-sessional, EGAP (English for General Academic Purposes) context. When learners move onto their departmental course, the extent to which they would be using a wide variety of EGAP language is uncertain. Yoon (2008) has actually suggested that learners do not necessarily have a problem with their subject language. If that is the case, one wonders how relevant the language contained within the BNC would be. It is also the case that learners would, once enrolled with their department, be exposed to the ESAP (English for Specialised Academic Purposes) through their course input and in the form of reading texts. This would build up lexical knowledge so it is again doubtful as to whether the corpus would be valuable at this stage.

A further issue, particular in the case of this experiment, is that of the mechanism of its use as a reference tool. On a pre-sessional course, because the focus is on improvement in English, learners can, and often do, receive a lot of help in terms of feedback. It has already been noted that this is something learners expect (Chandler, 2003). In the case of the University of Sheffield's pre-sessional courses, all students and teachers are supplied with a correction code to use. From the commentary given by the students and the pattern of use, it seems that the corpus was regarded as part of the feedback mechanism, not as a tool which could be searched when producing original pieces of written work. During the course of their subject courses, particularly at postgraduate level, learners would be far less likely to receive this level of support. If this is the case, then the occasion for corpus use as a corrective

reference tool is far less likely to arise. It night be then that only presenting the BYU-BNC as a reference tool does not best serve learners' long term needs. Whilst it might help them with some lexical revisions, perhaps a more appropriate use of the corpus would be as a research tool. Studies by Cobb (1997) and Lee and Swales (2006) have shown that there was some positive response to using corpora as research tools.

Alternatively, or additionally, perhaps a better use of student time with regards to corpora in general would be in the construction of their own corpora based on their own subject areas, as suggested by Cobb (1997) and Renouf (1997). This would perhaps necessitate a longer and more involved period of training, because learners would have to both learn how to use a corpus interface for searches and how to construct their own corpus. However, if they constructed the corpus based on their own subject area language needs, they would have something personalised and relevant to them individually. More importantly, they would potentially have something which they could continue to use during their course of study but which they could also add to as their course progressed. This would extend the life of the corpus and might make it more relevant. Despite its apparent complexity, this type of approach might encourage greater use and meet the criteria of giving the EAP learner control and responsibility for their own language and learning.

The evident lack of enthusiasm for the BYU-BNC amongst the learners might be indicative of a broader resistance to inductive learning in general. That is, rather than just a simple rejection of the BYU-BNC, its lack of use might suggest that learners had difficulties in adopting a more investigative and independent approach to their language learning. It may be that, as a result of previous learning environments, the students were simply not used to being put into the role of researcher responsible for the aspects of their own language learning. This potentially unexpected additional cognitive strain may have been a negative affective factor in learner progress. This might apply to all the students or, as with the learner

types mentioned above, only those with a particular learning style bias. If it is the case that the lack of corpus use was symptomatic of unwillingness or inability to learn inductively on the pre-sessional course, then it might be that learner expectations were not fully addressed. As much of the literature points out (Jordan, 2002; Banarjee and Wall, 2006; Alexander et al, 2008), the role of an EAP course is generally considered not just to be that of equipping learners with enough language or knowledge to pass the course, but also to equip them with the skills which will enable them to take responsibility for their future learning. It may be that in using the corpus as a reference tool within a series of cycles which were heavily dependent on the teachers input, learners remained largely field dependent and were unable to see the importance of being able to take responsibility for their own learning.

6.5 Research Implications

Whilst the experiment was able to address the research questions, its prosecution created more queries than it answered. Below, I reiterate the limitations of this study and at the same time note that the findings indicate directions for future investigation.

6.5.1 Range of Language

One area which might be useful to examine in order to broaden the application of the corpus is other areas of language which it could be used to address. The results from this thesis saw it used for specific types of lexical error, but corpus use could be applied to grammatical areas. The ability to isolate different elements of POS (part of speech) might mean, for example, that learners who are struggling with the use or form of particular tenses could benefit from consultation with the BYU-BNC. Closer investigation of ways in which the corpus could be used to address a wider variety of linguistic issues at EAP level could halt the decline in corpus use over a course and may encourage a more positive learner response.

6.5.2 Sample Size and Composition

Future research should involve larger samples. This would facilitate the collection of more data, making analysis, particularly on a statistical level, more reliable. These samples should also include learners from a range of cultural and linguistic backgrounds. This might results a completely different range of language being targeted through corpus consultation. Equally, it might result in a range of different attitudes towards corpus use being expressed.

Level is another variable to be considered. This study examined a group of learners whose level was relatively high, if one measures in terms of IELTS score and the fact that they were only half a band away from their unconditional offer score. Application of the BYU-BNC might differ if the level had been lower, although there is conflicting evidence as to whether learners who are more or less advanced receive a greater benefit from corpus consultation. Kenning (1996, p. 131), for example, said that it was more likely that advanced users of the language who already had a capability for autonomous learning would be able to successfully interact with corpus data. Yoon and Hirvela (2004), however, found that, when comparing two groups, one of an advanced level and one of an intermediate level, the students of a lower level had a more positive attitude towards using the corpus (p. 272) and it was posited that this might be due to a greater degree of enthusiasm for achieving the target language than the advanced students. Subsequent research should be particularly rigorous, then, in assessing the level of learners when taking into account any results derived from corpus investigation. Likewise, pedagogic practice should take into account the possibilities of utilising corpus data with different levels and what form this utilisation should take in order to be of greatest advantage to the learners.

Building a clear body of knowledge on the effects of corpus consultation at different levels with different types of student would be a useful future tool in gauging when corpus consultation is most likely to benefit learners and in what contexts it is best employed.

6.5.3 Length of Study

Studies of a length greater than six weeks obviously have a number of advantages; more quantitative data could be collected through a greater number of iterations, there might be the opportunity to encourage different aspects of corpus use amongst learners, it might become clearer what type of language the corpus was capable of helping with and a greater quantity and variety of qualitative data might be collected. However, the problem would again be that of the divide between insightful research on the one hand and practical information about where the use of corpus data might be encouraged in actual classroom situations on the other. This means that whilst lengthier studies are needed, studies which examine the potential benefits, or otherwise, of corpus use within more limited time frames are also called for. Large numbers of students only experience an EAP learning environment for a short period, as in a pre-sessional course, uses of the corpus need to be made practicable in this learning context. Any information gathered as to the positive or negative effects of corpus use on actual courses could encourage a more enlightened decision on the part of practitioners when deciding if, when and how to enable learners to employ corpus data in their learning.

6.5.5 Sharpened Research Tools

One evident limitation in this experiment is the lack of knowledge gathered about the precise patterns of interaction which learners made with the corpus. I have been able to account for which errors were corrected with the corpus by asking students to record error correction methods. This also facilitated an analysis of the effectiveness of different correction methods. However, I do not have quantitative data on the types of searches the learners were making, the number of searches they were making and the time it took to correct errors with the corpus. Information in these areas would allow future research to analyse patterns of corpus interaction and might highlight particular areas where corpus use was beneficial. Studies into

this area might build on the methods employed by Gaskell and Cobb (2004) and Perez-Paredes et al. (2011).

6.5.6 Corpus and Interface Development

One prominent finding of this experiment was that part of the reason for the decline in corpus use seemed to be related to difficulty, or inconvenience, of its use. This finding has a precedent in the literature, as mentioned in section 6.3.3. It is not only an issue for learners, but may also be a factor which accounts for the reluctance of practitioners to use corpora. Krishnamurthy and Kosem (2007) advocate the construction of interfaces aimed directly at particular types of learner. Areas for investigation might be simplified interfaces, a limited amount of concordances to avoid concordancing burnout and the tailoring of corpus interfaces to mobile devices.

6.6 Conclusion

The previous sections of this chapter have summarised the results of my investigation. Ultimately, my study has demonstrated that the use of corpus data can have some measurable benefit on the accurate revision of particular learner errors. Specifically, errors which seemed to stem from a pragmatic area and which involved the correct use of a word within its co-text and context rather than its meaning seemed to benefit from the attention of the corpus. This indicates that there may be a tangible benefit using the BYU-BNC in dealing with errors which might traditionally be regarded as untreatable in nature and which would otherwise require cultural knowledge unavailable to learners newly arrived on a pre-sessional course. The experiment has also shown that a short-term pre-sessional course can have perceptible improvements on students' linguistic accuracy and that future pedagogic practice might benefit from reconsidering the balance between macro and micro – linguistic elements. The

use of corpora as reference tools may be beneficial in promoting students' ability to process pragmatic information with a top-down approach.

However, if the use of a corpus is to be incorporated into pre-sessional courses, a number of considerations ought to be taken into account and claims for corpus effectiveness and its long term benefits need to be judged with caution. I would suggest that pre-sessional practitioners should try to ensure that the use of the corpus is closely allied with learner needs. Additionally, the method by which a corpus is introduced to learners needs to be informed by knowledge of learner type, level, background and expectations. Corpus use should be introduced to learners from a background which places an emphasis on deductive learning, for example, with considerable care. The potential for improvements in independent learning skills and language knowledge will not be realised unless appropriate forms of training and mediation are adopted by instructors.

This experiment has shown that knowledge of the advantages of corpus use might benefit not only from more research into the long term effects of corpus use under controlled conditions, but also the short term use of corpora with learners in those teaching contexts which are most likely to be encountered on English courses. This study does not dismiss the value of corpora for the development of teaching materials, nor does it argue that corpora are without value for most students. Rather, it argues that, based on a case study of a particular group of students, there are some observable benefits and limitations to the use of corpora which might inform the future use of corpora in EAP classrooms. Additionally, the findings that the corpus proved less than popular with learners and that use declined might be used to inform technological developments. Making corpus software more user friendly and attractive to non-specialists might well prove to be a means of generating a step change in classroom-based corpus analysis.

7

Contents of Disc Appendices

8

Bibliography

Acton, C and Miller, R. (2009) *SPSS for Social Scientists.* 2nd Edition. Basingstoke: Palgrave Macmillan

Anthony, L. (2010). *AntConc (Version 3.2.1)* [online]. Tokyo: Waseda University. Available at http://www.antlab.sci.waseda.ac.jp/

Alexander, O. Argent, S. and Spencer, J. (2008). *EAP Essentials: A Teacher's Guide to Principles and Practice*. Reading: Garnet Publishing Ltd.

Archibald, A. (2001) 'Targeting L2 Writing Proficiencies: Instruction and Areas of Change in Students' Writing Over Time', *International Journal of English Studies.* 1 (2): 153 – 174

Bailey, S. (2011) *Academic Writing: A Handbook for International Students*. 3rd Edition. Abingdon: Routledge

Banarjee, J and Wall, D. (2006) 'Assessing and reporting performances on pre-sessional EAP courses: Developing a final assessment checklist and investigating its validity', *Journal of English for Academic Purposes*. Volume 5, Issue 1. Pp50-69

Bardovi-Harlig, K and Bofman, T (1989). Attainment of syntactic and morphological accuracy by advanced language learners', *Studies in Second Language Acquisition* 11: 17-34.

Basturkmen, H and Lewis, M. (2002) 'Learner Perspectives of success in an EAP Writing Course', *Assessing Writing* 8 (1): 31-46

Benson, P. (2001) *Teaching and Researching Autonomy in Language Learning*. Harlow: Pearson Education Limited

Biber, D. (2006) *University Language: A corpus based-study of spoken and written registers*. Amsterdam: John Benjamins Publishing Company.

Biber, D. and Reppen, R. (2002) 'What does frequency have to do with grammar teaching?' , *Studies in Second Language Acquisition* 24(2): 199–208.

Bitchener, J. (2008) 'Evidence in Support of Written Corrective Feedback' , *Journal of Second Language Writing 17:* 102-118

Breyer, Y. (2009) 'Learning and teaching with corpora: reflections by student teachers', *Computer Assisted Language Learning 22(2*): 153-172

Bruce, I. (2011) *Theory and Concepts of English for Academic Purposes.* Houndmills: Palgrave Macmillan.

Bruton, A. (2009) 'Designing research into the effects of grammar correction in L2 writing: Not so straightforward', *Journal of Second Language Writing 18:* 136-140.

Bryman, A and Cramer, D. (2005) *Quantitative Data Analysis with SPSS 12 and 13*. Hove: Routledge.

Burgess, J and Etherington, S. (2002) 'Focus on Grammatical Form: Explicit or Implicit?' *System* 30: 433 – 458

Butler, C. (1985) *Statistics in Linguistics.* Oxford: Basil Blackwell Ltd.

Chambers, A. (2007) ' Popularising corpus consultation by language learners and teachers', *Language and Computers- Studies in Practical Linguistics* 61: 4-28

Chambers, A., & O'Sullivan, I. (2004) 'Corpus consultation and advanced learners' writing skills in French', *ReCALL, 16 (1):* 158–172.

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* 12:267–296

Charles, M. (2007) 'Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions', *Journal of English for Academic Purposes 6:* 289–302

Charles, M. (2012) 'Proper vocabulary and juicy collocations: EAP students evaluate do-it-yourself corpus building', *English for Specific Purposes 31:* 93-102

Cho, Y. 2003. 'Assessing writing: Are we bound by only one method?', *Assessing Writing*, 8 (3): 165–191

Cobb, T. (1997) 'Is there any measurable learning from hands-on concordancing?', *System, 25 (3):* 301–315.

Collins Wordbank (n.d) [online] Available at: http://www.collins.co.uk/page/Wordbanks+Online. Last accessed 14th October 2014.

Corballis, T and Jennings, W. (2009) *English for Management Studies in Higher Education Studies*. Reading: Garnet

Coxhead, A. (2000) 'A New Academic Wordlist', *TESOL Quarterly* 34 (2): 213-238

Coxhead, A. (2011) 'The Academic Wordlist 10 years on: Research and Teaching Implications', *TESOL Quarterly* 45(2): 355-362

Cresswell, A. (2007) 'Getting to 'know' connectors? Evaluating data-driven learning in a writing skills course' in Hidalgo,E. Quereda, L. and Juan,S. (eds), *Corpora in the Foreign Language Classroom,* pp. 267–287. Amsterdam: Rodopi.

Cullen, P. (2012) *Vocabulary for IELTS advanced*. Cambridge: Cambridge University Press.

Davies, M. (2004) *BYU-BNC*. (Based on the British National Corpus from Oxford University Press)[online]. Available online at http://corpus.byu.edu/bnc/. Last accessed 19th September 2014.

Dudeney, G. and Hockly, N. (2007) *How to Teach English with Technology*. Harlow: Pearson Longman.

Erickson, B.H. and Nosanchuk, T.A. (1992) *Understanding Data*. Second edition. Buckingham: Open University Press.

Ferris, D. (2002) *Treatment of Error in Second Language Student Writing*. Ann Arbor: University of Michigan Press.

Ferris, D. (2006) 'Does error feedback help student writers? New evidence on the short-and-long term effects of written error correction' in Hyland, K and Hyland, F. (eds), *Feedback in Second Language Writing: Context and Issues*, pp81-104. Cambridge: Cambridge University Press.

Ferris, D. and Roberts, B. (2001) 'Error Feedback in L2 Writing Classes: How explicit does it need to be?' , *Journal of Second Language Writing* 10 (3): 161 – 184

Fletcher, W.H. (2004) ' Facilitating the compilation and dissemination of ad-hoc web corpora' in Aston, G. Bernardini, S. and Stewart, D. (eds) *Corpora and Language Learners* , pp271 - 300. Amsterdam: John Benjamins Publishing Company.

Flowerdew, J. and Peacock, M. (eds) (2001) *Research Perspectives on English For Academic Purposes*. Cambridge: Cambridge University Press.

Flowerdew, L. (2005) 'An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: countering criticisms against corpus-based methodologies' , *English for Specific Purposes* 24: 321-332.

Flowerdew, L. (2009) 'Applying corpus linguistics to pedagogy: a critical evaluation' , *International Journal of Corpus Linguistics* 14(3): 393-417

Frankenberg-Garcia, A. (2005) 'Pedagogical Uses of Monolingual and Parallel Concordances' , *ELT Journal* 59(3):189 – 198

Frankenberg-Garcia, A. (2012a) 'Learners' Use Of Corpus Examples', *International Journal of Lexicography* 25(3): 273-296

Frankenberg-Garcia, A. (2012b) 'Raising teachers' awareness of corpora', *Language Teaching* 45: 475-489

Garner, R. (2005) *The Joy of Stats: A Short Guide to Introductory Statistics in the Social Sciences*. Plymouth: Broadview Press.

Gaskell, D., & Cobb, T. (2004) 'Can learners use concordance feedback for writing errors?' , *System* 32: 301–319

Gass, S. M., and Selinker, L.(1983) *Language Transfer in Language Learning*. Massachusetts: Newbury House Publishers, Inc

Gillett, A. J. (2011) *What is EAP?[online]* Available at http://www.uefap.com/bgnd/ Last Accessed 30th August 2014.

Gilmore, A. (2009) 'Using online corpora to develop students' writing skills' , ELT Journal 63(4): 363-372

Gollin, J. (1998) 'Deductive vs. Inductive Language Learning' , *ELT Journal* 52(1): 88-89

Granger, S. (2003) 'The international corpus of learner English: A new source for foreign language learning and teaching and second language acquisition research' , *TESOL Quarterly* 37: 538–545

Hamp-Lyons, L and Heasley, B. (2006) *Study Writing : A course in writing skills for academic purposes*. Cambridge: Cambridge University Press.

Hanks, P.(ed) (2009) *COBUILD Dictionary*. Glasgow: Harper Collins.

Harmer, J. (2001) *The practice of language teaching*. Harlow: Longman

Hartshorn, J; Evans, N; Merrill, P; Sudweeks,R; Strong-Krause, D and Anderson, N. (2010) 'Effects of Dynamic Corrective Feedback on ESL Writing Accuracy' , *TESOL Quartely* 44 (1): 85-109

Hedge, T. (2000) *Teaching and Learning in the Language Classroom*. Oxford. OUP

Hewings, M. (2012) *Cambridge Academic English: Upper-intermediate*. Cambridge: Cambridge University Press.

Hewings, M. and Thaine, C. (2012) *Cambridge Academic English: Advanced.* Cambridge: Cambridge University Press.

HM Government. (July 2013). *International Education: Global Growth and Prosperity [online]* Available at www.ukgov.bis. Last accessed October 20th 2013

Ho, J and Crookall, D. (1995) ' Breaking with Chinese cultural traditions: Learner autonomy in English language teaching', *System* 23(2): 235-44

Hoffan, S and Evert, S. (2006) *BNCweb (CQP-edition) Version 4.2* [online]. Available at http://corpora.lancs.ac.uk/BNCweb. Last accessed 4th October 2014.

Hoffman, S., Evert, S., Smith, N., Lee, D., and Bergland Prytz, Y. (2008) *Corpus Linguistics with BNCweb: a practical guide*. Frankfurt am Main: Peter Lang

Holec, H. (1981) *Autonomy in Foreign Language Learning*. Oxford: Pergammon.

Honey, P. and Mumford, A (1992) *The Manual of Learning Styles*. Maidenhead: Peter Honey.

Hunston, S. (2002) *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Huttner, J. (2010) 'The potential of purpose-built corpora in the analysis of student academic writing in English' , *Journal of Writing Research* 2(2) :197-218

Hyland, K. (2000) *Disciplinary Discourses.* Harlow: Longman

Hyland, K. (2002) 'Specificity Revisited: How Far Should We Go Now?' *English for Specific Purposes* 21: 385-395

Hyland, K. (2003) *Second Language Writing*. Cambridge: Cambridge University Press

Hyland, K. (a) (2009) *Academic Discourse*. London: Continuum International Publishing Group

Hyland, K. (b) (2009) 'Constructing proximity: Relating to readers in popular and professional science' , *Journal of English for Academic Purposes* 9 (2):116-127

Hyland, K. and Hamp-Lyons, L. (2002) 'EAP: Issues and Directions', *Journal of English for Academic Purposes* 1:1-12.

Hyland, K and Tse, P. (2007) ' Is there an "academic vocabulary"?' , *TESOL Quarterly* 41(2): 235-253

Hunt, K. (1965) *Grammatical Structures Written at Three Grade Levels*. Champaign: National Council of Teachers of English.

IBM Corp. (2012) *IBM SPSS Statistics for Windows, Version 21.0*. Armonk, NY: IBM Corp.

Inoue, C. (2010) 'Investigating the sensitivity of the measures of fluency, accuracy, complexity and idea units with a narrative task', *Papers from the Lancaster University Postgraduate Conference in Linguistics & Language Teaching* 4: 1-24

International Corpus of English (ICE). (n.d) [online]. Available at http://ice-corpora.net/ICE/INDEX.HTM. Last accessed 14th October 2014.

Jarvis, H. (2004) 'Investigating the classroom applications of computers on EFL courses at Higher Education Institutions in UK' , *Journal of English for Academic Purposes* 3: 111–137

Johns, T. (1991) 'Should you be persuaded: Two samples of data-driven learning materials' , *ELR Journal* 4: 1–16.

Jones, J. (1995) 'Self Access and Culture' *ELT Journal* 49(3): 228-34

Jordan, R. R. (1997) *English for academic purposes: A guide and resource book for teachers.* Cambridge: Cambridge University Press.

Jordan, R.R. (1999) *Academic Writing Course* Harlow: Pearson Longman.

Jordan, R.R. (2002) ‘ The Growth of EAP in Britain’, *Journal of English for Academic Purposes* 1: 69-78.

Kenning, M.M. (1996) ‘IT and autonomy’ in E. Broady & M-M Kenning (eds) *Promoting Learner Autonomy in University Language Teaching*, pp121-138. London: Association for French Language Studies / CILT

Korhonen, L; Laakkonen, I; Schneider, B; and Van Camp, R. (2007) *Corpus Project* [online] Available at https://kielikompassi.jyu.fi/kookit06/corpus/project.html Last accessed 16th September 2014.

Krishnamurthy, R. and Kosem, I. (2007) ‘ Issues in creating a corpus for EAP pedagogy and research’, *Journal of English for Academic Purposes* 6: 356–373

Lee, D. and Swales, J. (2006) ‘A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora’ , *English for Specific Purposes* 25: 56–75

Leech, G. (1997) ‘Teaching and language corpora: a convergence’ , in Wichmann, A; Fligelstone,S ; McEnery, T. and Knowles, G. (eds) *Teaching and Language Corpora*. Harlow: Longman.

Li, S. L., & Pemberton, R. (1994) ‘An investigation of students’ knowledge of academic and subtechnical vocabulary ‘ , *Proceedings of the Joint Seminar on Corpus Linguistics and*

Lexicology, pp. 183–196. Hong Kong SAR, China: Hong Kong University of Science and Technology.

Little, D. (1990) 'Autonomy in Language Learning' in Gathercole, I. (ed.) *Autonomy in Language Learning, pp.7-15*. London: CILT.

Little, D. (1991) *Learner Autonomy. 1: Definitions, Issues and Problems*. Dublin: Authentik

Lowie, W and Seton, B. (2013) *Essential Statistics for Applied Linguistics*. Basingstoke: Palgrave Macmillan

Makino, T. (1993) 'Learner self-correction in EFL written compositions', *ELT Journal* 47(4): 337-341

Mauranen, A. (2004)' Speech Corpora in the classroom' in Aston, G., Bernardini, S. and Stewart, D. (eds) *Corpora and Language Learners* , pp. 195-211. Amsterdam: John Benjamins Publishing Company.

McCarthy, M. (2007) *Accessing and interpreting corpus information in the teacher education context*. Revised version of a plenary paper presented at the QuiTE [Quality in Teacher Education] Annual Conference, London.

Mcarthy, M and O'Dell, F (2008) *Academic Vocabulary in Use*. Cambridge: Cambridge University Press.

Meyer, L. (1996) 'The contribution of genre theory to theme based EAP: Navigating foreign fjords' , *TESL Canada Journal 13(2):* 33-45

Nisbet, D; Tindall, E. and Arroyo, A. (2005) ' Language Learning Strategies and English Proficiency of Chinese University Subjects' , *Foreign Language Annals* 38(1):100 -107

Nunan, D. (1998) ' Teaching Grammar in Context' , *ELT Journal* 52(2): 101-109

O'Sullivan, I., & Chambers, A. (2006) ' Learners' writing skills in French: corpus consultation and learner evaluation ' , *Journal of Second Language Writing* 15(1): 49–68.

Perez-Paredes, P. and Cantos-Gomez, P. (2004) ' Some lessons students learn: self-discovery and corpora' , In Aston, G. Bernardini, S. and Stewart, D. (eds) *Corpora and Language Learners* , pp. 247 – 260. Amsterdam: John Benjamins Publishing Company

Perez-Paredesa,P., Sanchez-Tornelb,M., Alcaraz Caleroc, J.M., and Aguado Jimenez,p. (2011) 'Tracking learners' actual uses of corpora: guided vs non-guided corpus consultation' , *Computer Assisted Language Learning* 24(3): 233–253

Pierson, H.D. (1996) 'Learner culture and learner autonomy in the Hong Kong Chinese context' in R.Pemberton et al (eds) *Taking Control: Autonomy in Language Learning* , pp 49- 58. Hong Kong: Hong Kong University Press

Polio, C. (1997) 'Measures of linguistic accuracy in second language writing research' , *Language Learning.* 47(1):101-143

Polio, C., Fleck, C. and Leder, N. (1998) ' "If I Only Had More Time": ESL learners' changes in linguistic accuracy on essay revisions' , *Journal of Second Language Writing, 7(1)*: 43-68.

Poole, J. (2006) 'E-learning and learning styles: students' reactions to web-based Language and Style at Blackpool and the Fylde College ' , *Language and Literature* 15(3): 307-320

Renouf, A. (1997) 'Teaching corpus linguistics to teachers of English' in Wichmann,A., Fligelstone, S. McEnery, T. and Knowles,G. (Eds.), *Teaching and Language Corpora*. Harlow:Longman

Romer, U. (2004) 'Comparing real and ideal language learner input: The use of an EFL textbook corpus in corpus linguistics and language teaching' in Aston, G. Bernardini, S. and Stewart, D. (eds) *Corpora and Language Learners* , pp151 - 168. Amsterdam: John Benjamins Publishing Company.

Scott, M., (2013) WordSmith Tools.[online] Available at ***www.lexically.net/wordsmith*** Liverpool: Lexical Analysis Software. Last accessed 2nd April 2012.

Sheen,Y. (2007) 'The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles' , *TESOL Quarterly* 41:255–283

Schmitt, D and Schmitt, N. (2011) *Focus on Vocabulary 2. Mastering the Academic Wordlist*. New York: Pearson Longman

Sripicharn, P. (2004) 'Examining Native Speakers' and learners' investigation of the same concordance data and its implications for classroom concordancing with ELF learners' in Aston, G. Bernardini, S. and Stewart, D. (eds.) *Corpora and Language Learners* , pp. 233-247. Amsterdam: John Benjamins Publishing Company.

Stephenson, P. (2011) *Intellitext* [online] Web-based corpus interface, University of Leeds. Available at http://corpus.leeds.ac.uk/it/. Last accessed 4th October 2014.

Stevens, V.(1995) ' Concordancing with language learners: why? when? what?' , *CAELL Journal* 6(2):2-10

Storch, N and Tapper, J. (2009) 'The Impact of an EAP Course on Postgraduate Writing' , *Journal of English for Academic Purpose*s 8:207 - 223

Swales, J. M. (2002) 'Integrated and fragmented worlds: EAP materials and corpus linguistics' in Flowerdew, J. (ed.), *Academic Discourse* , pp. 150–164. Harlow: Longman.

Swales, J.M. (2005) *Genre Analysis. 2nd edn.* Cambridge: Cambridge University Press.

Swan, M..(2001) *Learner English.* Cambridge: Cambridge University Press.

Sun, Y.-C. (2007) ' Learner perceptions of a concordancing tool for academic writing' , *Computer Assisted Language Learning* 20(4):323–343

Tan, M. (2005) 'Authentic language or language errors? Lessons from a learner corpus' , ELT Journal Volume 59(2):126 – 134

The Regent of the University of Michigan. (2009) *Michigan Corpus of Upper-Level Student Papers (MICUSP)* [online] Available at micase.elicorpora.info. Last accessed 14th October, 2014

Thornbury, S. (1997) *About Language*. Cambridge: Cambridge University Press

Thurstun, J., & Candlin, C. N. (1998) 'Concordancing and the teaching of the vocabulary of academic English', *English for Specific Purposes* 17:267–280.

Truscott, J. (1996) 'The case against grammar correction in L2 writing classes ' , *Language Learning*, 46:327–369

Turnbull, J., & Burston, J. (1998) 'Towards independent concordance work for students: lessons from a case study' , *On-Call 12(2*):10–21.

Turner, J. (2004) 'Language as Academic Purpose ' , *Journal of English for Academic Purposes 3:* 95-109

Urdan, T.C. (2005) *Statistics in Plain English.* New Jersey: Lawrence Erlbaum Associates, Inc.

Watson Todd, R. (2001) ' Induction from self-selected concordances and self-correction' , *System 29:* 91–102

Watson Todd, R. (2003) 'EAP or TEAP?' , *Journal of English for Academic Purposes 2:*147-156

Widdowson, H.G. (1983) *Learning Purpose and Language Use.* Oxford: Oxford University Press.

Widdowson, H.G. (2000) 'On the limitations of linguistics applied' , *Applied Linguistics* 21(1): 3-25

Witalisz, E. (2007) 'Vocabulary Assessment in Writing: Lexical Statistics' in Lengyal, Z and Navracsics, J. (eds) *Second Language Lexical Processes: Applied Linguistic and Psycholinguistic Perspectives.* Clevedon: Multinlingual Matters. P101-116

Wolfe-Quintero, K; Inagaki, S and Kim, H-Y. (1998) *Second Language Development in Writing: Measures of Fluency, Accuracy and Complexity.* Honolulu. University of Hawai'I Press.

Yang, N.D. (1998) ' Exploring a new role for teachers: promoting learner autonomy' , System 26(1): 127-135

Yates, R. and Kenkel, J. (2002) ' Responding to sentence-level errors in writing' , *Journal of Second Language Writing* 11:29-47

Yoon, C. (2011) 'Concordancing in L2 writing class: An overview of research and issues' , *Journal of English for Academic Purposes* 10:130–139

Yoon, H. (2008) 'More than a linguistic reference: The influence of corpus technology on L2 academic writing' , *Language learning and Technology* 12(2):31-48

Yoon, H. and Hirvela, A. (2004) ' ESL student attitudes toward corpus use in L2 writing' , *Journal of Second Language Writing* 13: 257–283