‘TAKE HIM TO THE CLEANERS AND MAKE HIM DO YOUR HOMEWORK’: A CORPUS-BASED ANALYSIS OF LEXICAL STRUCTURES USED BY ENGLISH LANGUAGE LEARNERS

By

JUSTIN JAMES MCNAMARA, B.A., M.A.

Submitted to the University of Limerick for the degree of Doctor of Philosophy

Supervisors: Dr Anne O’Keeffe

Prof. Michael McCarthy

Submitted to the University of Limerick (June) 2020

For my beautiful Mam and Dad. You were so proud of this and you could not wait for me to finish it and to tell everyone. Well Mama and Papa, WE did it!!! You both taught me so much, but in particular the fact that I could achieve anything that I set my mind to, what you didn't realise is that without you I would have achieved nothing. You took my world with you and I will love you forever. Until we meet again, I hope you're dancing in the sky.

Abstract

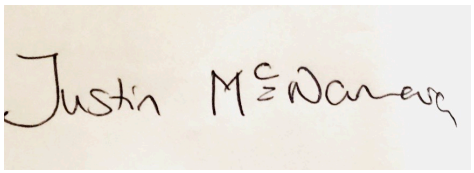
The present study is an empirical corpus based analysis of the use of four lexical bundles or strings by ESL students at a higher education centre in Ireland. The overall aim was to ascertain if students at both ends of the language learning spectrum used the following multi-word items in their speaking and writing: 1) Multi-word verbs 2) Delexical verbs 3) Collocations and 4) Idiomatic expressions. There are two levels of learners who took part in this study: A2 and C1. The learner's use of language was analysed over a period of twelve weeks. Recorded interactions and oral presentations in class were analysed as well as written homework and assignments. Integral to this study is corpus linguistics and the researcher's created Adult Corpus of English (ACE). This corpus-based methodology enabled the identification of the frequency and number of the four lexical strings used by the language learners. Overall, twenty four students agreed to take part in the research and as a result the corpus amounts to 170,000 words: 20,000 written and 150,000 spoken. The use of WordSmith Tools (2016) and manual sifting of the corpus identified that both cohorts clearly use the four lexical strings in their speaking and writing. Multi-word verbs such as *come back and put in*, delexical verbs such as *make an effort* and *do your homework*, collocations for example *spend time* and *write a letter* and idiomatic expressions such as *the grass is always greener* and *black and blue* were recorded, identified and tagged. It is argued that though the classroom is not the most natural of contexts the majority of language used is produced by the learner without prompting or explicit teaching. Overall, the C1 cohort was found to use the majority of opaque lexical structures while the A2 cohort used less and transparent strings.

Acknowledgements

- First and foremost, I would like to thank my supervisors, Prof. Michael McCarthy and Dr Anne O'Keeffe, for their immense help, encouragement, support and fundamentally, for believing in me throughout this journey. I will be forever grateful for each and every supervision and guidance over the years, and I am now even more thankful and proud to have you both as mentors and friends.
- To my Examiners: Prof. Ivor Timmis and Dr. Joan O'Sullivan, thanks for your interest, encouragement and great feedback on this thesis. I learnt so very much from you.
- To Dr Ann Higgins of MIC: you have been present and instrumental at so many pinnacle moments in my life, from educational to personal. You have played such a vital role in my education, from primary school teacher to principal and finally as faith would have it Viva Chair, I am forever grateful. However, I would also like to thank you for your friendship with my dearest Grandmother and Mother. I am proud to have you in my family.
- To my family: thanks for your continued love and support throughout every single day of my life and for making me believe that I can achieve anything. Thank you for your support and love, for putting up with me, and more importantly for not mentioning the 'T' word.
- To my friends: Niamh Morrison and Sharon Counihan, aka the Heavies, you have always been there with me through thick and thin, and that is not only during PhD times. Together, you have made the worst, roughest and toughest time of my life somewhat bearable.
- To my dearest MIC and PhD buddy Margaret Healy: you have become not just a colleague or friend but also a member of my family, thank you for your help, encouragement and love.
- To Dr James Binchy of MIC: I am will forever be grateful to you for your help in finalising this thesis: many, many thanks.
- Finally, thanks to my students, colleagues and friends at Limerick College of Further Education, Mary Immaculate College, Limerick Institute of Technology, The University of Limerick, and IVACS for making this research possible. I will be forever in debt to you all.

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.



5th June, 2020

Signed: _____ Date: _____

Justin McNamara

Table of Contents

1. Introduction **16**

1.0 Rationale for the present study 17

1.1 Locating the present study 21

1.2 Research questions 26

1.3 Overview of the thesis 30

2. Review of the literature: Vocabulary teaching and learning in a historical perspective **34**

2.0 Introduction 35

2.1 Earlier methodologies 36

2.2 The Grammar Translation Method 38

2.3 The Reform Movement 39

2.4 The Direct Method 42

2.5 The Reading Method/Situational Language Teaching 42

2.6 The Audiolingual Method 45

2.7 Some reassessments of the role of vocabulary 45

2.8 Second Language Acquisition Theory (SLA) and the role of vocabulary in language teaching 47

2.9 Idiomatic competence and performance within SLA 48

2.10 Communicative Language Teaching 50

2.11 The Lexical Approach of the 1990s 53

2.12 Corpus linguistics and the vocabulary control movement 57

2.13 The Newfound Importance of Learner Vocabulary Size 59

2.14 Measuring learner vocabulary 61

2.15 What constitutes a word? 65

2.16 English as an International Language (EIL) 67

2.17 English as a Lingua Franca (ELF) 69

2.18 The Common European Framework of Referencing for languages (CEFR) 71

2.19 Conclusion 74

Chapter 3 Conceptual framework: Vocabulary description 76

3.0 Introduction 77

3.1 The 'word' debate: multi-word verbs, delexical verbs, collocations and idiomatic expressions 78

3.2 Multi-word verbs 79

3.3 Delexical verbs 80

3.4 Collocations 85

3.5 Idiomatic expressions and figurative language 87

3.6 Conclusion 92

Chapter 4 Methodology 94

4.0 Introduction 95

4.1 Corpus linguistic: origin and definition 95

4.2 The Adult Corpus of English (ACE corpus) 98

4.3 Methodological framework 107

4.4 Wordlists 108

4.5 Keyword lists 110

4.6 Concordance lines 112

4.7 Normalisation 114

4.8 Cluster analysis 114

4.9 'In search of the unknown:' The observer paradox 119

4.10 Analytical framework 121

4.11 Conclusion 123

Chapter 5 Multi-Word Verbs **124**

5.0 Introduction 125

5.1 Multi-Word Verbs description 125

5.2 Identifying and classifying types of multi-word verbs 126

5.3 Corpus analysis of multi-word verbs: the BNC as a reference corpus 127

5.4 Analysis: Multi-Word Verbs in the ACE corpus 129

5.5 Register and one word substitution 152

5.6 Conclusion 154

Chapter 6 Delexical verbs **158**

6.0 Introduction 159

6.1 Defining delexical verbs 159

6.2 General features of delexical verbs 161

6.3 Corpus analysis of delexcial verbs the BNC as a reference corpus 164

6.4 A closer examination of students' delexical verb use 181

6.5 Conclusion 190

Chapter 7 Collocations **193**

7.0 Introduction 194

7.1 Four definitions of collocation 195

7.2 Collocation overview 196

7.3 The types of collocations 197

7.4 Data analysis and the ACE corpus 199

7.5 Collocations in the written component of ACE 209

7.6 Case studies 215

7.7 Conclusion 218

Chapter 8 Idiomatic Expressions **221**

8.0 Introduction 222

8.1 An anecdote from the ACE corpus 223

8.2 Fixed Expressions and Identifying an Idiomatic Phrase 225

8.3 Characteristics of Idioms 227

8.4 How many idioms are there in English? 229

8.5 Forms and Types of Idioms 231

8.6 Data Analysis of the idiomatic expressions in ACE 233

8.7 Analysis of the functions of idioms in the ACE corpus 251

8.8 Conclusion 266

Chapter 9 Conclusion **269**

9.0 The ACE corpus and results 270

9.1 The four lexical strings 272

9.2 Returning to the Research Questions 276

9.3 Relating the findings to theory and practice 282

9.4 Limitations of this research 284

9.5 Suggestions for Further Research 286

10 Bibliography **290**

11 Appendices **315**

List of tables

Table 2.1 Vocabulary size of students of English

Table 2.2 The total number of phrases known by learners of each CEFR level in the English Vocabulary Profile (EVP)

Table 2.3 The number of multi-word verbs known across the CEFR levels 63

Table 3.1 Learner Dictionaries and examples of Delexical Verbs

Table 3.2 Types of Fixed Expressions and Idioms

Table 4.1 the four datasets of the ACE corpus (rounded numbers)

Table 4.2 Spread data, by learner and by mode (spoken and written) in the sample period in the classroom in ACE

Table 4.3 Comparison of the top 20 most frequent words of the ACE and BNC corpora

Table 4.4 ACE corpus Keyword List using the BNC as a reference corpus

Table 4.5 the 10 most frequent two-word clusters in the ACE corpus (per million words)

Table 4.6 10 most frequent three-word clusters in the ACE corpus (per million words)

Table 4.7 Top 10 four-word clusters in the ACE corpus (per million words)

Table 5.1 Raw and PMW frequencies of Top 20 Multi-Word Verbs in the BNC

Table 5.2 The number of individual learners using multi-word verbs in the ACE corpus

Table 5.3: Breakdown of total number of occurrences of MWVs across levels and spoken versus written data (raw and PMW)

Table 5.4: MWVs broken down by type and by level

Table 5.5: Top five most frequent MWV forms across levels (raw frequencies).

Table 5.6 Occurrence of multi-word verbs in ACE, plus the CEFR level at which they appear in the English Vocabulary Profile (Capel, 2012).

Table 5.7 the top twenty most frequent multi-word verbs in the ACE spoken corpus

Table 5.8 *Come back* across both cohorts

Table 5.9 The twenty most frequent multi-word verbs in the ACE written corpus

Table 5.10: Summary of the EVP profiles of the MWVs used by learners in their writing

Table 5.11 Multi-word verbs uttered in ACE corpus and their possible one word equivalent

Table 6.1: Five patterns with delexical verbs

Table 6.2 Examples of delexical verbs and their single-word equivalent

Table 6.3 High frequency verbs in the BNC and ACE corpora

Table 6.4 Number and percentage of students using delexical verbs across both cohorts

Table 6.5 Frequencies (raw and normalised) of delexical verbs in A2 and C2 data, across spoken and written data.

Table 6.6 Delexical verbs total frequencies (raw and PMWs) across speaking and writing in ACE

Table 6.7 Most frequent (> 4 occurrences) of delexical verbs patterns in the Spoken ACE, by level compared with the EVP expected level.

Table 6.8 Examples of delexical patterns with single occurrence in Spoken ACE, by level and EVP

Table 6.9 Comparison of the most frequent delexical patterns (> 4) in Spoken ACE with BNC

Table 6.10 Most frequent (> 4 occurrences) of delexical verbs patterns in the Written ACE, by level compared with the EVP expected level.

Table 6.11 Comparison of the most frequent delexical patterns (> 4) in Written ACE with BNC

Table 6.12 Student attempts and errors at delexicality

Table 6.13 Student <$15> (A2 cohort)

Table 6.14 Student <$17> (A2 cohort)

Table 6.15 Student <$4> (A2 cohort)

Table 6.16 Student <$5> (C1 cohort)

Table 6.17 Student <$24> (C1 cohort)

Table 7.1 Number and percentage of students using collocations across both cohorts

Table 7.2 Overall frequency of collocations in ACE by level (raw and PMW)

Table 7.3 Frequency of collocations by level in spoken and written ACE sub-corpora

Table 7.4 The 20 most frequent collocations in the ACE spoken corpus compared with the BNC and their EVP level

Table 7.5 The 20 most frequent collocations in the ACE written corpus

Table 8.1 The percentage of Idioms in English (adapted from Schmitt, 2013).

Table 8.2 Types of Idioms (Carter, 1987).

Table 8.3 Number and percentage of students using idioms across both cohorts

Table 8.4 Breakdown of frequencies of idiomatic language in ACE speaking and writing, by level and percentage.

Table 8.5 Distribution of idiomatic expression by form, PMW frequency and sub-corpus

Table 8.6 Types of idiomatic expression, by PMW frequency and level

Table 8.7 Top 20 idiomatic expressions in the ACE data, by level, compared with BNC and EVP.

Table 8.8 Functions of idioms in the ACE Corpus, by level across speaking and writing

Table 8.9 Colloquial Irish expressions in the Spoken ACE corpus.

Table 9.1 Summary of percentage use of the four lexical string features

Table 9.2 Summary of total frequencies of four features, by level, sub-corpus

Table 9.3 Raw number of occurrences by feature type, by level (based on Fig. 9.3)

Table 9.4 PMW frequency comparisons across four lexical string features

List of Figures

Fig. 2.1 Sweet's reform movement curriculum

Fig. 3.1 The compositionality of idioms

Fig 4.1 Screenshot of Wordlist in ACE corpus, alphabetically ordered**.**

Fig. 4.2 Concordance list for the word *I* in the ACE corpus

Fig. 4.3 Concordance lines for the top key word, *am,* the in the ACE corpus

Fig. 5.1 Multi-word verbs use by individual learners in the ACE Corpus across

A2 and C1 learners by percentage, across spoken and written sub-corpora.

Fig. 5.2 Breakdown of all uses of MWVs across three types

Fig. 5.3 The five most frequent multi-word verbs in the ACE corpus (%)

Fig. 5.4 Concordance screenshot sample of *come back* (A2)

Fig. 5.5 Concordance screenshot of sample *go* + particle in C1 spoken sub-corpus

Fig. 6.1 The cline of idiomaticity of delexical structures

Fig. 7.1 Concordance for *of course*

Fig. 7.2 Concordance list for *thank you*

Fig. 7.3 Concordance list for thank you

Fig. 7.4 Concordance list for a few

Fig 8.1 Structure of a communication according to Malinowski

Fig 8.2 Types of idioms used in the corpus (Raw Frequency)

Fig 9.1 The division of the ACE corpus

Fig. 9.2 Summary of average number of students using four lexical string features, by level

Fig. 9.3 Summary of raw number of occurrences by feature type, by level

Fig. 9.2 Collocations in the ACE corpus

List of Appendices

Appendix 1: Specific learner outcomes for FETAC A2 316

Appendix 2: Specific learner outcomes for FETAC C1 322

Appendix 3: Metadata for all participants 337

Appendix 4: Letter of Consent 342

Appendix 5: CEFR level descriptors and IELTS Band 8 and 9 Descriptors 346

Appendix 6: Transcription conventions 350

Appendix 6: Writing task samples 359

Chapter 1 Introduction

It has been estimated that the 850 words of Basic English have 12,425 meanings (Paul Nation, 1983: 11).

1.0 Rationale for the present study

Two past and present scenarios have led to my interest in the topic of Applied Corpus Linguistics: firstly, being an English as a Second Language teacher for seventeen years and secondly my own personal experience of learning German as a second language.

- During my experience of studying German at both secondary school and university, I developed an interest in vocabulary and how one acquires new words in another language. Starting as a real beginner, I was fascinated with increasing my vocabulary size, acquiring fluency and being able to appropriately use and to retain newly learned German words.
- As Firth (1935: 37) claims, 'words do not merely occur alone', and much recent research (reviewed and discussed in subsequent chapters) points to the central role of multi-word strings of various kinds in the creation of meaning, lexical storage and fluent production. Since words do not occur alone in native speaker usage, it was felt that it would be interesting to try to discover how words are used in various types of combinations in the data of the learners and to ascertain what lexical patterns and combinations English as a Foreign Language (hereafter, EFL) learners were able to use. This led to the multi-word framework for the main analytical chapters of this thesis.
- The Common European Framework of Referencing for Languages (hereafter, CEFR) (Council of Europe (COE), 2008) clearly outlines the language exponents an EFL learner should know at various levels. The CEFR and its 'can do' statements is readily available, easily accessible and

used worldwide as a graduated reference point for proficiency markers in over 30 languages. Despite the international application of the CEFR descriptors, it is not without criticism (see below). Specific to this research which is set in the context of English as a Second Language Learner, (hereafter, ESL, learners), it is lacking as there remains a clear and rather large gap in the literature. Currently, there is no framework available that outlines the language features of an ESL Learner specifically. This was an integral incentive for this research underpinned by the belief that ESL learners' competency markers are not identical to those of EFL. Specifically relating to how this research in terms of vocabulary learning, the CEFR states that items such as: 1) Delexical Verbs 2) Multi-Word Verbs 3) Collocations and 4) Idiomatic Expressions (the lexical bundles that make up the data analysis chapters of this thesis) are components of the lexicon of the advanced level learner and therefore an indication of proficiency in English. However, based on experience over the years in ESL classrooms, I was personally intrigued to investigate my intuition that these items are present (and used to varying degrees of competence) at lower levels. Therefore, this study came about through my genuine practice-based curiosity as to whether lower-level ESL learners could productively use the aforementioned features associated with more advanced learners within the CEFR.

- Schmitt (2013) highlighted the difference between vocabulary size and vocabulary depth and identified how it was just as important to know how to use a word, as it was to know its meaning. According to Schmitt (2013) and McCarthy (2006), one of the key features of vocabulary is the collocations that accompany a unit in language in use. Based on Schmitt's (2013)

classification, I feel that this research will provide valuable data of an ESL learner's depth of knowledge.

- The English Vocabulary Profile (Capel, 2007) and The English Grammar Profile (Mark and O'Keeffe, 2010) further inspired this research. They approach the question of language competency form the point of view of what a learner actually *can* do with the language rather than what they *cannot* do. These two projects and their products provided a major reason for this thesis since I am analysing both the successful and attempted (with errors) use of the four lexical features in order to ascertain what an ESL learner at both ends of the spectrum can actually do. Successful use means without a single error but I decided to include attempts with errors as it shows that the student is trying to use the features (albeit with errors and mistakes). The findings of the occurrences in this research will be compared to that of the English Vocabulary Profile and the English Grammar Profile in order to make an accurate comparison between language learners.
- It is further interesting to note that the college at which the research was conducted follows a syllabus developed by the Further Education Training and Awards Council of Ireland (FETAC). Delexicality, multi-word verbs, idiomatic and figurative language and collocations do not feature anywhere in the language module descriptors (See Appendixes 1 and 2). Therefore, it was felt that it would be interesting to determine whether the students could use these features of language without being taught them explicitly[1].

[1] A methodological note that will be discussed further in the analysis: it is noted that the researcher conducted one lesson on phrasal verbs and idioms. Be that as it may, the majority of phrasal verbs and idioms were elicited from the learners. In addition, both cohorts had lessons on the topics of asking for and giving directions which resulted in exponents such as *take a left* and *go straight on* etc.

- Another rationale for the study is to counterbalance the generic nature of the CEFR through a localised study in an ESL context. Localisation means that real learners with real profiles can be examined. This means that the collection of learner metadata could be used to add depth to the contexts of learner language use[2] (see Appendix 3). These questions also served to provide the researcher with vital information about the learners such as their daily exposure to the language (e.g. by knowing a learners' length of stay or hobbies, one can explain the use of certain language).
- The rationale for the study is underpinned by using real learners and real language use. Therefore a corpus-based approach was considered the best means of empirically analysing their data. The Adult Corpus of English (ACE) was therefore compiled (see Chapter 4). A corpus-based approach is an efficient means of arriving at both a frequency counts and recurring patterns. It also allows for the use of contextual evidence when examining the learners' vocabulary use (e.g. who is the learner?, what is their background?, what is the classroom activity or task that generated this language use?). The corpus-based findings generated from this corpus provide specific insights, both quantitative and qualitative, and allow for an overall discussion of the language produced by these learners.
- Finally, it was felt that a comparative analysis between the ACE corpus and a larger well baseline native-speaker corpus would provide a benchmark for a more grounded evaluation of the learners' proficiency and the robustness or otherwise of the CEFR levels. Such a comparison also provides valuable results in terms of comparing an ESL speaker to a native speaker from a

[2] The following metadata is available on all learners: 1) full name (anonymised) 2) gender 3) date of birth 4) nationality 5) length of stay in Ireland 5) occupation 6) hobbies.

developmental (rather than binary) perspective. As Chapter 4 will detail, the British National Corpus (BNC) was chosen as the benchmark against which spoken and written learner language could be compared where relevant within the analysis[3].

In summary, the rationale for this thesis is a personal one, from the point of view of a language teacher wishing to empirically analyse a specific group of ESL learners' use of and competence in formulaic language. It is driven by a hypothesis that the CEFR descriptors for key aspects of formulaic language use for EFL are not in line with the reality of ESL learners attempted and sometimes correct use of these items at lower levels that are set out in the CEFR.

1.1 Locating the present study

A Hallidayan (1990) view of language as words, meanings, systems and functions has been the driving force behind this study. Halliday (1990) and Sinclair (1996, 2000) take a Neo-Firthian (Firth, 1957) stance on language and both propagate Firths' earlier view that 'words do not occur alone' (Firth, 1957: 204), but rather, words occur in the company of other words and as a result, meaning is not always identified by one individual unit. Sinclair's (1996) *Model of Extended Units* clearly identifies meaning as being evident from the combination of individual words. According to Sinclair (1996, 1998), understanding the meaning of a word is not just about grasping a dictionary definition. As an alternative, to really grasp a word, we must look at the broader picture and identify the collocational patterns of the word. Those who hold a Hallidayan

[3] As detailed in Chapter 4: The British National Corpus (BNC) is a contemporary 100 million word collection of spoken (10%) and written (90%) data from native speakers of British English from a wide range of sources. The written section is drawn from newspapers, periodicals and journals from a wide variety of sources while the spoken section contains informal conversations, business meetings and radio chat shows.

position see language as composed of groups of words or phrases which often function with one unitary meaning.

Since the advent of corpora as a means of statistically measuring language in use, the Hallidayan view of language has become ever more influential. Now more than ever, corpora can show us the language patterns and the frequencies of such patterns in everyday language use. Frequently used features of language are today measured with ease thanks to corpora. In terms of advancing our understanding of vocabulary, through corpus analysis, Sinclair (1991), Carter and McCarthy (1997) and Wolter (2002) (among others) have identified four lexical clusters as frequent features of native speaker discourse: multi-word verbs; delexical verbs; collocations; and idiomatic expressions.

This study is situated methodologically as an empirical study that is informed by corpus data. Most of all it is informed by the learners that generated the data in the corpus. The researcher is a teacher seeking to answer a real-classroom question from his practice. To this end, he has built a corpus so that he could address the question of how ESL learners used formulaic language features in both their spoken and written English. Essentially, this is a learner corpus study, within a specific case study localised to a specific group of ESL learners. Mainstream learner corpus studies are situated almost entirely in the context of advanced university-level writing contexts (see overview of learner corpus research by Granger, Gilquin and Meunier, 2015). Most studies of learner language focus on writing rather than speaking (McEnery, Brezina, Gablasova and Banerjee, 2019). Most studies of learner language are synchronic across a large sample of individual writing responses to specified tasks because largescale longitudinal learner data is challenging to collect (Meunier, 2015). Diachronic learner corpora can only be undertaken as case studies simply because of the near impossibility of tracking the same

(large) cohort of learners across all levels of their learning from Beginner to Proficiency level. As Meunier (2015) notes, cross-sectional studies allow researchers to look at points along the learning pathway, e.g. a particular year of study, or CEFR level of competency. This study takes spoken and written data from two cohorts of learners over a 12-week period. This study, therefore, can be described as one that gathers cross-sectional data longitudinally within a case study of two ESL learner cohorts. Students were recorded speaking in class over the course of a 12 week programme for a total of 40 hours. This is combined with their written language to form the 170,000-word ACE corpus (See Chapter 4, for further details).

The learners who participated in this study come from a range of backgrounds. They are all non-native English students at an adult education college in Southern Ireland. These learners come from countries including Spain, Italy, France, Germany, Poland (see Chapter 4). All participants took English classes with me nice hours per week, over two academic semesters. At the time of research, the students had been living in Ireland for varying durations, for example, one student (<$24> from the C1 group) moved to Ireland 11 years prior to the time this research was conducted while student #4 had been living in Ireland for 8 weeks at the time of research. The college is a private institution where students pay to attend a course. The language courses at the college are part time, two evenings per week and attracted highly motivated adult learners many of whom were working full-time in the city and self-funding their English course[4].

As discussed, these students were ESL rather than EFL learners and the only language used in the classroom was the target language: English. The difference between EFL and ESL students is fundamental to this study as the CEFR is based on EFL learners and there is no currently available framework of the ESL learner. Furthermore, the fact

[4] Taking into the account the fact that students paid for the course partly explains a high motivation. It is also noted that there was a high attendance at each class; students rarely missed a session.

that these students were living and working in Ireland meant that the language they used primarily on a day-to-day basis was English to communicate socially and in the workplace. As we shall discuss further in chapter 4, the participants were placed by CEFR level using the in-house ESOL placement test. The two levels in the study are: the Elementary cohort (A2) and advanced (C1).

As already mentioned above, central to the research is the CEFR and the level descriptors devised by the COE (1991). The CEFR provides a framework for identifying what an EFL learner can do at each level of the spectrum. However, and fundamental to this study is the notion that a foreign language learner is not the same as a second language learner so the CEFR (based for the most part on a generic notion of a foreign language learner), while useful has limitations for the ESL context because in some respects, it cannot accommodate the differing learning context. ESL learners have more every day learning experiences especially through the medium of spoken language. As already mentioned, the students taking part in this study are living in Ireland and are therefore English as a second language (ESL) learners. Currently, there is no extant framework comparable to the CEFR to determine what an ESL learner can and cannot do with the language. So, it was considered interesting to focus on ESL learners' use of certain lexical features within a corpus-based case study over a period of time.

The primary aim is to illuminate aspects of ESL students' use of the vocabulary. Specifically, this study focuses on multi-word items in their speaking and writing. It is therefore interested in use of vocabulary beyond the single-word items. Its focus is on 1) Multi-word verbs 2) Delexical verbs 3) Collocations and 4) Idiomatic expressions. The rationale for the choice of these items is based on the high frequency of occurrence of these features in native speaker discourse (Carter and McCarthy 2001; 2006) and the

fact that they are linked to advanced competence in the CEFR Descriptors (Council of Europe, 2008). These four types of lexical patterns will form the analytical lens for this study (chapters 5 to 8). Furthermore, these features were chosen based on the complex nature and character of the notion of 'word' (as outlined in Chapter 2), since each feature fits within the confines of the notion of a multi-word unit. For the purposes of the present study, the term *lexical item* will be favoured over the term *'word'. Lexical item* will include single and multi-word strings across the four types of word-string chosen for analysis.

Another theoretical underpinning for looking at patterned use of language is the growing understanding that second language acquisition (SLA) conforms to a usage-based model of learning within SLA studies (Tomasello, 2003; Ellis & Larsen-Freeman, 2006; Ortega, 2013). As Ellis & Larsen-Freeman (2006: 577) notes it is'part of a communicatively-rich human social environment'. For ESL learners, much of their knowledge of language comes from experiencing and using in the second language environment that they live and work in. Within the usage-based model, it is held that through exposure and use, the mind makes sense of the frequencies of forms found in the language experience, as well as the meanings to which these forms are put (Pérez-Paredes, Mark and O'Keeffe, In press). A central idea in the usage-based model is the idea that, language learners learn the patterns that they repeatedly hear and understand them in use. Though the current study is an empirical one, the usage-based theory is one that we will return to in the concluding chapter to reflect on whether the findings from this study of ESL learners are in line with this emerging model of second language learning.

Though theoretically unrelated, the notion of idiomatic competence Liontas (2015) also underpins this study. Being able to use multi-word units, idioms and other frequently

used formulaic units in a target language, as we shall discuss further in chapter 2 (see 2.9) needs to be taken into account when we consider communicative competencies. This is important to the context of this study because of its focus on ESL learners. According to Liontas (2015) becoming idiomatically competent within the target *culture* is part of the process of Idiomatiziation (see chapter 2, section 2.9).

1.2 Research Questions

In addition to theoretical contributions that the findings of this study might bring, it also has practical and applied outputs. Firstly, there are numerous analyses and examples of research based on intermediate level English language students, but there is relatively little available in regard to the lower and higher levels of competence (The Centre of Literacy, 2008). While empirical, corpus-based work on grammar and vocabulary is available across the six levels of the CEFR through the *English Grammar Profile* (EGP) (O'Keeffe and Mark, 2017) and the *English Vocabulary Profile* (EVP)(Capel, 2012), this is based on written learner data from the Cambridge exam suite. The EGP and the EVP will prove useful as reference points in this study, the results from this study will add some empirical insight into the profile of A2 and C1 ESL learners, from both spoken and written data collected within the same group of learners over a learning period. Research on the lower A2 learners also remains minimal in regard to the preoccupations of the present thesis, namely the degree to which and nature of the acquisition of multi-word units. Consequently, it was thought that authentic data and an original corpus compiled of recordings of ESL learners from both sides of the spectrum (beginner and advanced) would benefit on-going research and further enable the evaluation of existing research. Also, the continuously-developing field of corpus linguistics emphasises data gathered from real-life events and situations. In this respect,

the ACE corpus contains genuine and naturalistic speech used in an ESL classroom setting. In existing research on spoken learner language, the majority of the language produced by the learners is derived through elicitation tasks, such as discourse completion tasks and roleplays (see O'Keeffe, Clancy and Adolphs, 2020). As mentioned above, most learner corpora comprise written data in the context of university learners, based on once-off samples (Granger, et al, 2015). In the present study, learner data comes from natural classroom activity, including ad hoc discussions, small talk with the teacher, as well as output from structured tasks and communicative activities and simulated scenarios in the real world. As a communicative classroom, the language used by the students was often spontaneous and everyday requests and interactions which would appear outside of the classroom frequently arose (see Walsh, 2012). As discussed previously, the research questions that motivates this study comes from my experience as a teacher. They are not a theoretically-derived though they can be placed within theories (as discussed above and in chapters 2 and 3). The questions centre on the observation that ESL learners seem to be using multi-word units even at lower levels and this goes against received expectations within the CEFR and within syllabi. In fact, very often, these items do not even appear explicitly on syllabi. Thus, the following are the core research questions of the present study (these are discussed in-depth in Chapter 4):

- Research Question:

  To what degree can adult learners of ESL use multi-word lexical items (multi-word verbs, delexical verbs, collocations and idiomatic/figurative language) in their speaking and writing?

  Sub-Questions:

i) What is the difference between the lexical competence of the A2 level student and the C1 level student?

ii) Which level (A2 or C1) use the majority of multi-word units and language strings in their speaking and writing?

iii) Is there a progression from more high-frequency transparent strings towards low frequency opaque strings?

The answers to the above research questions will contribute to theory and practice in the following ways:

1) findings from this micro-study of ESL learners over a period of time in the classroom will underscore the importance of the work of Sinclair (1991) and Halliday (1991) (discussed above) which views language as composed of groups of words or phrases functioning with one unitary meaning, through recurring collocational patterns.

2) within Applied Linguistics, this study will add to other corpus-based studies that advance our understanding of patterned vocabulary acquisition in the classroom (Carter and McCarthy, 1997; Wolter, 2002, among many others) across: multi-word verbs; delexical verbs; collocations; and idiomatic expressions. While grammars have described these features based on nature speaker use (e.g. Carter and McCarthy, 2006), this study will do so in the micro-context of ESL learners in the classroom.

3) in terms of learner corpus studies, this research will help to offset the lack of case study diachronic data (as identified by Granger et al, 2015; McEnery et al, 2019). It will use data that is not drawn synchronically from advanced university-level learners (Meunier, 2015). It will look at both spoken and written data from two cohorts of learners over a 12-week period, thus gathering cross-sectional data longitudinally over 12 weeks, within a case study of two ESL learner cohorts.

4) theoretically, as discussed above, this study may inform the growing understanding that second language acquisition (SLA) conforms to a usage-based model of learning within SLA studies (Ellis and Larsen-Freeman, 2006) and it will do so by using language data that is communicatively-rich within the classroom but also rich from the perspective of experience. The learners in this study live their lives in an English-speaking environment and culture. Thus the patterns that they frequently experience should form part of their multi-word patterned output. The patterns that they use, in other words, will mirror their language experience as immigrants living and working in Ireland.

5) from a practical and professional perspective, as previously mentioned accurate knowledge and use of the lexical strings: multi-word verbs, delexical verbs, collocations and idioms are seen as exponents used by higher level learners and native speakers of English. The CEFR places them at B2 and above while the IELTS examination reference them in the descriptors of the 6.5 band and above. If the learners of the ACE corpus use such features in their speaking and writing then it could be argued that levels as low as A2 actually do use these four features productively.

6) Furthermore, another practical outcome will be to identify the difference between the lexical competences across cohorts at different levels of competence in relation to the four lexical strings types. This could enhance our profile of what to expect of learners in terms of their use of these strings in their speaking and writing. Ultimately, this inform ESL syllabi and materials design, especially for lower levels.

7) another empirical contribution relates to idiomaticity in that this study explores the degree to which learners can use opaque language as well as transparent idiomatic. It may expose an under expectation about A2 use of opaque forms.

8) Finally, at a broader level, this study both informs and is informed by aspects of language variation. Within the analysis, there is comparison between the learner data and the British National Corpus (BNC). This will highlight convergence and variation in the language used by ESL learners living in Ireland compared to the British English used in the BNC.

1.3 Overview of the Thesis

Chapter 2: Review of Literature: Vocabulary Teaching and Learning in a Historical Perspective

In this chapter, the literature on language teaching and learning is reviewed. This review is chronological; beginning with English language teaching methods prior to the 1800's and ending with current and contemporary methods. Both the role and significance of lexis in the various language teaching methodologies is analysed. The chapter also focuses on the advent of corpus linguistics and the implications of this approach for vocabulary teaching and learning.

Chapter 3: Conceptual Chapter: Vocabulary Description

Following the Literature Review (Chapter 2) which surveyed the available literature and identified, through time, vocabulary as a neglected component of English language learning, this chapter identifies and argues for the importance of lexis in language learning today. Chapter 3 focuses on the new-found interest in vocabulary in the more contemporary classroom and highlights the acquisition of multi-word lexical items as an integral part of the language learning process. It will be shown that the four data analysis topics discussed in this research are frequent features of native speaker

discourse and therefore are central for the language learner. This chapter will describe the forms, features and functions of the four multi-word lexical items.

Chapter 4: Methodology

In this chapter the compilation process of the ACE corpus is detailed. The section will focus on the manner in which both the spoken and written data were gathered. The chapter also describes Wordsmith Tools (Scott, 2016) as the primary software used in the present study to analyse corpora. The chapter will also highlight (through examples from the ACE corpus) the benefits to research of tools such as wordlists, keyword lists, cluster lists, concordance lines, normalisation and lemmatisation. Finally, the chapter identifies limitations and problems encountered during the implementation of the methodology in the present study.

Chapter 5: Multi-Word Verbs

This chapter deals with multi-word verbs in terms of definition, form and function as manifested in native speaker discourse. The multi-word verbs used both in the speaking and writing of the learners of the ACE corpus are examined to ascertain (based on the research questions) whether the learners of the ACE corpus use multi-word verbs in their speaking and writing and if so which level uses the majority. It analyses in detail the variant structures of multi-word verbs along with their syntactic features. Bolinger's (1971) extensive work on the syntactic elements of multi-word verbs is also referenced.

Chapter 6: Delexical Verbs

This chapter examines delexical verbs in reference to definitions in the literature before providing the current researcher's own definition. The chapter analyses in terms of form and function the delexical verbs found in ACE the corpus and answers the research questions to what degree the learners use the lexical features and which level uses the majority.

Chapter 7: Collocations

The main aim of this chapter is to identify what if any transparent and idiomatic collocations the A2 and C1 students of the ACE corpus use in their speaking and writing in order to answer the research question: to what degree do learners of the corpus use collocations and which level uses the most. This chapter reviews the work of Firth (1957) known widely as the 'father of collocation' and overviews and defines the lexical feature of collocation. This chapter also identifies the use of collocations by the native speaker of English.

Chapter 8: Idiomatic Expressions

This chapter defines, identifies and characterises idiomaticity as a category for analysis. It looks in detail at the syntactic, semantic and pragmatic features of idiomatic expressions. The primary aim of the chapter is to analyse by form, category, and function the idioms used by the two cohorts of learners of the ACE corpus and to see what level uses the most idioms. Sinclair's (1991) Open Choice and Idiom Principle is outlined as an aid in interpreting extended strings in the usage of the native speaker before the chapter looks at the interpretation process and how this is manifested in language learners.

Chapter 9: Conclusion

This chapter presents the various conclusions based on the learners' use of the four core features that complete the data analysis. It identifies that, despite the four lexical features being regular features of native speaker discourse and indications of an advanced level of language proficiency, learners do use them to an extent even at the lower A2 level. This provides valuable insights into the language levels for teachers,

learners and researchers, who previously may have neglected these lexical items until learners are at a higher level. The chapter also returns to the original research questions to evaluate the responses provided by this thesis, identifies the limitations of the present study and outlines recommendations for further research.

Chapter 2 Review of Literature: Vocabulary teaching and learning in a historical perspective

> Though a linguist should pride himself to have all the Tongues that Babel cleft the world into, yet, if he have not studied the solid things in them as well as the Words and Lexicons, he were nothing so much to be esteem'd a learned man, as a Yeoman or Tradesman competently wise in his mother Dialect only.
> (Milton, 1644)
>
> The human mind cannot help but make meaning.
> (Walshe, 1988 cited in Rignall, 2016).

2.0 Introduction

This chapter will review the research relevant to the topic of vocabulary, both in terms of linguistic description and language pedagogy. The changing discussions of vocabulary over time, in particular with reference to English language teaching, will be surveyed. First and foremost, vocabulary was neglected, hidden and in some cases almost non-existent in the language classroom up to the 1940s. Described as the *Cinderella* of language teaching (Meara, 1980), the teaching of lexis remained for much time relegated to a secondary position in favour of the retention of accurate grammatical structures. Many have explained the reasoning behind this relegation and this chapter will examine the rationale of various scholars who examine the ideologies at work in the language classroom of the time.

Publications focusing on the various past and current approaches to English language teaching and, pivotal to this study, the role played by vocabulary in the language classroom, will be further reviewed. The position of teaching and learning vocabulary in the following English language teaching methods will be investigated: 1) The Grammar Translation Method 2) The Reform

Movement 3) The Direct Method 4) The Reading Method or Situational Language Teaching 5) The Audio Lingual Method 6) Communicative Language Teaching 7) The Natural Approach, and 8) The Lexical Approach and of utmost importance to this study, the role of Corpus Linguistics in vocabulary teaching will also be surveyed. It is the aim of this chapter to overview the place of vocabulary as a neglected aspect of language teaching in the past and to show that it was not until the vocabulary revival of the 1980s that it found a more central role in the English language classroom. The research published on the vocabulary revival will be highlighted and the impact the emergence of Corpus Linguistics had on language study will be examined. The ever-growing field of Corpus Linguistics is central to this research and its growth as a tool to underpin vocabulary learning in contemporary language classrooms will also be examined.

2.1 Earlier Methodologies

2.1.1 English Language Teaching prior to 1800.

Towards the end of the middle ages, English replaced French as the second language of England (Howatt, 2004). Previously, it was claimed that the people of the time in fact knew very little French and Latin taught through English became the norm (see John of Trevisa, 1385). Predicting a need and demand for the learning of English as a foreign language, William Bullokar's *Pamphlet of Grammar* in 1586, was one of the first scholarly descriptions of the language. The English language that appeared in language teaching materials from then on was based mainly on texts and sample dialogues, very similar to the existing methods of teaching Latin. Drilling, memorization of phrases, and phrasebooks was the norm for any student learning the language in the early 16th century. Students were frequently tasked with creating their own texts modelled on and often copied from great authors of classical antiquity. Reformulating, paraphrasing and summarizing such classical 'greats' provided syllabi for the English language classroom (Ascham, 1870). Neither grammar nor vocabulary were explicitly taught in the classroom, a method reflected in the 'anti-grammar' of Joseph Webbe who stated 'no man can run speedily to the mark of language that is schackled and

ingiv'd with grammar precepts' (Webbe, 1620: 9). Webbe's notion was repeated by the reform leader Sweet (1899) and again by the Direct Method teachers of the 20th century (see section 2.5).

The latter part of the 17th century saw the introduction of works such as Guy Meige's *New Method of Learning English* (1685) and *Nouvelle Methode* (1685). Despite the success of such publications and the first monolingual French and English dictionaries published at this time, English language courses continued to focus on translation, memorising of dialogues, with minimal teaching of vocabulary.

The predominance of grammar over vocabulary was to continue long into the 18th century. A firm example of the materials of the time is Lowth's (1762) *A short Introduction to English grammar, with critical notes,* which remains an influence in English education today (Aitchison 2001). Lowth (1762) aimed to provide a standard of English grammar and was somewhat commercially successful with numerous copies sold throughout Britain, the USA and Germany. Murray (1795) and Cobbett (1819) were to follow in the footsteps of Lowth and develop his ideas with Murray earning the title of 'father of English grammar'. Murray (1795) used activities and practice grammar exercises identical to those used in the Grammar Translation period (see section 2.3). While Cobbett (1819) saw literary writings as an ideal model of language containing concrete and correct sentence structures, he claimed that as humans our actions come from our thoughts and in order to communicate thoughts we use words, and these words are structured through the correct use of grammar.

English continued to spread world-wide from 1600, but the methodology for English language teaching remained static. Howatt (2004) described this as 'remarkable', he noted the language was being spoken 'in every port from Gibraltor to Fiji' (2004: 127). Apodictically, English was by no means the world's most widely-spoken language but it was fast becoming the most frequently used. At this time, varieties of the language such as Indian and American English and Caribbean English emerged (Howatt, 2004; Harmer, 2015). In Europe, English became a specialised subject in schools. In terms of the approach to teaching English in schools, students were taught grammar through Latin paradigms. As a result, Latin's status had still not waned within education, so much so that it became the *Lingua Franca* of the university sector (Ljosland, 2011).

The following section will focus on the different methods of teaching English from the 18th century up to the present day, and how these relate to vocabulary pedagogy. It must be noted however, that many of these methods existed simultaneously and are seen as counter methods by the researcher. As can be seen, the English language teaching methods explored here did not occur in a linear or chronological manner but rather were developed during similar periods. This section will examine the role and place of vocabulary in each method and highlight how the learning of words has emerged from the linguistic darkness and is now an important component of language learning.

2.2 The Grammar Translation Method

The late 18th and early 19th centuries continued to see the priority given to grammar and the learning of language structures as seen in the earlier 17th century. A new method of language teaching began in Prussia at the end of the 18th century and was self-explanatorily titled The Grammar Translation Method (hereafter GTM). Language students of this era focused primarily on grammatical structures and translation activities. Grammar was taught in isolation; translation was central, both in teaching and in student practice exercises. The aim was that language would be made easier to learn in comparison to earlier approaches (see Mart, 2013; Austin, 2003; McCarthy, 1988). The GTM stressed accuracy and completeness in writing sentences; the spoken language was neglected (Larsen-Freeman, 2000). Lexis taken from the high culture literary texts was selected for memorisation, but only if it illustrated the grammatical rules being taught (Ellis, 1992). Bilingual vocabulary lists were also used as a means of instruction (Richards and Rogers, 2001). Students read and wrote classical materials to aid them in passing standardized examinations. It was assumed that the learners would never actually use the target language, but they would benefit from the mental exercises of drilling, conjugating and translating (Howatt, 2004). They were provided with detailed explanations of the grammar in their native language and they memorized paradigms and bilingual vocabulary lists (Kelly, 1969).

Throughout the 19th century, vocabulary pedagogy was based on etymology and simple definitions of words; bilingual dictionaries became a common tool in the language classroom, but as a reference point (Turner, 2010). Reading was carried out by students in sections, with two or three columns of new vocabulary and mother tongue translations, which were then followed by a test (Willis, 1997). The Grammar Translation method survived well into the 20th century in schools throughout Europe and the USA, but it faced multiple criticisms (Coady and Huckin, 1997). Zimmerman (1997) noted that there was a lack of realistic language and situations in which to use the language. Works such as Rivers (1981) also criticised this method of teaching, claiming that it neglected realistic oral language.

2.3 The Reform Movement

As previously seen in section 2.2, vocabulary had somewhat of a passive function in the GTM method; the lexis practised was taken from memorised dialogues and translations of Latin literary texts. There was no productive approach to vocabulary and as a result it remained secondary (Meara, 1990). Many linguists of the 19th century expressed their opinions on the need for a new method of language teaching. Two such names were Marcel (1937-1896) and Gouin (1831-1896). Though these scholars were virtually unknown at the time, modern researchers have noted the merit of their innovative ideas (see, for example, Carter and McCarthy's 1988: 39 and 48 review of Gouin's work (1892).

Marcel viewed child language learning as a model for second language teaching, a stance that would be echoed by Krashen in 1981 and second language acquisition theory more generally (Krashen, 1987; Crystal, 1997). Perhaps the better-known of the two, Gouin, claimed that the most successful language learning happens through using the target language to accomplish events and actions (Gouin, 1892; Tauber, 2016). For Gouin, topics should be used as a means of organising and presenting oral language (Tauber, 2016). He created the renowned *Gouin Series*, which focused on context that creates clear sentences and actions (Gouin, 1892). It is clear that educators were

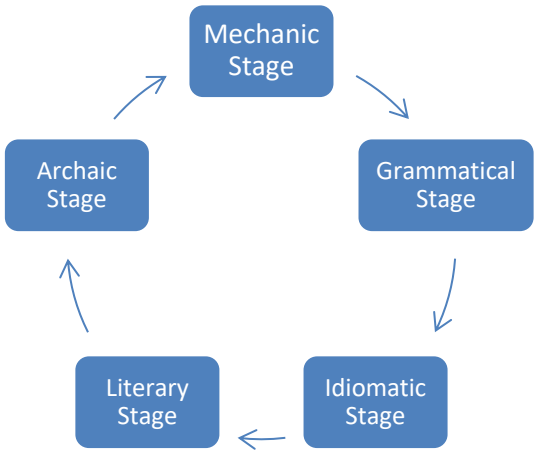
beginning to identify the need for speaking over reading, grammar and working with classical literary texts. Journals and articles were published in which a change in language teaching methods was called for (Vietors, 1882; Regel, 1897; Junker, 1904, for example).

In the 1850s, public examinations controlled by universities were established (Howatt, 2004). The results of students taking such examinations were to lead to determining both the content of the new syllabus and the methods used by the teachers (UCLES, 1858). The founding of both *The Oxford* and *Cambridge Local Examinations* in 1858, which enhanced the status of Modern languages, in particular English, by including them in the curriculum, had a profound effect on language teaching. However, this was to prove both positive and negative for the English language. The English language seemed to lose its academic prestige through its association with the examinations and as a result was 'downgraded' to a 'softer' option course in many academies (Howatt, 2004). This association actually stifled the Reform Movement, but still a need for a monolingual teaching methodology and the use of the target language as the language of the classroom remained, as for example advocated by Sweet (1894).

In 1899 Sweet, the 'father of the reform', published his classic: *The Practical Study of Languages*; which was followed by Jespersen's 1904, *How to Teach a Foreign Language.* Both authors radically challenged the rigidness of the GTM and argued for a new approach to English language teaching. The main principles of this international Reform Movement were the primacy of speech over the other language skills. Oral production and communication were now to play the most vital role in language learning with the introduction of spoken activities focused on accurate pronunciation.

Be that as it may, the learning of vocabulary remained relegated to a lesser status but this time in favour of a focus on phonetics and transcription (Espinosa, 2003). Simple and useful words were taught at different levels while the isolated word lists of the past were abolished and replaced by contextualised vocabulary. Fig 2.1 below illustrates Sweet's (1889) system of language teaching during the Reform Movement:

Fig 2.1-Sweet's Reform Movement Curriculum (1889)



Sweet's (1889) new curriculum began with the **Mechanical Stage**: here students studied the phonetic alphabet and transcriptions of the relevant sounds. Next, is the **Grammatical Stage,** where grammatical structures and very basic vocabulary were taught. The third stage, the **Idiomatic Stage,** is where learners explored vocabulary in greater depth; they would now acquire some 'vital' phrases. Only after thorough analysis of complex texts were the **Literary** and **Archaic** Stages addressed; this was usually reserved until university level where students studied philology (Zimmerman, 1997). Clearly, this new curriculum, with its idiomatic stage marked an improvement in the status of vocabulary. Words were now associated with reality, rather than archaic usage and syntactic patterns. Lexis was selected according to usefulness and simplicity. Sweet (1889) viewed practical vocabulary in relation to the environment and incorporated everyday household objects into the target lexicon. For Sweet (1889), the learning of 3,000 common words was seen as an adequate target. He claimed that we do not speak in words but in sentences and that these sentences are the 'smallest unit of language'. Furthermore, Sweet (1889) discussed the possibility of the development of lists of

vocabulary based on statistical measures. With the Reform Movement a need for change was identified and that need led ultimately and indirectly to the development of more communicative methods of language teaching, the first being the Direct Method.

2.4 The Direct Method

Sauveur (1875) led the thinking which created the Direct Method. He observed that interaction and conversation were at the heart of native speaker language and proposed that this should be reflected in the language classroom. With Sauveur's (1875) influence, the needs of students instead of the teacher were to play a greater role in language teaching. The Direct Method eliminated direct translation and instead conducted lessons via the target language (Richards and Rogers, 1986). Core verbs and common adjectives became regular fixtures in the Direct Method classroom (ibid). Everyday vocabulary was taught without translation (ibid). Moreover, concrete, simple and familiar lexis was explained, while abstract vocabulary was taught through semantic associations (ibid).

While Sauveur (1875) developed the concept of the Direct Method, it was the private language school chain initiated by (and named after) Berlitz (1903) which made it famous. In fact, without Berlitz, the Direct Method would have been virtually unknown (Marcinik, Marc and Martinez, 2010). Berlitz language schools' methodology placed a strong emphasis on oral production, student interaction, avoidance of overt grammar explanation and the development of question and answer techniques. The importance of the role of vocabulary in language teaching grew from this period. The Direct Method influenced other subsequent, related approaches in terms of ethos, such as the Reading Method in the US or Situational Language Teaching in the UK

2.5 The Reading Method/ Situational Language Teaching

From the 1920s onwards, the English language teaching profession accelerated quickly (Smith, 1999; 2003; 2005b). However, the private school sector, such as Berlitz (1903), remained dominated by the Direct Method and an over-reliance on the teaching of phonetics (Richards and Rogers, 1986; Fotos, 2005). In response, notable figures such as Jones (1918) and Palmer (1925), for example,

promoted new ideas. One such idea focused on the importance of vocabulary with the production of the *Carnegie word list* (West, Palmer and Faucette 1934). This was a 2000-word general list, which highlighted commonly used English vocabulary. Its creation led to a boost in the teaching of English as a foreign or second language. The publication of the *Carnegie word list* contributed greatly to the growing interest in vocabulary and resulted in a project which Howatt (2004) describes as 'brilliant': Ogden's (1934) *Basic English.*

Ogden aimed in this basic list to create a complete language in 850 words and a few grammar rules (Ogden, 1934; Templer, 2017). Notwithstanding, the prospect of learning a complete language in such little time appealed to many and *Basic English* continued to be referred to long after its effective demise (see Carter, 1998: 23-25). Simultaneously, a new method of language instruction emerged: **The Reading Method** in the USA and **Situational Language Teaching** in the UK. This new method focused primarily on reading skills, and was in part a response to the 1929 *Coleman Report* (Rivers, 1981), which evidenced serious inabilities in the reading skills of American students. Reading now became predominant in American classrooms. Coincidentally, the same idea appeared in the UK. Michael West (1930) stressed the need for improved reading ability in students and identified how an improvement in vocabulary skills would lead to better reading ability. West (1930) accentuated the view that the primary goal of learning a language is the acquisition of the target vocabulary and the practice of using such vocabulary. Be that as it may, the textbooks available at the time largely ignored this (West, 1930: 514). West (1930) noticed how classroom activities that spent time on non–speaking activities and a lack of useful vocabulary practice resulted in foreign language learners not even having a basic vocabulary of 1000 words. West (1930) continued to recommend the importance of vocabulary for language learners and endorsed the use of Thorndike's word frequency list (1932). He claimed that such lists served an authentic scientific basis for the vocabulary taught in the language classroom. Consequently, in 1953 he published *A General Service List of English Words.* Although now dated, it remains one of the most widely used word lists available. Here, and for the first time vocabulary and the learning thereof was considered an integral component of

language learning. West's *General Service List of English Words* (1953) later informed Coxhead's (2000) *Academic Word List*.

West (1953) defended the priority of reading over the other skills, claiming that the skill leads to the receptive knowledge of up to 60,000 words (West, 1953). West's language tasks consisted of simplifying vocabulary from literary texts by replacing the words with modern equivalents; this task became known as the **lexical selection principle**. For West, there were three stages of Reading Strategy:

1) **Vocabulary stage**: up to a maximum of 1,500 words were introduced.

2) **Development of Skills:** here the focus was on the skills, while vocabulary remained constant.

3) **Strategic Reading:** skimming and scanning.

(West, 1953)

As a result of figures such as Hornby (1898-1978), Gatenby (1944), Morris (1945) and French (1948-50) embracing this new method, vocabulary came further to the forefront. Hornby, Gatenby and Wakefield published the *Dictionary of Current English* in 1952, which led to a new commitment in the teaching of the English language. Fries (1952) and Hornby (1950) are today remembered for their involvement in the creation of the Situational/Reading Approach, which also became known as the Oral Approach. This new focus on spoken English and the control of vocabulary was never entirely abandoned and continued to be used throughout the English language teaching profession (Richards and Rogers 1986).

Situational language Teaching used a structural syllabus but also brought lists of vocabulary to the classroom. However, vocabulary was only chosen if it contributed to the attainment of grammatical structures (Frisby, 1957). Drilling continued to play a vital role, particularly in the earlier stages when learners listened and repeated word for word what the teacher said. This was followed by answering questions and commands that ultimately would lead to a freer production of the learned language structures. The approach attracted criticism as learners had no control over the learning content; they simply listened and repeated (Coady, 1995). The method was criticised as failing to

account for the creativity of language and the uniqueness of every individual utterance (see Camiss, 2012).

2.6 The Audio Lingual Method

The Audio Lingual Method was mainly a reaction to the lack of speaking skills produced by the Reading Approach. The new method, which also became known as the Army Method because of its advocacy by the military (Larsen-Freeman, 2000), was based on the behaviourists' view that any human being could be trained through reinforcement and habit (Skinner, 1959). Like the earlier Direct Method (see section 2.5), under this approach, students were taught the language directly without the use of the L1. Teachers would present the correct grammatical model of a structure and the students would repeat it over and over again (in a repetition drill) until it was correct. It reflected Skinner's verbal behaviourism theory (1959) as positive and negative reinforcement were used in relation to the students' answers. Fries (1945) and Bloomfield (1942) viewed this behaviourist form of teaching as a core component of second language learning. They saw repetition drilling as the principal means of language acquisition. The hope was that students would learn the language communicatively and be equipped with the necessary topics to engage in a conversation. This method led to a further increase in focus on the teaching of vocabulary. The contrastive analysis work of Lado (1957), in particular, meant that the target language (L2) and the learners' first language were contrasted across phonology, vocabulary and grammar so as to identify similarities (positive transfer) and differences (negative transfer) (Lado 1957). In terms of the place of vocabulary, this meant that a body of work emerged comparing English lexical items with other languages and this included the identification of false cognates which were likely to cause negative transfer. These items were part of the key vocabulary items systematically included in the Audio Lingual syllabus (Lado 1957).

2.7 Some reassessments of the role of vocabulary

Throughout the 1940s, 1950s and 1960s, vocabulary retained its secondary status in the teaching of foreign languages. According to Fries (1945), learners needed only a basic vocabulary that would aid

them in producing the correct language structures, a view that was echoed by other American teachers and linguists. Rivers (1968, 1972 and 1983) highlighted reasons for the salience of language structure rather than vocabulary:

1) It is almost impossible to predict what vocabulary a student might need in their daily lives.
2) Too much vocabulary focus leads learners to feel that learning a language is merely accumulating words.
3) It is difficult for learners to retain lists of vocabulary and meanings.
4) Acquisition of the mother tongue succeeds with the knowledge of a limited number of vocabulary items until structures are first mastered.

(Rivers, 1983: 4)

McCarthy, writing in 1984, claimed that this view was still alive at the time of writing in many ESL course books (McCarthy 1984). Interestingly, this was a period of great demand for English language learning but the methods discussed above produced speakers with a very limited vocabulary. Meara (1980) referred around this time to vocabulary as occupying a "Cinderella" status in language teaching, while Carter and McCarthy (1988) saw vocabulary as being a problem of grading and selection which had not been properly addressed. The dominance of structure began to come under wider criticism. However, Wilkins (1972) saw explicit vocabulary teaching as only offering success if the need of the students was not to learn the language quickly and accurately. Though his work began a change in the thinking of language teachers, Wilkins did admit that teaching the structures first and then teaching vocabulary second was a 'sound approach'. Nonetheless, vocabulary, Wilkins (1972) felt, should not be completely ignored. In fact, he propagated that 'while without grammar little can be conveyed, without vocabulary nothing can be conveyed' (1972: 111), a view that was finally beginning to be seriously considered.

Nation and Coady (1988), Twaddell (1972 and 1973) and Bright and McGregor (1970) all argued for the teaching of vocabulary learning strategies rather than the explicit teaching of words. They felt it

was virtually impossible to teach any student all the vocabulary they would need to know. Clearly, the role of vocabulary needed to change. As Lord (1974) echoing Meara (1980), emphasised, vocabulary should not remain as the *Cinderella* of language teaching, an outcast in pedagogical terms, driven by fear of an unlimited list of words to be acquired. Instead, Lord (ibid.) and Richards (1976) argued that vocabulary should focus on the needs of the learner rather than attempting to encompass the whole native speaker lexicon.

Thus far, we have looked at the place of vocabulary over the years in different approaches to teaching English. We have also summarised some of the reasons for the neglect of vocabulary in language teaching and looked at the works of authors who pleaded for change. Now we will look at the growth of Second Language Acquisition theory (SLA) in the 1980s and 1990s and how it affected the place of vocabulary.

2.8 Second Language Acquisition theory (SLA) and the role of vocabulary in language teaching

As a reaction to the behaviourist paradigm, which underpinned the Audio-Lingual Approach, and the related Contrastive Analysis work (Lado 1957), a more cognitive paradigm of learning emerged in the 60s. Within this view, error analysis was seen as a clear window into the learning process and needs of the students (Corder, 1967). This led to many studies of learners in control and test groups. Granger postulated the merits of such language gathering processes later in 2002: 'it is difficult to control the variables that affect learner production in a non-experimental context' (2002: 6). Through such monitoring, learner errors in grammar, pronunciation and to some extent vocabulary were identified. They brought particularly focus to phonology and grammar, for example, identifying some errors as intralinguistics (part of language development) and some as interlinguistic (caused by interference from the learners' L1) (Corder, 1967). This, in turn, informed syllabi and again prioritised areas of grammar and phonology over vocabulary (see also Myles, 2005).

The most influential output of this period came from the work of Selinker (1972), who put forward the notion of *interlanguage* (Selinker, 1972). This was a milestone in SLA theory as it gave learner

language a name and an independent status and placed it not as an imperfect copy of the target language but rather a rule-governed system in its own right (Selinker, 1972). The notion of *fossilisation* of learner errors (ibid) was also core to the concept of interlanguage. These concepts related to lexis, grammar and phonology but most focus was given to fossilisation in grammar and phonology over lexis.

Psychologists such as McLaughlin (1984) and Wells (1986) looked at how mother tongue language, both vocabulary and grammatical structures, are acquired during childhood, and some methods try to emulate this acquisition process in the learning environment of a second language, e.g. see discussions in Krashen (1981), Klein (1986) Ellis (1997) and Nunan (2001). Although the imitation of first language acquisition in second language acquisition has contributed in some part to the interactivity of language classes today, it has also caused problems, and one such problem is in terms of vocabulary and the learning of new lexis. It was assumed by many that the main focus of language should be on grammatical structures and the production of correct, native-sounding sentences. In line with this is a widely-held view within SLA theory, namely the Natural Order Hypothesis, which assumes a natural order of the acquisition of grammatical items (Krashen 1982). This view holds that the acquisition of grammatical structures occurs in a predictable sequence which could not be overridden by pedagogy. Despite some criticism (see Gregg 1984), the natural order hypothesis drove syllabi, thus relegating vocabulary once again to a secondary status. Recent developments in corpus linguistics and the analysis of learner data have led to a questioning of the natural order (Murakami, 2013); however, the hypothesis retains a strong following at the time of writing.

2.9 Idiomatic competence and performance within SLA

What constitutes L1 and L2 language acquisition has been discussed and debated in numerous works over time (see Campbell and Wales, 1970; Ellis, 1985, 1994; Krashen, 1982 and Liontas, 2015). Successful Second Language Acquisition is clearly governed by the notion of Communicative Competence (see section 2.10 below) and the ability to use acquired language in a similar manner to

that of a native speaker so much so that this ability is often viewed as being competent in the second language. According to Canale and Swain (1980) Communicative Competence is divided into four components: 1) Grammatical Competence, 2) Sociolinguistic Competence, 3) Discourse Competence and 4) Strategic Competence. Integral to this study, Liontas (1999) and Liontas (2015) adds Idiomatic Competence to this list and views the mastery of idiomatic language as mastery of the target language and notably the culture. Idiomatic usage is the extent to which a language learner demonstrates their knowledge of idiomaticity, while idiomatic use focuses on the speaker's ability to use idioms for communication in various social settings. Liontas (1999 and 2015) claims that Idiomatic Competence is the ability to understand and use idiomatic language appropriately (Liontas, 2015). To be able to use idiomatic language in a variety of sociocultural contexts in a manner similar to a native speaker is seen as a language skill in itself (see Canale and Swain, 1980 and Bachman, 1990). Be that as it may, Liontas (2015) asserts that idioms, multi-word verbs, and lexical bundles remain underrepresented in the field of SLA theory and research (Liontas, 2015), which is surprising considering their high frequency in native speaker discourse. As mentioned earlier, the use of lexical structures, such as the four highlighted in this study, is frequently equated with knowledge of the culture of the target language. We note that this is important to the context of this study because it focuses on learning English as a second language. According to Liontas (2015) part of the process of Idiomatiziation or becoming idiomatic in a language means being able to function as idiomatically competent within the target culture.

Fundamental to this study is the low representation of studies that examine ESL learners' use of idiomatic language. The majority of research to date does not examine how idioms are learned over time or how learners of different levels of proficiency acquire and use idioms. According to Liontas (2015: 623) 'it is hypothesized that the acquisition of idioms is qualitatively different when idioms are learned in a natural environment than when such learning takes place in a formal educational setting'. We can hypothesise that the natural acquisition outside of the language classroom by ESL learners happens at a subconscious level, somewhat akin to native-like learning in contrast to formally

learned language patterns in the classroom which usually require conscious learning processes (see Han and Finneran, 2013). Of salience also is the point that a number of studies have revealed that the *production* of L2 idiomatic language is more challenging than their *comprehension* or interpretation (see Laufer, 2000; Liao and Fukuya, 2004; Conklin and Schmitt, 2008).

2.10 Communicative Language Teaching (CLT)

In contemporary language classrooms, Communicative Language Teaching (hereafter CLT) appears to be the accepted, or at least desired, norm. Its development began as a reaction to the questioning of the prominence of overt grammar teaching in language acquisition (Richards, 2006). A shift developed, whereby accuracy remained integral but a need for language that could assist in promoting fluency was acknowledged. Moreover, widespread criticism that the classical and traditional methods created language users who were unable to communicate in English was the main driving force behind CLT (Rivers, 1983 and Richards and Rogers, 1990). This method is largely based on the notion of Communicative Competence (Hymes, 1972). Hymes (1972) stressed that communicative competence existed over, and even above, linguistic competence and was a required component of natural communication. Richards and Rogers (1986) described CLT as an approach rather than a method. They saw it is a 'philosophy of teaching' that considers communicative 'ability' as more important than the reproduction of learned grammatical structures. CLT encompasses a functional syllabus where the language exponents taught have a clear function often used to complete a given task; such as *making requests/decisions* or *going to the supermarket*. Learners were given the necessary language and pedagogical support to complete such interactions, task-oriented learning often in pairs or groups became a regular feature of the classroom, and grammar structures for many took second place.

A combination of *Grammatical Competence, Sociolinguistic Competence, Discourse Competence* and *Strategic Competence* were, according to Canale and Swain (1998), the sub-competencies that underlie the overall notion of communicative competence. To this end, syllabi which aided learners

in not only mastering the core semantic notions of a language, but also incorporating functional tasks to achieve communicative goals, were introduced. Such a syllabus is commonly known as the notional-functional syllabus (Wilkins, 1972, Laufer, 1986). In class, students are exposed to native speaker like, situationalised occurrences of the language embedded in the culture of the target language. The learner took part in activities based on themes and situations deemed to be appropriate to their communicative needs and aspirations. The new found interest in authentic language in use and the identification and teaching of the language that the learner would regularly need to communicate further lead to the development of descriptors for language learning which contained both structural and communicative-functional components.

Teaching the four language skills with the primary goal as communication replaced the repetitive drills of the archaic 'army methods'. The new communicative curriculum incorporated authentic language through interactive tasks, which mirrored possible, real life situations and as a result according to Widdowson (1978), fostered fluency rather than accuracy. In CLT, students are encouraged to use the target language from the very beginning, with varied activities and strategies used to aid the learning process. Such activities from the outset promoted learning while the primary function of the task became communication. There was a surge in the publication of communicatively-oriented literature and textbooks in these domains. In terms of vocabulary, in 1978, Judd published *Vocabulary Teaching and TESOL: a need for re-evaluation of existing assumptions*, where he called for the need of vocabulary to become a skill in itself. For Judd (1978), lexis should not be secondary in favour of the acquisition of grammatical structures and neither should it be merely a supplement to reading and listening texts. Judd (1978) advocated for vocabulary to be taught, in context, at the beginning of any English language course. In addition, he claimed that it was not enough to see words as labels for objects and ideas, instead the language class should address every element of the word; from 'meaning', 'pronunciation' and 'spelling' to its behaviour and relationship with 'other words'. According to Carter and McCarthy (1988), Judd's (1978) ideas 'combine the

desire to give vocabulary a proper status and the need to give the learner the breadth of resource that the lexicon can offer' (1988: 46).

The late 1970s saw learners attempting to communicate in the target language with the teacher and each other. No longer did they simply reiterate or "parrot" taught phrases. Instead, they took part in information gap activities, roleplays, simulations, interviews, games and pair work on a regular basis (Harmer, 2007). The idea was that the learners would be equipped with the language necessary to communicate in the English-speaking world (See Simmons, 2010). McCarthy (1990) also highlighted the importance of vocabulary by stating that in order to guarantee communicative competence, learners must have the necessary words to express their desired meanings.

Many teachers of the time began to see vocabulary as a fundamental component of language learning and ultimately viewed the acquisition of words as just as necessary as the acquisition of accurate grammatical structures. McCarthy (1984) pleaded for a reassessment of the status of vocabulary in language teaching and this was reflective of many scholars' plea for change, including Corder (1960), Graver (1963), Barnard (1971, 1975), Land (1975) and Robinson (1977), while figures such as Richards (1976), Judd (1978), Celce Murcia (1979), and McKay (1980) also published articles on the importance of vocabulary to the second language learner. McCarthy (1984) saw the importance of raising awareness of how vocabulary works in authentic communication and identified the ability to use vocabulary effectively in communication as a language skill in itself. In addition, Carter and McCarthy (1986) and Meara (1988) promoted a greater interest in vocabulary in the communicative classroom.

However, it must be noted, that there was not a universal shift to CLT; other language teaching approaches mentioned above were still in use throughout the world and continued to influence each other. Counter to the call for change in the status of vocabulary, publications including Bright and McGregor (1970) Madsen (1983), Weir (1990) Spolsky (1995) and Oller (1979) all continued to

advocate the production of a correct, grammatical sentence as the most important element of language learning.

2.11 The Lexical Approach of the 1990s

The lexical approach to language teaching put lexis at centre-stage, and the lexicon became seen as a crucial element of language use and language learning (Richards and Rodgers, 2001). Vocabulary began to influence not only ideas about language skills but also came to be seen as central to the very nature of structural configurations and communication itself. Several language teaching approaches advocating a new perspective on vocabulary were discussed under headings such as the Lexical Syllabus (Willis, 1990), Lexical Phrases and Language Teaching (Nattinger and Decarrico, 1992), and the Lexical Approach (Lewis, 1993). Carter and McCarthy (1988) had argued: 'Language practitioners need not shy away from lexis as a boundless chaos; organisational principles are available and simply wait to be more fully exploited' (1988: 38) A means of doing this is to organise and grade teaching and learning of the language as used in real life. As a result, studies of native speaker discourse and its vocabulary were more extensively drawn upon by ELT researchers and practitioners than ever before (Long, 1983; Carter, 1987; Carter and McCarthy, 1988; Nation, 1990; Nation and Waring, 1997; Schmitt and McCarthy, 1997).

The identity and nature of multi-word units, language chunks and lexico-grammatical strings of various kinds as a primary feature of native speaker discourse took centre stage in language research and gradually influenced pedagogy. Language chunks, or patterns of groups of words which occur together on a frequent basis, were seen to encompass a large proportion of the language used by the native speaker. The earlier work of Firth in 1957 would now become an invaluable support for language researchers, who could now argue that one does certainly 'know a word by the company it keeps' (Firth, 1957: 11). Nattinger (1980) claimed: 'Perhaps we should base our teaching on the assumption that, for a great deal of the time anyway, language production consists of piecing together

the ready-made units appropriate for a particular situation and that comprehension relies on knowing which of these patterns to predict in these situations' (Nattinger, 1980: 341).

The basic idea behind the Lexical Approach (that would be described in detail by Michael Lewis in 1993) was that a vital part of successfully acquiring and learning a second language was the ability to understand, retain and accurately produce language 'chunks'. Lewis (1993) stated 'language consists of grammaticalised lexis, not lexicalised grammar'. Clearly, according to Lewis, vocabulary was just as important for the language learner as grammar, and the two were in effect inseparable. The grammar-vocabulary dichotomy was invalid, as the English language contains a vast number of recurring lexico-grammatical patterns. Under the Lexical Approach, collocation patterns, colligational configurations and fossilised chunks were all to be integrated into the language syllabus. Lewis and adherents of the Lexical Approach were more concerned with language as being composed of other structural elements and not what was traditionally viewed as grammar. Lewis (1993) commented how many language teachers noticed that high level, advanced students' grammar was practically perfect, but it was their lexis and the ability to accurately 'chunk' words together, into coherent phrases and expressions, that identified them as non-native speakers (Lackman, 2011). Lewis (op. cit.) claimed that it was paramount for ELT teachers to help their students to identify and be aware of the lexical structures that occur in language. If a learner is aware of the collocations, chunks and combinations that make up the language, then that learner will more productively make use of such structures. As Sinclair stated, 'a lexical mistake often causes misunderstanding, while a grammar mistake rarely does' (Sinclair, 1996: 16).

A further dichotomy that had previously existed was the predominance of writing over speech; under this new approach, developing speaking skills and communicative competence took primacy over the ability to write accurately. In the language classroom, students were now encouraged to use the lexico-grammatical resources of the language to their fullest in the service of promoting fluency. Tasks that used authentic materials which encouraged the learner to use relevant vocabulary to complete this task were now the norm (see Martinez, 2002). Grammar remained a component of

language learning but it became subordinate to the challenges of lexis. Increasing the vocabulary size of the learner was now favoured and tests (such as those outlined in 2.12 below) to ascertain the number of 'words' a language learner knows, were used more than ever (see Nation, 2001 for example).

Under the Lexical Approach materials and methods used in the classroom showed a great differentiation through the levels of competence; materials appropriate to elementary level students were 'radically different' (Willis, 1990) to those of the advanced learner. The primary aims of the lexical classroom were that the students' ability to concentrate, appreciate, argue an opinion, and co-operate would increase due to the lexical approach's new found focus on the context and co-text in which the language occurred. Fundamental to the approach and as mentioned earlier, is making the learner aware of the behaviour of the language they are learning; as a result, the study of language chunks, collocations and grammatical colligation all became primary features of the lexically-oriented EFL classroom (see also Sinclair and Renouf, 1988). Courses and teachers began to focus on providing their learners with a large and pragmatically useful number of single- and multi-word lexical items.

However, this is not to say that grammar disappeared entirely in the development of the Lexical Approach; instead it took a secondary position in order of importance and in the majority of major published courses grammar and vocabulary were, and continue to be, taught simultaneously. Grammar retained a function in the syllabus but a reduced one, while lexis acquired an increased and primary focus. The importance of grammatical lists in the classroom continued but they were now accompanied by associated lexical groupings (Nunan, 1989). Willis (1990) identifies the lexical approach as using lists of linguistic exponents, as syllabi have always done, but he states that the lists now being used are drawn from the language which speakers actually use and is therefore of greater relevance than earlier types of lists. Willis (ibid.) claims:

> In order to recreate a picture of the target language…The lexical syllabus does itemise language. It itemises language minutely, resting on a large body of research into natural language. On the basis of this research it makes realistic and economical statements about what is to be learned. But the methodology associated with the lexical approach does not depend on itemisation.
>
> (Willis, 1990: viii).

The lexical approach espoused an analytical syllabus; such a syllabus is different to the synthetic syllabus of the grammar-translation and audio lingual teaching methods. The preceding, syntactic-based syllabi promoted language learning as a gradual process of accumulating each individual component of a structure before the whole structure is acquired. In sharp contrast, the analytical syllabus of the lexical approach, according to Nunan (1989: 28), presents learners with language 'chunks which may include structures of varying difficulty' where the essential element of language learning is 'communicative purpose'. Conditions and authentic situations in which the language is used form part of the materials used in the language classroom; realia such as menus, timetables and surveys brought into the classroom function to merge both the lexical approach and communicative language teaching. The lexical syllabus was comprised of items to be studied on a lexical rather than grammatical basis, such as the delexical constructions of *have*, *get* and *put* (discussed further in chapter 5 below). Increased attention was paid to the base form of frequent lexical verbs, as the lexical approach promotes the need for students to possess a large repertoire of lexical verbs in their base form as the 'addresses' for knowledge of their lexico-grammatical combinations (Carter and McCarthy, 2016). Collocations assumed an important role in the syllabus, while pragmatics and the study of the functions of language in authentic use brought groups of words, lexical linking, relations between synonyms in discourse and the study of idiom and metaphor into the syllabus and materials (McCarthy, 1991).

Swan (1985) points out that in order for the lexical approach to be feasible there must be a balance between the lexis and language functions being used; he states 'functions without lexis are no better than structures without lexis' (Swan, 1985: 39). However, as both communicative language teaching

and the lexical approach apodictically espouse the use of authentic language materials in the classroom, language teachers and materials designers needed a source of such information. Such a primary source would come in the form of insights from corpus linguistics. As McCarthy (2001) points out, 'the language of the corpus is, above all, real, and what is it that all language learners want, other than 'real' contact with the target language' (McCarthy, 2001: 128). Thus, the Lexical Approach and corpus linguistics became inextricably entwined.

2.12 Corpus Linguistics and the Vocabulary Control Movement

Perhaps the earliest record of corpora is the line by line hand written analysis of the Bible by Padua in 1195 (see chapter 5) but it was not until more contemporary times when lexicographers identified the usefulness of corpora analysis for language learning (McCarthy and O'Keeffe, 2010). Quirk's (1960) *Survey of English Usage,* which provided analysis of spoken British English, inspired Kucera and Francis (1961) to create the first modern electronically readable corpus at Brown University, the *Brown Corpus* (1961). This is a one-million-word collection of American English texts which is still in use today. Later, Mifflin and Kucera (1969) published the *American Heritage Dictionary,* the first known dictionary based on corpus findings. Subsequently, the *LOB Corpus of British English* was published in 1970.

Corpus linguistics fast provided innovative resources for language learning and teaching in the late 1990s and early 2000s. The COBUILD dictionary (Sinclair, 1987) is an example of a dictionary driven by corpus findings. In addition, reference grammars emerged, based on corpus analysis (e.g. COBUILD Grammar, 1990, Biber et al, 1999; Carter and McCarthy, 2006). Corpus analysis made it possible to examine real language in use in context and this greatly enhanced vocabulary material. One example of this is the *Vocabulary in Use* series (McCarthy and O'Dell, 1994). As already mentioned, the main attraction for the use of corpora as tools for learning and teaching was, as McCarthy (2001) notes, that the language that is exploited is 'real' and 'authentic'.

Evidence from corpora such as the COBUILD Corpus (Sinclair, 1987) (4.5 billion words) the British National corpus (Oxford University Press, 1991) (110 million words), the Cambridge International Corpus (Cambridge University Press, 2000) (almost 2 billion words) the CANCODE corpus (McCarthy, 1998) (5 million words) and the MICASE corpus (University of Michigan, 1998) (1.7 million words), among many others, have all brought to light patterns of lexical structure predominant in everyday communication and academic spoken language, respectively. Corpus analysis also showed that much of what was perceived earlier as a manifestation of grammar is actually vocabulary at work (Schmitt, 2000). Rather than words simply filling the slots provided by the grammar, they were shown to be powerful drivers of the structural configurations and recurring patterns that surround them. This notion of *lexico-grammar* brought vocabulary into the spotlight within ELT research and pedagogy. However, it was not until the late 20th century that identifiable corpus-informed work began on vocabulary in pedagogy as manifested in publications such as: Sinclair and Renouf (1988), Nation (1990), Oxford (1990), Ellis (1997), Meara (1997), Schmitt and McCarthy (1997), Sinclair (1997), McCarthy (2002), Renouf and Banerjee (2007) and McCarthy, O'Keeffe and Walsh (2010)..

The reliability and originality of corpus findings shifted the pedagogical focus from grammatical to lexical and lexico-grammatical features, promoting vocabulary as a primary element of language use and learning. The rapid development of corpora resulted in one of the most significant advances in vocabulary studies in recent years (Schmitt, 2000) as corpora allowed researchers, learners and teachers to access great amounts of real language, shifting the focus away from invented examples. All major English dictionaries today incorporate evidence from corpora, for example The *Cambridge International Dictionary of English* (1995), *the Longman Dictionary of Contemporary English* (1995), *the Oxford Advanced Learners Dictionary* (1995), and *the Longman Language Activator* (1994). Publicly available software to analyse corpora appeared from the late 1980s and onward, for example *The Oxford Concordance Program* (1987), Scott's Wordsmith Tools (1996), Barlow's *Monoconc* (1996), as well as *Antconc* (Anthony, 2010). Through the use of such software, collecting

and analysing authentic language was now within the reach of non-technical experts, such as syllabus designers, materials writers and teachers and learners in classrooms.

2.13 The Newfound Importance of Learner Vocabulary Size

Approaches to vocabulary and opinions on its importance have radically changed since Meara's (1980) 'Cinderella of language teaching' comparison. Gyllstad (2013: 11) described the newfound interest in vocabulary as a 'formidable explosion in terms of activity in publications'. This 'explosion' in interest in lexis is accompanied by the interest in the knowledge of the language used by the native speaker, by non-native users and by learners and, as already argued in this chapter, corpus linguistics has provided a means to study what these various language users say and write on a regular basis.

Arguably, the most vital information acquired from corpus findings is frequency counts. Frequency counts show how vocabulary is distributed in lay and specialised use. Estimates of the total size of the English word-store vary from 54,000 word families (Nation and Waring, 1997) to millions (Bryson, 1990) and in this seemingly daunting situation, corpora have aided language educators in understanding which word families are in frequent use and therefore most useful to be learned. Crystal (1980) estimated that an educated native speaker can actively use 60,000 words and understand up to 75,000. Nation and Waring (1997) state that a native speaker has a repertoire of 20,000 word families. Meanwhile, O'Keeffe, McCarthy and Carter (2007) note that native speakers survive on a day-to-day basis with a relatively small number of core word forms of around 2,500 items (see also Goulden, Nation and Read, 1990). Mackey (1965) claims that non-academically educated adults typically possess a vocabulary size of 10,000 words, while a professional scientist may have a receptive vocabulary size of around 80,000 words, whereas university students studying English as a second language generally have the potential of achieving a maximum vocabulary size of 10,000 words and typically fall far below this level. Clearly, the estimates of how many words a speaker 'knows' (understands or can use) is continuously debated.

Alongside studies of native-speaker vocabulary, interest in learner's vocabulary size and range likewise developed (Laufer and Nation, 1990; Schmitt, 2000; McCarthy, O'Keeffe and Walsh, 2010). The need to develop learners' vocabulary size has led to the creation of numerous vocabulary size tests, which are used on a daily basis by institutions around the world. Nation and Laufer's *Vocabulary Levels Test* (Nation, 1983; Laufer, 1999; Schmitt et al, 2001) was designed in 1983 to give an estimate of learner vocabulary size for general English and Academic English. Nation (1983) designed this test as a diagnostic test of vocabulary size and levels, for use by students and teachers, with Read (1988) claiming it to be 'reliable'. The test assesses vocabulary sizes from 1,000 to 14,000 words, where takers must choose the correct definition of a word from a multiple choice list of four. For Nation, the ability to use the first 2,000 words in English was necessary to successfully communicate in everyday scenarios. The knowledge of the 3,000-word level was where learners were able to read authentic English texts, and knowledge of the 10,000-word level would give an understanding of specific academic vocabulary (Schmitt, Schmitt and Clapham, 2007). The levels test originates from a number of different frequency levels taken from Lorge and Thorndike's list (1932), and provides a profile of the vocabulary the learner actually knows. Such tests are now increasingly based on modern corpus evidence.

The *Eurocentres Vocabulary Size Test* (Meara and Jones, 1992) was originally designed in 1992 for use with learners of English as a second language as a response to a commission from Eurocentres: a group of language teaching schools in Switzerland. The rapid turnaround of students in such schools created a problem in placing students in correct language levels. Designed to work as a placement test it was to replace the previous *Joint Entrance Test* (JET). It gives an estimate of the overall vocabulary size of learners which is used to place students together with others with the same vocabulary size. It takes ten minutes to complete and it is self-scoring. Learners are presented with a group of words, some genuine and others invented, and students then must claim if they know the words or not (Meara, 1990). As with Nation's (1990) test, the vocabulary questions are divided up into blocks based on the frequency level of the English lexicon from the first 1,000 up to 10,000

words and is based on Single Detection Theory Models of the 1970s (Zimmerman et al, 1977 and Kling and Riggs, 1971). Although criticised for highlighting only passive knowledge of the tested words, both the speed and the correlation of the results with those of the JET examination has led to it remaining as the key vocabulary test for Eurocentres schools.

In 1995, Laufer and Nation created *The Lexical Frequency Profile*, a system used to measure the percentage of words in a text belonging to different levels of frequency. It is designed to identify the number of words a learner needs to know in order to fully understand a text that they are reading, which Laufer (1997) claims to be 95% of all words. The lexical proficiency of the learner would inform the teacher as to whether a text or other material being used in class is appropriate to the students' level. The test divides the analysis into bands. If 95% of the words appearing in a text are part of the 2,000 high-frequency words of the English language, then the text is considered to be of low difficulty, as it is assumed that the majority of language courses would focus on these words. If 95% of the words belong to the first 5,000 words of the language, then the text is considered to be of medium difficulty. Finally, if more than 5% of the words of a text have a frequency under the level of 5,000 it is regarded as a difficult, advanced level text. This test can be accessed by learners and teachers online, where a piece of text can be entered so that the test can automatically analyse its difficulty. All of these developments rest heavily on insights from corpus linguistics.

2.14 Measuring Learner Vocabulary

Learners of a second language frequently equate knowing the language with being able to speak it fluently and the ability to identify words and meanings (Nation and Laufer, 1995). Many students feel that the greater their vocabulary size (in terms of the number of individual words), the more fluent they are (See Nation, 1990).

Table 2.1 below shows the vocabulary size a learner of English should have at each level across the Common European Framework of Reference (CEFR) (See section 2.15 of this chapter for an

explanation of the CEFR). In addition, the table aligns each level CEFR level to an official ESOL examination.

Table 2.1 Vocabulary size of students of English (Milton, personal communication 2011)

Vocabulary size (max 10,000)	**Cambridge**	**Test of English as a Foreign Language (TOEFL)**	**International English Language Testing System (IELTS)**	**CEFR LEVEL**
9000		630	8	
8000/9000	Certificate of Proficiency in English (CPE)	620	7	C2
7000/8000	Certificate in Advanced English (CAE)	600	6.5	C1
6000/7000		550	6	
5500/6000		500	5.5	
4500/5500	First Certificate in English (FCE)	450	5	B2

About 4000	Preliminary English Test (PET)	350-400	4.5	B1
About 3500	Key English Test (KET)	300	4	A2

As can be seen from Table 2.1, students of English at CEFR A2 level should possess approximately 3,500 words and a student at a C2 level should know about 8,000/9,000 words (Milton p.c., 2011). The A2 and C1 levels are the two cohorts examined in the present study. Interestingly, the levels can be mapped to the International English Language System (IELTS) examination at a band four and seven respectively. Meanwhile, it is typically accepted that young native speakers will add around 1,000 word families to their vocabulary size a year as they mature and progress through the education system (Schmitt, 1997; Crystal, 2007; Milton 2011). Moreover, Schmitt (2013) notes that speech is less dense than written discourse so as a result he postulates that if a non-native speaker has a lexicon of between 2,000 and 3,000 word families then that speaker can survive in everyday conversation. On the other hand, in order to read a wide variety of texts, Schmitt (2013) estimates that the same speaker is required to have a vocabulary size of between 8,000 and 9,000 words. In addition, in a paper presented at the TESOL conference in the USA in 2006, McCarthy identified that in order to reach an advanced level of competence, the language learner needs to possess a receptive vocabulary of around 6,000 words. This 6,000 word level McCarthy (2006: 2) refers to as a 'good threshold' at which learners are ready to embark on an advanced level programme'

Adding further to this list, and integral to the present research, is the work of the English Profile project (EP). The project was devised to analyse and identify the language that learners at each level of the CEFR levels actually use (in corpus data). One of the strands of the EP research is the *English Vocabulary Profile* (EVP) (Capel, 2012). This online resource describes both receptive and

productive vocabulary knowledge of learners based on the Cambridge Learner Corpus as well as language teaching materials and syllabi (see Capel, 2012). The EVP also identifies the phrases, phrasal verbs and idioms which learners at the various CEFR levels know.

The EVP provides a global benchmark for learner lexical proficiency at different levels. As Capel (2012: 1) notes in relation to the EVP, 'it offers extensive information about words, phrases, phrasal verbs and idioms and currently includes just under 7,000 headwords'. Therefore, the research addresses the increasing understanding that words are not fully meaningful when viewed as single entities but only achieve their meaning potential in combinations such as the four language features which encompass the data analysis chapters of the present study. The EVP follows a 'can-do' rationale based on the CEFR, using the language that students know, as evidenced in their written production (and, to a lesser extent, their spoken production), rather than should know, and is composed of both Academic and General English, taken primarily from the Cambridge Learner Corpus, as well as other sources.

Table 2.2 below illustrates the number of words up to B2 level and the number of phrases known by learners at each CEFR level (see Capel, 2012: 10) according to the EVP.

Table 2.2 The total number of phrases known by learners at each CEFR level in the EVP (Capel, 2012)

Level	**No. of words**	**No. of phrases**
A1	601	47
A2	925	147
B1	1,429	335
B2	7,711	561
C1	8,000	380
C2	1,600	722

From Capel's (2012) research, it is clear that the learner data show an increase in the number of phrases known at each of the CEFR levels. That is, until they reach C1 level, where the figure drops back down to 380. The figure rises again to 722 at C2 level. Be that as it may, when comparing the total figures at both C1 and C2 levels, it is clear that they contain more phrases than the B levels.

Table 2.3 The number of multi-word verbs known across the CEFR levels (Capel, 2012)

Level	**Number of phrasal verbs**
A1	4
A2	27
B1	135
B2	277
C1	99
C2	184

We see a comparable set of occurrences for phrasal verbs across the CEFR levels and again there is a decrease after the C level, Table 2.3 above illustrates this.

2.15 What constitutes a word?

In relation to measuring vocabulary, McCarthy and Carter (1997) raise the question 'what exactly should be counted as a word'? Based on its definition in the Collins COBUILD Dictionary, is a word a single unit of language, which, when written, has a space at either side? (Collins, 2009) If this is the case, what labels could be attached to linguistic phenomena such as multi-word units, phrasal verbs, delexical verbs, collocations and idiomatic expressions (The four components focused on in this thesis)? In addition, and to add further to the problem, Crystal (1987: 1) posed

the question of identifying acronyms and abbreviations as English words and established that the *Acronyms, Initialisms and Abbreviations Dictionary* by the Gale Research Company (1987) lists over 400,000 abbreviations. This includes high-frequency units such as *flu*, *hi-fi* and *FBI.* As a result, it is evidently a difficult task to firstly count the number of 'words' in the English language and secondly to ascertain how many 'words' a person has at their disposal. We do not simply store words as single items in our mental lexicon. Instead, research shows that the human brain stores words in multiple 'addresses', which include clusters and associative networks of various kinds (Wolter, 2002: 22).

Words are therefore not single isolated entities. Units of meaning may consist of a single word or a string of words, whether related grammatically, collocationally or semantically (see Section 6.2 on Sinclair and the 'idiom principle').

To assume that to know a word is merely to know its definition, or translation falls far short of an adequate description of word knowledge. Macro points out:

> Language is just not made up of individual words but has units smaller than words: phonemes and morphemes. It also has units bigger than words: clauses, utterances, sentences, paragraphs, monologues, dialogues, whole books and words are complex things in relation to other words and to us humans that use them. Words imply and entail other words. Words may even trigger our thought processes and therefore the utterances and sentences we provide (Macro, 2003: 62).

Furthermore, one of the most influential papers published in the area of second language vocabulary, Richards' (1976), outlines eight conditions of what it means to know a word. These conditions have been both accepted and criticised by many (see Meara, 1996, for example). According to Richards (1976: 83), linguists of his time considered the following eight characteristics to be fundamental to understanding a word:

1. The native speaker continues to expand and develop his/her vocabulary throughout their life whereas very little grammatical development occurs.

2. Knowing a word also means to know the probability of encountering that word in speech or writing and also to be aware of the words that are likely to occur with it.
3. Knowing a word means knowing its suitability to contexts, situations, genres and registers.
4. Knowing a word means knowing the syntactic behaviour of that word.
5. Knowing a word means knowing the form of that word including all derivations and conjugations.
6. Knowing a word implies knowledge of its associations and relationships with other words of the language.
7. Knowing a word means knowing the semantic values of a word.
8. Knowing a word means knowing the many different meanings of that word.

Evidence from the development of tests as detailed above, along with the evidence provided by corpus analysis of native-speaker data (and latterly, learner data), has meant that commitment to the importance of vocabulary in the teaching and learning of a second language has increased over time. The teaching and learning of vocabulary is no longer relegated to a secondary position; it has now become equal to grammar and phonology (and indeed has overtaken phonology in many cases) in terms of focus in many syllabi and published teaching materials. Meanwhile, corpus linguistics has contributed to the refinement of tests based on ascertaining the number of words a learner knows or should know.

2.16 English as an International Language (EIL)

As discussed elsewhere, as this study focuses on students living and working in Ireland, it is therefore focused on English as a Second Language (ESL) study as opposed to English as a Foreign Language (EFL). In the earlier sections of this chapter, we reviewed vocabulary teaching in relation to methodologies over the years but the core understanding of "a learner" inherent in these methods is usually a foreign language learner. Here we briefly consider this in relation to the learners in this study. According to Harmer (2007) EFL students are usually learning the language to use it in

interactions with any other English speaker in the world, be they native or not and they have usually studied the language in their home country. ESL students, in contrast, are living in the target language country and need the language to interact and, in essence, to function in that community. Harmer (2007), among others, suggests that the language input needs to be more differentiated for these different types of learning contexts (see also Pennycook 2016 on the gatekeeping role of English internationally). It is fair to say that drawing a binary between ESL and EFL is nowadays too simplistic as English is seen as a Lingua Franca in so many contexts. Other terms such as English as a Lingua Franca (ELF) and English as an International Language (EIL) are frequently more relevant to contemporary uses of English globally and at all levels of education.

From an EIL perspective, language is seen as neutral and not actually belonging to any set group or region (see Roux, 2014). As a model, it also encompasses the notion of language variation within the context, community or society in which it is spoken. Fundamental to EIL is the notion that speakers, native or otherwise, are not identical and that all languages have variations and dialects. The idea of the native speaker as the core model of language is challenged under EIL, so much so that the British Council (2014) state that the number of speakers of English as a second language far exceeds those speaking it as a first language. However, Jenkins (2006a: 171) argues that 'the belief in native speaker ownership persists among native and non-native speakers'. EIL scholars argue that 'intercultural competence' needs to be viewed as a core element of 'proficiency in English' (Sharifian, 2009: 4) and argue for a study of different speech communities. The EIL model is also applied within critical perspectives on variation where the aim is to understand the role of English in the formation and maintenance of societal power and dominant culture(s) (see for instance Fairclough 2001, who notes that language has arguably become the primary medium of social control and power).

2.17 English as a Lingua Franca (ELF)

A lingua franca is the language that is spoken between two people who do not share the same first language. Globally, English is the predominant language used as a lingua franca (see Firth, 1996; House, 1999; Seidlhofer, 2001). According to Crystal (2003), roughly only one out of four users of English in the world is a native speaker. Hence, the majority of conversations in English are between non-native speakers of the language. Firth (1996) stated: 'what is distinctive about ELF is that, in most cases, it is a 'contact language' between persons who share neither a common native tongue nor a common (national) culture, and for whom English is the chosen foreign language of communication' (Firth, 1996: 240). Within the sociolinguistic study of the spread of World Englishes, various models have evolved (and will continue to evolve) to capture the changing situation. The best known model is that of Kachru (1992). His influential model is based on three concentric circles. The Inner Circle (original sociolinguistic bases of English as it emerged and spread through diaspora, e.g. UK, Ireland, USA); The Outer Circle (Englishes that came through colonial rule where it was used as a lingua franca, e.g. Nigeria, Malaysia, India) and the Expanding Circle (countries where English is not a colonial legacy but where, nonetheless, it is increasingly used as a medium for commerce, trade, education, science etc. e.g. China, Israel, Saudi Arabia). Scholars such as Jenkins, (2000); McArthur (1998); Melchers and Shaw (2003) note English is used by many more speakers than Kachau's (1992) Inner Circle, Outer Circle and Expanding Circle allow. An alternative dynamic model is proposed by Schneider (2007) which views use of English less in relation to geo-political history. This model is more focused on sociolinguistic concepts of *identity.* This is not a sociolinguistic study however it is recognised that identity formation has a link to learning in an ESL context.

The work of Seidlhofer (2004) on ELF is often cited in the context of English language teaching. The idea that native speaker rhetoric is the model and desired variety of language remains the orthodox, according to Seidlhofer (2004). The native speaker is seen as 'custodian' of what is excepted in English language use (2004: 339). Notably, Seidlhofer (2004) and associates set about the systematic

study of the nature of ELF so that this would inform teachers, material designers and students alike as to norms that moved beyond the native speaker orthodoxy (examples of this are found in Jenkins, 2000 and Seidlhofer, 2004). Furthermore, corpus linguistics has recently been used to explore the language exponents and structures used in interactions between speakers of different mother tongues. The English as a Lingua Franca in Academic Settings Corpus (2003) and the General Vienna-Oxford International Corpus of English (2004) shine a light on ELF empirically.

Overall, it is clear that ELF encompasses a number of varieties of English and the majority of researchers in the field argue that native speaker-like language use is not 'the be all and end all' for learners and teachers of English. As O'Keeffe, McCarthy and Carter (2007: 153) argue language teaching is 'a complex activity taking place in different contexts and involving teachers and students with different cultural expectations'. Linking this to the present study we focus on a group of students who are in an ESL context in Ireland and they are also using English as a Lingua Franca when interacting among themselves. In terms of vocabulary acquisition then, there are different motivations at play, they are not learning a language for school exams or certification in an EFL sense; they are learning English to live, work and make friends in an English speaking country where they will be moving between being ESL, ELF and EIL roles depending on the interactional context. Their motivation to learn new vocabulary will have to be influenced by this. For example, they have clear real-world motivations for using language. Additionally, from a socio-cultural theory perspective (Lantolf and Thorne, 2006), their peers both in class and outside will mediate their learning and be part of their zone of proximal development. Some of these peers will be native speakers of English, some will be expert users of English while others may be at a level above or below their own level. For example, in Chapter 8, we discuss the idioms and we note the use of *I'm knackered* by one of the learners. This is a clear example of the socio-cultural context of language use adding to the lexicon of the learner because we know this phrase did not come from a formal syllabus. For the learners in this study, English is a 'resource' (Wink et al. 2016) that they draw on for many different motivations.

2.18 The Common European Framework of Referencing for languages (CEFR).

The Common European Framework of Referencing for Languages (CEFR) was first launched at a Council of Europe (COE) symposium in Switzerland in 1991. It was designed to inform syllabus and material designers by providing reference level descriptors (commonly referred to as 'can do' statements) as to what language learners can do at each level of proficiency. Available in 40 languages, the CEFR describes the abilities of learners in each language skill across six levels of proficiency: A1 Beginners, A2 Elementary, B1 Intermediate, B2 Upper-Intermediate, C1 Advanced and C2 Proficiency. CEFR 'can do' statements are intuitively derived with the aim of being broadly applicable to language being learnt (i.e. they are not language specific and they are not empirically-derived). For example, under the heading of *Interaction Speaking*, the list of Descriptors provided for the lowest levels, A1 and A2, under the theme of *Transactions to obtain goods and services* are:

A1 Can ask people for things and give people things.

Can handle numbers, quantities, cost and time.

A2 Can deal with common aspects of everyday living such as travel, lodgings, eating and shopping.

Can get all the information needed from a tourist office, as long as it is of a straightforward, non-specialised nature.

Can ask for and provide everyday goods and services.

Can get simple information about travel, use public transport: buses, trains, and taxis, ask and give directions, and buy tickets.

Can ask about things and make simple transactions in shops, post offices or banks.

Can give and receive information about quantities, numbers, prices etc.

Can make simple purchases by stating what is wanted and asking the price.

Can order a meal.

The CEFR is increasingly a standard reference point for examination boards, publishers and teachers as it provides a convenient consensus for assigning levels of proficiency to learners (O'Keeffe and Mark, 2017). Anderson (2007: 660) notes that the six levels of the CEFR 'have become a common currency in language education, and curricula, syllabuses, textbooks, teacher training courses' and 'not only examinations, claim to be related to the CEFR.' The CEFR proved instrumental in the context of the new Europe of the 21st century as it was used as the basis for many test development projects in Central Europe, 'as those countries emerged into the post-Communist world and saw the need for examination reform in order to give educational certificates credibility in the new Europe of the 21st century' (Anderson, 2007: 660).

Despite the many advantages that there are to having general consensus about what a particular level of competence generally means, the CEFR is not without its critics (for example, Fulcher, 2004; Weir, 2005; O'Sullivan, 2011), especially because it concerned itself primarily with the *validity* of performance-based descriptors of competency at each level. Its creators dismiss the second pillar of assessment, *reliability,* as a mere 'technical term' (Council of Europe, 2001: 177). As Osbourne (2014: 54) puts it, the question of *reliability* hangs over the CEFR 'How do I know that my Level B1 is your Level B1?' (see Figueras et al., 2005). Others have criticized the framework for being too Eurocentric in nature and also that it is being used to force teachers to work in particular ways (Anderson, 2007).

Another criticism of the CEFR is that because it was created as a generic framework (i.e. language neutral) it means that competency statements that are quite 'underspecified' (see Milanovic 2009; Hawkins and Filipović, 2012). Callies and Zaytseva (2013) also make the observation that the descriptors and *can-do* statements of the CEFR often appear too general to be of practical value. They note that this has led to an increasing awareness of the need to develop linguistic descriptors (Neff van Aertselaer and Bunce, 2011) or 'criterial features' (Hawkins and Buttery, 2010). The English Profile project (of relevance to the present study, see below) was an attempt to provide

empirical detail about learner English to complement and supplement the intuitively-derived language-neutral CEFR for languages.

Others have criticised the validity of the CEFR: 'The CEFR levels are neither based on empirical evidence taken from L2 learner performance, nor on any theory in the field of linguistics or verbal communication' (Hulstijn, Alderson, Schoonen, 2010: 15). Because the CEFR was designed as a generic framework for competence across languages, it lacks a specific L1 informed dimension. The CEFR does not take account learners' differing backgrounds such as gender, age, nationality, profession, length of stay in the country and daily interaction with the language and culture. De ng (1988) sums it up best when he states 'what we need to know if we want to develop good [proficiency] scales is how somebody acquires language, that is, what the developmental stages in language acquisition are' (De Jong, 1988: 74).

The English Grammar Profile (EGP) (O'Keeffe and Mark, 2017) and *The English Vocabulary Profile* (EVP) (Capel 2010) are two examples of endeavours that sought to use empirical evidence from learner corpora across stages of acquisition in English to enhance the descriptors for grammar and vocabulary, respectively. The EVP is of particular use as a reference point in the present study. This study offers a case study of learners in the ESOL context to test the notion of competency against.

As mentioned throughout this study, the CEFR claims that the four lexical strings that this research focuses on occur at B1 level plus but this is sometimes at odds with findings here as shall be discussed in the analysis chapters. When empirical research is carried out on language features in the CEFR, it is not uncommon that the results are at odds with the level specified in the framework. Carlsen (2014), for example, analysed the occurrence of connectives and found that the CEFR placed them at B2 and above whereas her study found them occurring at B1 level, this resulted with her asking the question 'is revision of the scale warranted' (Carlsen, 2014: 25).

An important point in relation to the present study is that it draws on the CEFR from two perspectives. First, the framework is used to check where the four features under scrutiny are placed within it (or within the EVP). Second, it compares this with the profile and distribution of their use in the cohorts that are taking part in the study, by CEFR level (A2 and C1). However, as is commonly the case in language teaching contexts, institutional imperatives dictated that there was funding to run two classes at A2 and C1. This was an imperative outside of the control of the researcher but it is a common reality for language teachers. In reality, teachers are often faced with teaching an Elementary class where, for example, the learners may range in competence from A1 (or true beginner) to B1. We will return to this point in Chapter 4 when we discuss the methodology. For now, it is a point that reflects the messy reality of operationalising the notion of levels in line with a neat six tiered framework.

2.19 Conclusion

Chapter 2 has chronologically surveyed the literature on vocabulary teaching and learning and has highlighted the fact that the place of lexis in the English language classroom was relegated to a secondary position over many years. As evidenced by this chapter, earlier language teaching methods, most notably the Grammar Translation, favoured the teaching of grammar and the translation of literary texts. Other methods, such as the Audiolingual Approach, while not teaching grammar overtly, involved the repetition of drilled phrases and dialogues (as habits to be formed) rather than placing importance on vocabulary teaching.

This chapter has also considered the English learning context of this study (ESL) and other related concepts (EFL, EIL, ELF and so on). While this is not a sociolinguistic study, there are important considerations nonetheless not least of all because our frameworks (e.g. CEFR) and syllabi and often driven by one notion of a learner and are focused on attaining native-speaker-like competence. This chapter also reviewed the CEFR. Though it has many limitations, it does offer a starting point for describing competence. It is a framework that lacks consideration of ESL contexts and one that more

describes EFL competence. However, this study will be able to add to a growing body of empirical work through its case study of ESL learners.

In summary, the chapter has analysed the treatment of vocabulary in research, in second language teaching and language learning. The chapter has highlighted how vocabulary was a neglected aspect of English language teaching under the various methods. It then identified that there was a requirement for a change in English language teaching methods as the needs and wants of the learner began to change. With the Reform Movement and the subsequent advent of Communicative Language Teaching in the 1980's, an interest in the importance of vocabulary for a second language learner emerged. This chapter has traced the rise of vocabulary in terms of focus and importance during the 1990s Lexical Approach and the chapter culminated with a survey of work on measuring the vocabulary size of the language learner. Also highlighted in the chapter is Corpus Linguistics as a tool for analysing language in use, a development which has unquestionably aided in bringing vocabulary to the forefront of English Language instruction. As has been shown, it is now no longer the Cinderella of language learning.

Chapter 3 will focus on vocabulary description and will act as the conceptual framework for the current study. Chapter 3 will identify the literature published on vocabulary in contemporary times and also focus on the four lexical features selected for the present study: 1) Delexical Verbs 2) Multi-Word verbs 3) Idiomatic expressions and 4) Collocations, all of which will be discussed in much greater detail in the data analysis chapters 6-9 of this thesis.

Chapter 3 Conceptual Framework:

Vocabulary Description

> Chunks are automatically produced strings of words which we use repeatedly. Corpus evidence shows us just how frequent the most commonly used chunks are (McCarthy and O'Keeffe, 2014).

3.0 Introduction

Having reviewed the literature and charted the rise in focus of vocabulary in English language teaching in the previous chapter, this current chapter focuses attention on the debate as to what is a 'word'. Through contemporary publications, it will become clear that the frequently quoted explanation of a word as a single unit of language is, in effect, flawed. This chapter will highlight through the works of scholars such as Firth (1957), Halliday (1966; 1985) Sinclair (1966; 1983), McCarthy (1987), O'Keeffe, McCarthy and Carter (2007) and Timmis (2015), that lexical strings such as chunks, fixed phrases and collocations are in fact units of language in their own right and behave in a similar manner to a single unit. In addition, it will be shown that such lexical combinations are acquired, stored and recalled by a speaker as one holistic unit. Naturally, this does frequently create a problem in measuring and analysing vocabulary size of a language user. This chapter will focus on the literature published on four multi-word units:

- Multi-Word Verbs
- Delexical Verbs
- Collocations
- Idiomatic Expressions

Competency in the use of these lexical features is frequently attributed to native speaker and proficient non-native talk. As chapters 4 to 8 will detail, these four items will form the basis of this thesis where they will be examined in the ESL learners' language use, at difference levels of competency.

3.1 The 'word' debate: multi-word verbs, delexical verbs, collocations and idiomatic expressions

Multi-word verbs, delexical verbs, collocations and idiomatic expressions have all been shown to occur at a high frequency in native speaker communication (Howarth, 1998; Nesselhauf, 2003). Conklin and Schmitt (2008: 74) go so far as to describe these bundles as 'making up a large part of any discourse'. It could be argued that a native speaker communicates by choosing the correct word combination and configuration from their pre-acquired lexicon where they are stored as 'semi-pre-constructed phrases' (Sinclair, 1991:17). A native speaker speaks in language chunks rather than individual units of language and therefore uses Sinclair's 'idiom principle' (a set of pre-constructed combinations available in the lexicon of the speaker) more than the 'open choice principle' (the slot and filler model where virtually any unit of language can be inserted) (see section 2.12.1 below).

Since the advent of the Lexical Approach to language teaching (circa 1990's), a new found interest in the learner has arisen. As teachers are now focusing more on communication in the classroom, syllabi have recently been created based mainly on the needs and wants of the language learner. Through analysis of learner needs, it is clear that the ability to communicate in a similar manner to a native speaker emerged as a fundamental motivation for the language learner. Corpus Linguistics and frameworks such as CEFR clearly highlight these four units as prominent in native speaker English discourse. Ergo, they are imperative for the second language speaker. Nesselhauf (2003: 223) clarifies that they 'not only enhance accuracy but also fluency'. Be that as it may, each continues to prove difficult for even the most advanced learners, so much so that an error in the use of such multi-word units often identifies the learner as a non-native speaker (see Granger, 1998; Lorenz, 1999; Nesselhauf, 2003; 2005). The following subsections will discuss previous research on the topics of multi-word verbs, delexical verbs, collocations and idiomatic expressions, the four data analysis topics of this present study.

3.2 Multi-Word Verbs

An example of a recurring lexico-grammatical pattern is the multi-word verb or verb and particle construction (Baldwin and Villavicencio, 2002). The term: *phrasal verb* first appeared in Smith's (1925) work *Words and Idioms* and this coinage has become an umbrella term within ELT practice to cover a span of multi-word units (though technically it only refers to one type as detailed below). A phrasal verb is comprised of a verb plus one or more other items, usually a prepositional or an adverbial particle. Examples of phrasal verbs are *give up, go away, put up with*. Such a construction was initially conceived as more than one word each with an individual meaning; however, a more contemporary opinion is that all components work together to form a single semantic unit. Recently, the term phrasal verb has been subsumed within the general purview of multi-word units or multi-word lexemes (see Linn, 1999; Schmitt, 2000; McCarthy et al, 2001). Courtney (1983) and Carter and McCarthy (2006) further distinguish between three main types of multi-word verbs, a distinction that has since been adopted by many and will be discussed again in Chapter 6 of this present study. This linguistic distinction appears as follows:

1. **Prepositional Verb** - this type contains a verb followed by a preposition for example, *to go on.*

2. **Phrasal Verb** - this type is composed of a verb followed by an adverb, for example, *to take away.*

3. **Phrasal prepositional Verb** - this type is structured with a verb followed by both an adverb and a preposition, for example, *to put up with.*

Mc McCarthy and O'Dell (2004) postulate that there are over 5,000 phrasal verbs in the English language with 1,000 of them in regular use. These high-frequency features are constructed in

different ways. Firstly, they can be transitive (with an object) for example *we ran into Mark in London*, or intransitive (without an object) for example *my alarm went off*. Secondly, they are composed of a lexical verb such as *break, bring, call, do, go, give, take* and a particle that is either an adverb or preposition or as seen above both: *away, up, on, off, after*. However, the lexical verb's frequency is less when it occurs in such a combination than when it occurs alone as a main verb. McArthur (1989) stated they have 'always' represented a 'vigorous part' of the English language as they make up roughly 'one third' of its vocabulary and new ones are coined on a 'daily basis'. Despite such a frequency of occurrence in the English language, they continue to be difficult to define. Gardiner and Davies (2007) go so far as to state 'linguists and grammarians struggle with nuances of phrasal verb definitions'. A more archaic view saw a phrasal verb as a unit which contained just an adverbial particle (Live, 1965; Bolinger, 1971; McCarter and Atkins, 1974).

In terms of interpretation, the meaning of such a multi word verb is interpreted through the overall and combined meaning of the lexical verb and its collocating particle. Phrasal verbs tend to add a richness and often a metaphorical sound and nature to the language but they also act as an inclusive device in language interactions. It has been identified, that a central function of phrasal verbs in the culture in which they are expressed, makes the language less formal, less face threatening, and serve to include the language user in the local culture of the expression (Halliday, 1990, Sinclair, 1991). As an integral feature of colloquial and informal English, they make the language sound natural, familiar and unstilted. Phrasal verbs will be discussed in detail in data analysis Chapter 5.

3.3 Delexical Verbs

Thus far, we have seen how the definition of *word* has changed from the archetypical view of a single unit easily defined in a dictionary to incorporating numerous units combined together to form a single semantic meaning. This subsection focuses on the literature published on the topic of delexicality. Delexical verbs represent a further feature of collocabilty. This canonical lexical string with examples such as *make an effort* and *have a look* is a ubiquitous feature of the English language, which has

received scholarly attention dating back as far as the Pre-Prague school era (see Mathesius, 1913 and Nickel, 1968). It embraces the relationship of high frequency verbs with their collocations with nouns, prepositional phrases and particles (see O'Keeffe, McCarthy and Carter, 2007).

In his *Papers in Linguistics* (1957), Firth coined the term collocation by propagating a 'new definition of word'. He claimed that words do not 'occur alone' and that in order to truly understand a word we must look at the other units with which it 'co-occurs'. Be that as it may, it was much later and in the 1990's when Firth's notes became central in language learning. The substantial increase in phraseological studies of the 1990's (Goldberg, 2006) placed Firthian ideologies at the centre of linguistic study. As a result, delexical verbs were identified are a clear example of words occurring in unison and they have proven to be a substantially frequent feature of English language in use (Sinclair and Renouf, 1988; Timmis, 2003; O'Keeffe, McCarthy and Carter, 2007).

A vast majority of the work on delexical verbs coincides with the advent of corpus linguistics and the ability to both analyse authentic language in use and measure combinations of words in an arguably easy way. In research, delexical verbs have recently become ever more topical as a frequent feature of English native speaker language. Based solely on corpus analysis, Hanks (1990) found that the commonest of occurrences of the verb *take* in English are in fact delexical (Hanks, 1990: 135). Notwithstanding, such verbs are found more commonly in native speaker speech than non-native; Howath (1998) proposes that the ratio of occurrence is 13 percent against 21 percent. Such a propagation is further highlighted in the plethora of comparative research conducted on the topic of delexical use. Such research frequently compares and contrasts the use by native and non-native speakers of English (Fan, 1991; Allenberg and Granger, 2001; Siyanova and Schmitt, 2007 McCarthy and O' Keeffe, 2011).

Delexical verbs are verbs whose collocating items determine the meaning and the interpretation of the lexical verb. The verb itself is seen as semantically indeterminate where it loses most of its lexical content (hence it is delexical). Examples would include the difference in sub-sense of the verb *have*

in collocations such as *have a party, have a car, have a baby, have a headache* (McCarten, 2007). Here it is clear that there is a wide scope of semantic differences in the use of the same verb, but in different combinations. The verb *have* is lexically light and therefore, as Sinclair (1991) notes, 'delexicalised.' Sinclair (1991) adds that the emphasis of a delexical verb is on its meaning in combination rather than its function alone.

COBUILD Grammar (Sinclair, 1990) defines delexical verbs as 'a verb which has very little meaning in itself and is used with an object that carries the main meaning of the structure' (1990, xix). Moreover, around the time of the emergence of vocabulary out of the relegation (as discussed in Chapter 2), Moon (1987) illustrated that the collocation of a delexical verb was different to that of a phrasal verb. Moon (1987) clarified with examples of the core lexical verb *make* combined with objects such as *make an effort*, *make oneself at home* and to *make it*. These occurrences of *make* are unmistakably delexical since according to Carter, McCarthy and O'Keeffe (2007: 36) 'the lexical content of the verb has to be combined with the lexical content of the words it is collocating with in order to achieve its meaning potential.'

Such structures normally contain a high-frequency lexical verb, but it is the particle and therefore the overall meaning that reduces its frequency of occurrence. Generally, the more frequent a word is, the less independent meaning it has, as it is more likely than not occurring as part of a combination, albeit delexical, phrasal, idiomatic or collocational (see Carter and McCarthy, 1988: 153; Cowie, 1998: 135).

Delexical verbs are structured in three different ways:

1) **Verb + Noun** e.g. *to have an argument.*

2) **Verb + Adjective** e.g. *to go red.*

3) **Verb + Adverb** e.g. *to take it slowly.*

According to Altenberg (1993), the verb-noun delexical collocation is in fact one of the most frequent in the English language. Examples of such a delexical verb occurs at the core of a vast number of native speaker texts and speech with the Oxford English Corpus (1928) citing the most frequent lexical verbs in a delexical construction as all occurring inside the 100 most frequent words of English. For instance, *have* is the ninth word, *do* is the nineteenth, *go* is 49th, *make* is 52nd, *take* is 60th and *give* 97th in the most frequent words in English. The basic rules of English imply that these verbs will occur in combination with some form of collocate, the majority of which have been shown to be delexical (See O'Keeffe, McCarthy and Carter, 1997). Live (1973) noted delexicalisation as an 'increasing trend' and our ability to examine the phenomenon has substantially risen since the ready availability of language corpora. Live (1973) highlighted the manner in which such a cluster splits the verb and collocate and places the meaning on the collocate. Views such as Live's (1973: 31) assertion that the lexical verb is substantially devoid of meaning have been challenged first by Stein (1991) and, more recently, by Allan (2017) whose academic English corpus (1997) clearly shows that the verb is a lot more than a 'mere auxiliary'.

Leading English language dictionaries, such as *The Oxford Advanced Learner's Dictionary of Current English (2000), The Longman Dictionary of Contemporary English (1995),* and the *COBUILD dictionary (1995),* all include various delexical examples. In addition, a range of specialised publications cite delexical definitions, for example: *The BBI Dictionary of English Word Combinations, The LTP Dictionary of selected Collocations* and *the Oxford Collocations Dictionary for Students of English.* Furthermore, based on findings from the British National Corpus*, the Longman Grammar of Spoken and Written English* (1999) also includes a chapter that deals with the three most frequent delexical verbs *have*, *make* and *take.* However, it has been shown that despite their frequency in speaker language and their necessity to second language learners, delexical verbs are not a primary focus in many learner dictionaries (see Coffey, 2006). In his research, Coffey (2006) examined five different learner English dictionaries and found few delexical references overall:

Table 3.1 Learner Dictionaries and examples of Delexical Verbs

Dictionary	**Delexical Examples**
Longman Dictionary of Contemporary English (2003)	40
MacMillan Dictionary for Advanced Learners (2002)	29
Cambridge Advanced Learners Dictionary (2003)	28
Oxford Advanced Learners Dictionary (2003)	26
Collins COBUILD English Dictionary for Advanced Learners (2003)	12

Further indicating delexical verbs as a high frequency lexical item, McCarthy and O'Keeffe (2011) conducted a study based on the use of delexical verbs across three different spectrums: 1) Learner language 2) Academic Language 3) Conversation in Use. The data for their research was taken from the Cambridge Oral Examinations and The Cambridge Limerick and Shannon Corpus (CLAS). They analysed the use of delexical verbs such as *get, have* and *make* and found very little discrepancy throughout the three language genres. They found that *have* is somewhat imbalanced while *get* was used the least delexically by the English language learners observed.

Delexical verbs also exist on a cline of idiomaticity from clear and transparent, for example *make an effort* to opaque and idiomatic, as in the example *go crazy* (More, 2005). Sinclair (1987: 323) refers to the cline of idiomaticity as 'a progressive delexicalisation' while Cowie and Mackin (1975: x) contrast the idiomatic and transparent sides by observing that it is 'not sharply drawn but [rather] hazy and imprecise'. Such a cline of idiomaticity has resulted in debate as to the features of delexical verbs. For instance, Aisenstadt (1979) identifies delexical verbs as non-idiomatic restricted

collocations whereas and in sharp contrast, Sinclair and Renouf (1988) identify them as idiomatic phrases. Others view idiomaticity as not a simple case of either or but rather a case of degree (Cowie and Mackin, 1975; McCarthy 1999; Wang, 2003).

Delexical verbs have important functions in language. Asher et al (1994); Biber et al, (1999) and Lock, (2005) highlight the following primary functions:

1) They make the interaction less face threatening.
2) They are informal or neutral in style.
3) They aid in creating a sense of language community among speakers.
4) They function as an inclusive device.

3.4 Collocations

In 1957 Firth propagated and coined the idea that the meaning of a word depends on the 'company it keeps.' The co-occurrence of units of language collocating together is for Firth and the Neo-Firthians a key aspect of meaning and stresses that words cannot be considered in isolation. Firth (1957) termed this phenomena as collocation. In explaining collocation in 1966, Bazell et al published *In Memory of JR Firth,* containing essays by scholars on the topic of collocation (Sinclair, Halliday and Mitchell). Firstly, Halliday (1996) cites *strong tea* versus *powerful tea.* In English, one is accustomed to drinking *strong tea* rather than *powerful tea,* although there is no apparent difference in structure of both phrases; the difference is in the probability of co-occurrence. McCarten (2007) gives the frequency of occurrence of *heavy rain* in English versus the improbable *heavy sun* in her explanation of collocation. In addition, both Benson (1985) and Biskup (1992) identify grammatical and lexical combinations. A grammatical collocation is when the dominant word fits together with a grammatical word for example a noun, verb or adjective followed by a preposition. On the other hand, a lexical collocation is when two lexical words are combined, characteristically, for example a particular noun followed by a particular verb.

Moreover, Neo-Firthian Sinclair (1991) elaborated the *open choice* and *idiom principle* (see Chapter 5 for further detail). Sinclair (1991) identifies collocations as occurring in two ways:

1) Language is creative and therefore a wide variety of language components can come together to form a coherent message. So, language is seen as open and possessing no restrictions in its configurations. This was referred to by Barnbrook (2007) as 'the normal way' of describing language and was also known as the 'slot and filler' model (Sinclair, 1991). The open choice principle encompasses sentences as slots available for filling by lexical items.

2) Sinclair's second ideology is the *idiom principle.* Here regularities in the combination of language forms are identified and stressed. Units of meaning are composed of more than one individual word. Such units are patterned occurrences of lexical items with associated syntactic forms.

Research undertaken by Sinclair (1991) and Nattinger and DeCarrico (1992) suggests that there is more lexical patterning and collocational occurrence in language than had previously been identified. Schmitt (2000) attributes collocations to the means by which the mind tends to chunk language to make it easier to process and interpret. According to Schmitt (2000: 79) there are different levels of collocational complexity:

3) **Fixed Idiomatic Expressions**: *bite the dust, shoot the breeze,* and *rain cats and dogs.*
4) **Invariable Collocation**: *from head to toe*.
5) **Collocation with limited choice at one point:** *give/allow/permit access to*.
6) **Collocations with limited choice at two points:** *as dark/black/ as coal/night/ink*.

Renouf and Sinclair (1988) argued that the more frequent a word is in English, the less independent meaning it has. Renouf and Sinclair (1988) see words as acting together with other words,

contributing to the creation of a linguistic chunk, while Schmitt (2000) claimed that the term ***word*** was too general to identify the different formations and complexities of vocabulary.

Moon (1998) identified the following types of collocations:

a) **Cranberry Collocations**: unique and not found in any other formations, for example: *in retrospect* and *to and fro*.
b) **Defective Collocations**: they are difficult to interpret, for example: *at least* and *in time*.
c) **Bound Collocations**: for example: *to foot the bill*.
d) **Phraseological Collocations**: they are the weakest form of collocations for example: *on show* and *in action.*
e) **Simple Formulae**: they have a function in discourse and are often emphatic. For example: *alive and well* and *you know*.
f) **Sayings:** these are well known quotations or catch phrases. For example: *an eye for an eye.*

Moon's (1998) work and other work on collocations will be dealt with further in chapter 8 of this thesis.

3.5 Idiomatic expressions and figurative language

A further form of multi-word unit, which is dealt with in Chapter 8 of the present study, is idiomaticity or formulaic language. In archaic generative linguistics, idioms were mostly discarded as language anomalies (for example see Fraser, 1970). But, with the advent of corpus linguistics and the ability to analyse real language in use, the primacy of idiomatic expressions in English has become evident. According to Moon (1997), 'idioms are a form of multi-word unit that is an item comprising two or more words where the meaning is not always interpretable from each individual component'. Expressions such as: *written all over his face*, *to kick the bucket* and *to hit the roof* are a frequent feature of the native speaker lexicon. Moreover, there are over 6,000 idioms

in the *Oxford Dictionary of Idioms* (2009) and 5,000 cited in the *Cambridge Dictionary of American Idioms (2003).* Be that as it may, there are in fact numerous publications (Sinclair, 1991; Sorhus, 1997; Howarth, 1998; Biber et al 1999; Erman and Warren, 2000; Foster 2001; Rayson (2008); Oppenheim, 2000; Foster, 2001 for example), which debate the frequency in occurrence of idioms in the English language. In fact, estimates range from 58% coverage to an occurrence of 'once in every five words' (See Chapter 8 for further detail).

While researching his book *Figurative Language*, Pollio (1990) gathered 200,000 words of recorded data from English native speaker interactions in settings such as political debates, psychotherapy sessions, as well as compositions written by students and adults. Pollio (1990) and his researchers found that the participants in the study used idioms at an average rate of 4.08 per minute. Later and at the University of Georgia Athens, Cooper (2010) transcribed the dialogues from three recorded hours of popular, monolingual, English television programmes. He found that idioms occurred in this context at a rate of three per minute and that the idioms uttered were crucial to understanding the storylines of the shows. Despite conflicting views on the frequency of occurrence of idioms in English, the research clearly shows that idioms are an integral feature of English and are used by the native speaker on a regular basis. Idioms have been shown to express different functions in language; they are often 'engaging', 'casual', and 'colourful' and make the language sound almost poetic (Simpson and Mendis, 2003). Idioms are more often than not register specific and a thorough understanding of the context and indeed even the culture in which they are expressed is necessary to interpret them correctly (see Nattinger and Decarrico, 1992; Moon, 1998; Wray and Perkins, 2000).

Idioms hold vital socio-interactional functions and are frequent features of jokes, anecdotes, story-telling, advertisements, movies and songs. Fernando (1978) commented on the central role of idioms in language when he stated

> The elimination of idiom from languages would not alter their essential structural design but if man were bereft of the power of wit of which idiom is one manifestation, language would approximate more nearly to systems of animal communication and the algorithmic operations of automata (Fernando, 1978: 341).

So without the 'colour 'of idiomatic expressions Fernando (1978) sees language as plain, boring, monotonous and even robotic. Further to this, McCarthy (1998) states that idioms are far from random and not just mere 'quirks' of native speaker discourse. They are features of real language in use that have specific language functions. Idioms make the language stimulating to listen to, act as evaluative responses and are a form of social inclusion and solidarity between groups who strive to ensure their language is casual and non-face-threatening (Malinowski, 1923; Firth, 1957; Halliday, 1985; Carter and McCarthy, 1988; Moon, 1992; Sinclair, 1991 and 2000). Moreover, idioms are an integral part of a conversation as they are on the one hand used to highlight shared views, opinions and collaborative ideas of speakers and on the other hand are used as discourse boundaries to change a topic or finish a conversation in a polite and non-threatening way (Drew and Holt, 1998). According to Moon (1997) the main features of an idiom are that they are:

1) Institutionalised in the culture in which they are spoken.
2) Fixed in structure.
3) Non-compositional.

Ergo, in order for a structure to be an idiom it must be accepted and used regularly in the culture of the language in which it is spoken. Moreover, it must be fixed in structure with no variations and the meaning should not be derivable from its constituents. Semantically, idioms clearly exist on a scale of transparency from transparent in meaning to figurative. Simpson and Mendis, (2003), Gibbs (1987), Kaysar and Bly, (1995) and McCarthy and O'Dell (2010) show how examples such as *to play your cards right* and *to let the cat out of the bag* are transparent and opaque respectively.

Furthermore, according to Carter (1987), there are numerous types of figurative expressions. Table 3.2 below identifies with examples the various forms (See Chapter 8 for further types of idiomatic expressions):

Table 3.2 Types of Fixed Expressions and Idioms

Type of Fixed Expression	**Example**
Idioms	
Irreversible Binomials	*Spick and span; red tape*
Full Idioms	*To rain cats and dogs, to smell a rat*
Semi-Idioms	*Dead drunk, the party kicks off at nine*
Proverbs	*A bird in the hand is worth two in the bush*
Stock Phrases	*When all is said and done*
Catchphrases	*That's another fine mess you got us into*
Allusions/Quotations	*To be or not to be that is the question*
Idiomatic Similes	*As daft as a brush; as drunk as a skunk*
Clichés	*Long-time no see; bottoms up*
Connectives	*To sum up; finally*
Conversational Gambits	*Guess what!; I wondered if I could have a word*
Stylistic Formulae	*Ladies and gentlemen; regarding my recent request*
Stereotypes	*It's not what you think!; I thought you'd never ask*

In addition, Grant and Bauer, (2004) divide idioms into three main categories:

1) **Figuratives,** for example multi-word units such as *a bird's eye view.*

2) **Onces** which contain one element that is non-compositional, for example by *a long shot/chalk.*

3) **Core Idioms,** for example the *bee's knees.*

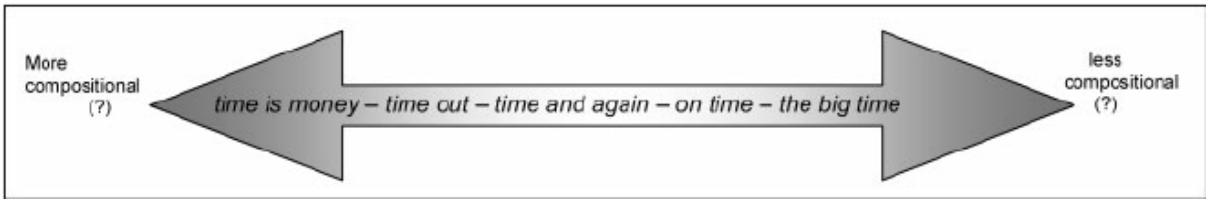
Notwithstanding, with such a vast array of features, idioms remain a problematic area for the second language learner. Firstly, receptive and productive use of idioms is often affected by negative transfer from the L1. Mc Lay (1987), Wray (1999), Nesselhauf (2003), and Spöttl and McCarthy (2004) all conducted research which shows the negative transference the L1 can cause to the language learner and identify idiomatic expressions as a core error (See Chapter 8 for further detail).

Secondly, a further issue is created in that idioms are embedded in the culture of the English language. Perhaps a language learner is not a long time member of such a culture (and as a result not a native speaker) and is reluctant to 'mimic' the customs of that culture (see King, 2000 for a full discussion).

Thirdly, there are numerous variations of idiomatic expressions both on a lexical and a grammatical level that create a stressful task for a learner. Such variations though interpreted easily by the native speaker are a major hindrance to the language learner. This hindrance, Prodromou (2005) titles The Paradox of Idioms. The paradox is the apparent simplistic way a native speaker can automatically understand and productively use an idiom, a skill which the majority of language learners find practically impossible.

Finally, Idiomatic expressions are interpreted differently by both the native and non-native speaker. As already mentioned in this chapter, Sinclair's *Open Choice* and *Idiom Principle* (1991) are the most cited means of interpretation. Prior to Sinclair's (1991) contributions, Swinney and Cutler (1979) reference Bobrow and Bell (1973), Fraser, (1974) and Heringer (1976) who all advocate the *Idiom Mode of Processing*. They stress that idioms are stored in an 'Idiom List' completely separate from literal exponents in the mental lexicon; when an expression is not understood, this idiom list is activated. Such a manner of interpretation has proven a difficult task for a language learner to achieve.

Fig. 3.1 The Compositionality of Idioms



(Martinez, 2015)

Besides these complications Pawley and Syder (1983) refer to 'the puzzle of native-like selection'. This selection focuses on the idea that if a language learner wishes to communicate like a native speaker then accurate knowledge of the structures that govern the language is not enough; that learner also needs to have a confident grasp of the core lexical patterns which the native speaker frequently uses, of which idiomatic expressions are only one manifestation.

3.6 Conclusion

In sum, this conceptual framework has surveyed literature published on the topics of lexical strings and language chunks and described in detail the features of the four multi -word units of the thesis. First of all, the chapter clearly identified how native speakers of English process and recall language from their lexicon in chunks rather than single units of language. It has been shown that such combinations are stored in the mind and automatically retrieved by the native speaker upon necessity. Secondly, this chapter identified that to replicate such a manner of acquisition and production remains a constant difficult task for the language learner. This has been shown to be a primary reason as to why such lexical combinations pose a problem to the learner. Following that, the chapter chronologically surveyed the available literature on the four lexical strings that make up the data analysis sections of this current research. The literature on multi-word verbs, delexical verbs, collocations and idiomatic expressions were charted. Definitions, characteristics, functions and frequencies were all highlighted with the integral aim of highlighting how words do not occur alone but in combinations.

In sum, this conceptual framework chapter has posited the significant role that multi-word units play in English and has argued that lexical meaning is on a frequent basis encoded in strings of more than one word. Apodictically we can see that a word is no longer seen as a mere single unit to be defined by means of a dictionary; instead words can form part of multi-word units such as the four types discussed in this research. The chapter has selected and described four of the most common, core types of multi-word phenomena (Multi-Word Verbs, Delexical Verbs, Collocations and Idiomatic Expressions) in terms of their forms and functions. The chapter has provided a broad framework for the analysis of the ACE learner corpus. The above framework enables us to judge (a) how the distribution in the non-native speaker data relates to the core forms and functions in native speaker usage, and (b) what difficulties, if any, the learners seem to be having, and (c) how near or far from native speaker usage the non-native data is at different CEFR levels. The next chapter will describe the data and the methods used to examine the learner language in order to answer the central research questions and sub-questions discussed in Chapter 1.

Chapter 4 Methodology

> Technology has been the major enabling factor in the growth of corpus linguistics, but has both shaped and been shaped by it. The ability to store masses of data on relatively small computer drives meant that corpora could be as big as one wanted.
> McCarthy and O'Keeffe, (2010: 6)

4.0 Introduction

The current chapter sets out to detail the data (corpora) and methodology utilised in the present study. The chapter identifies the various components of the 170,000 word Adult Corpus of English (ACE) corpus. It also details key functions of the corpus-analytical software *Wordsmith Tools* (Scott, 2016) that were used in the analysis to create the word lists, keyword lists, concordance lines and clusters from the data of the ESL learners. The chapter will also detail processes of normalisation and comparison (where the British National Corpus (BNC) was used as a baseline corpus).

4.1 Corpus Linguistics: origin and definition

Corpora: definition and use

Reppen et al (2002: 2) defines a corpus as a 'principled collection of naturally occurring texts (written or spoken) stored electronically… that help to identify both linguistic and situational co-occurrence patterns'. As Biber and Conrad (2009) and Mauranen (2003) demonstrate, a corpus can be used to identify the most frequent and important linguistic exponents of a language. Furthermore, corpus linguistics identifies the language used in the 'real' world that can both inform and act as exemplars in language teaching materials. A number of works by various experts in the field of Applied Linguistics continue to stress the importance of corpus linguistics as a tool for language analysis, instruction and learning (McCarthy and O'Dell, 2004; Schmitt, 2004; Sinclair 1996; McCarthy, McCarten and Sandiford 2015). According to Gardner and Davies (2007), such works have helped to combine corpus linguistics with language instruction and as a result, have 'begun to bridge the gap between corpus-based findings and fruitful instructional practices' (2007: 352). However, it was

through the works of American Structuralists such as Harris, Fries and Hill in the 1950s that authentic language and use became the core of what linguists and practitioners studied. Halliday (1993) claimed that corpora have led to a qualitative change in our understanding of language since they function to combine both data gathering and data theorising.

Historic context and lineage

The term *corpus* is a direct translation from Latin meaning the word *body* and is used to describe this method of analysis as a *body* of text. Despite a surge in use, it is important to note that corpus-type analysis is not a new activity. In fact, analysis of large bodies of text(s) dates as far back as the 12th century. Biblical scholars manually indexed the Bible, line by line and page by page and Anthony of Padua (1195-1231) is accredited with the first known concordance of the Bible (McCarthy and O'Keeffe, 2010). Nevertheless, in the early days, the utilisation of corpora as a language analysis tool was not limited solely to religion. In 1787, Becket analysed and concordanced the linguistic features of the poetry and plays of William Shakespeare. This, according to McCarthy and O'Keeffe (2010: 12), 'provides an immense resource for the literary scholar'. Also, Roberto Busa (the *Index Thomisticus,* 1949) created an electronic and lemmatised index of the complete works of Saint Tomas. The vital information gathered by the means of corpus analysis has resulted in the publication of language learning materials such as dictionaries. As the result of many years of manually analysing a paper corpus, Samuel Johnson in 1755 published the first dictionary of the English language. Johnson's 'corpus' was taken from the attested language from 1560 to 1600. In addition, the entries in the Oxford English Dictionary (OED) of the 1880s were directly drawn from three million hand written slips of paper. Later, in *A Modern English Grammar on Historical Principles* (1909-49), Jespersen identified and analysed the linguistic features drawn from the works of the literary authors: Chaucer, Shakespeare, Swift and Austin.

Corpus analysis has been simplified by the developments in technology and the advent of the personal computer (Roberts, 1975). Henceforth, the archaic paper data of earlier years were replaced with larger readily available computerised databases, easily accessible with the click of a mouse. The first

computerised corpus was the Brown Corpus (data gathered in 1961), containing 1 million words of American English; the Brown Corpus, though edited, presents an insight into the written English used in various American genres during this period (see McCarthy and O'Keeffe, 2010). It contains 2,000-word samples from press reports, government documents and fiction, and accurately presents the language used across the various genres.

Size and Representativeness

A recurring challenge with corpus analysis is choosing or building a corpus that will accurately represent the language genre the researcher wishes to analyse. Let us take the present study for example; the main topic of research is ESOL learners' use of lexical strings; therefore an analysis of a corpus compiled of data from EFL learners would not be representative. This notion of representativeness is core to corpus design and is discussed by Leech (1992: 16) who states corpora are 'generally assembled with particular purposes in mind, and are often assembled to be (informally speaking) *representative* of some language or text type'. Corpus linguists such as Biber (1993), Sinclair (1996), Johansson (1998) and Tognini-Bonelli (2001: 2), among others, underscore the importance of gathered corpus data 'according to explicit design criteria'. Moreover, Hunston (2002: 30) identifies time as an integral aspect of corpus representativeness; she asserts that all language changes over time and that each and every corpus should be revisited as 'any corpus that is not regularly updated rapidly becomes invalid'.

Corpora vary in size from thousands to millions and even billions of words. For example, The Bank of English Corpus (BOE) comprises 500 million words, while The British National Corpus (BNC) contains 100 million words of spoken and written language while the English Web corpus 2015 (enTenTen15) comprises 15 billion words. While a corpus is designed to accurately represent language there is, however, no one corpus to suit all language analysis purposes. According to Hunston (2002: 3), a corpus alone is 'merely a store of used language' (Hunston, 2002:3). In order to exploit a corpus, analytical software, such as *Wordsmith Tools* (Scott, 2016) (used in this present study), is required, as discussed further below.

4.2 The Adult Corpus of English (ACE)

4.2.1 The corpus: general overview

The Adult Corpus of English (ACE) consists of transcripts of three-hour English as a Second Language classes, recorded over the course period of one 12 week term, at a further education (FE) institute. The FE College, located in an Irish city, offers 78 full-time courses, ranging from disciplines such as Broadcast Journalism to Sports, Recreation and Fitness, and the College also offers a further 180 part-time courses in topics from Languages to Childcare. The College's main cliental are school leavers and mature learners, which the Director attributes to the fact that 'further education in Ireland has recently and rightly been recognised as a distinct sector achieving parity of esteem with primary, secondary and third level education' (Director[5], 2015).

The ACE corpus was compiled during the spring term of 2010 to 2011 and amounts to 150,000 words of spoken and 20,000 words of written data. The spoken data is composed of classroom interactions with the teacher and other students. It also includes oral examinations and presentations (see Table 4.1).

Table 4.1 the four datasets of the ACE corpus (rounded numbers)

ACE Dataset	A2 Speaking	A2 Writing	C1 Speaking	C1 Writing
No. of Words	64,000	7,000	86,000	13,000

Additionally, the written data includes class assignments, essays and homework. The researcher, at the time, worked as an ESL and German as a Second Language teacher at the College. The English language classes at the College were conducted through the medium of English, encompassing interactive techniques of the communicative language teaching method (CLT). Overall, the receptive skills of reading and listening and the productive skills of speaking and writing, along with grammar

[5] Anonymised reference

and vocabulary, were integral components of the overall syllabus (see Appendix 1 and 2 for Learning Outcomes).

4.2.2 Participants

At the time, there were two cohorts of ESOL learners at the College. These were labelled as *A2* and *C1*:

1) A2 (CEFR) Elementary

2) C1 (CEFR) Advanced.

Students were divided into their appropriate groups based on a written placement test (on the opening night of term) which required them to complete a form with personal details, plus hobbies and undertake a writing task. The tasks were descriptive:

- Describe a famous person of your choice,
- Describe your favourite film.

Based on the writing samples from the above tasks, two teachers (one of whom was the researcher) ranked the learners by level. Essentially, they sought to divide the total cohort of ESL students into two groups. These groups were set at A2 and C1 levels. It must be noted however, that there was only funding for two levels. Therefore, as is often the reality in language schools where budgets only allow for a limited number of class levels, the teachers tried to divide the groups in terms of their approximations to these levels. Most learners were at or around these two levels. However, it is recognised as a limitation of this study that some students could have been placed below or above the level of A2 or C1 were there funding for more class levels. It is noted that students were placed in the nearest appropriate level. That is, for example, no Beginner was placed in the C1 level, or no B2 level learner was placed in the A2 level etc. Within the analysis chapters, where students are below the level, it has been noted in qualitative commentaries. The information gathered about individual students also played a role in placement (see Appendix 3). For instance, information about length of

stay in Ireland, occupation, hobbies and etc. provided the teachers with vital sociocultural information and insight into the degree of interaction with the English language outside the language classroom. The classes were divided in the following way (see Appendix 3):

A2 cohort: 9 students.

Programme: The Further Education and Training Awards Council (FETAC) level three.

C1 cohort: 15 students.

Programme: The Further Education and Training Awards Council (FETAC) level five.

Furthermore, these FETAC examinations formed the basis for the topics taught during the lessons and also the final examinations at the end of the twelve-week programme (see Appendix 5 for the CEFR can-do statements and Appendices 1 and 2 for the FETAC learner outcomes).

4.2.3 Ethics

The students were fully informed of the study and asked if they would be interested in taking part, all twenty four students agreed to take part and completed the accompanying speaker information sheet and consent form (see Appendix 4). Learners were fully aware that they were being recorded and that their written samples were being held for analysis. Learners had the option to withdraw from the study at any time in accordance with the norms of research integrity.

4.2.4 Data gathering

The researcher recorded entire three hour classes, every Tuesday evening, from September 2010 to April 2011. The spoken component includes oral presentations, oral examinations, classroom interactions (typical of a communicative classroom). The written component of the corpus comprises written assignments, homework and open-ended grammar tasks (see below for examples). In total, 42 hours of spoken data and 114 samples of writing were collected in the sampling period (see Appendix 3 for individual learner breakdowns).

All classroom recordings were gathered in audio mode only, using a Dictaphone. Learners were aware at all times that they were being recorded[6]. Though they had the choice to opt out from participating in the research, all 24 learners took part for the full duration of the study period. The data was stored securely at all times. All data was anonymised during the transcription process. Table 4.2 summarises the spread of data (by learner and mode)

Table 4.2 Spread data, by learner and by mode (spoken and written) in the sample period in the classroom in ACE

Speaker	Cohort	Spoken word count	Written word count
<$1>	A2	1,065	548
<$2>	C1	10,022	958
<$3>	C1	4,028	1020
<$4>	A2	4,012	697
<$5>	A2	3,426	867
<$7>	C1	5,245	1158
<$8>	C1	12,458	959
<$9>	C1	18,879	1059
<$10>	A2	3,425	717
<$11>	C1	3,246	1118
<$12>	C1	12,645	858
<$13>	A2	5,014	718
<$14>	A2	3,658	958
<$15>	C1	4,424	964
<$16>	A2	5,467	808

[6] While video recording would have been preferable, it was not permittable within this institutional context.

<$17>	C1	17,269	858
<$18>	C1	5,625	960
<$19>	C1	1,253	913
<$20>	A2	6,865	660
<$21>	C1	7,262	814
<$2 2>	C1	2,206	758
<$23>	A2	6,258	670
<$24>	C1	6,248	960
Totals		**150,000**	**20,000**

It is noted that some contributions are much greater than others. This is not unusual in any class but varying word counts across learners is also explained by inevitable absences from some of the 12 class sessions. For obvious reasons, it is also the case that C1 learners contributed longer turns and ultimately more spoken data.

4.2.5 Data Processing: transcription, sub-corpora formats and manual coding

In terms of the spoken data, all recordings were transcribed in full according to the transcription conventions in Appendix 6. The conventions are adapted from those used in the Cambridge and Nottingham Corpus of Discourse in English (CANCODE) (McCarthy, 1998; Carter, 2004) and the Limerick Corpus of Irish English (Farr, Murphy and O'Keeffe, 2004). Each learner was given a speaker tag or 'an anonymous identity' from $1 to $24. The tag then operated as an identifier for all spoken and written contributions from each student (see Table 4.2 and Appendix 3). In addition, the hand-written work was typed into electronic format (with no alternations to errors) and added to the relevant student file.

The next step in processing the data was to manually sift through it. This meant a close reading and re-read of all 150,000 words of spoken data and all 20,000 words written data so as to manually identify and highlight all occurrences of the four lexical features under examination (Multi-Wlord Verbs; Delexical Verbs; Collocations and Idioms). This was done in hard copy; items were highlighted through colour coding. This stage generated four separate lists that were tallied and stored in excel files for further analysis (see analysis chapters).

Because the focus of this study is on learner language, a number of steps have to be taken to conFig. the data into different folder formats and thus create sub-corpora. The first step in this process was to isolate each learner's contribution and save it as a text file, with an identifier that linked to the metadata. Depending on the type of analysis or comparison, different sub-corpora formats were needed:

Format 1: All learner data by learner

- 24 folders, i.e. one per learner into sub-folders labelled $1 to $24.
- Each of the 24 folders contained all spoken and written data for each learner together (12 spoken files and 12 written files per learner);
- These data could also be sorted into A2 and C1 levels.

Format 2: Spoken and written data by learner

- 2 folders: one for spoken and one for written data.
- Within each of these folders, learner data was contained in a sub-folder, labelled $1 to $24.
- These data could also be sorted into A2 and C1 levels.

Format 3: Spoken versus written data by learner

- 24 folders;
- Within each folder where to sub-folders, labelled by learner and mode, e.g. folder for learner <$1>: "<$1> spoken" (12 spoken files) and "<$1> written" (12 written files)
- These data could also be sorted into A2 and C1 levels.

The feature list identified manually and stored in excel could then be checked automatically across the whole dataset (format 1); the modes (format 2) and individual learners spoken versus written output (format 3). This was done using *Wordsmith Tools* (2016). As described below, the British National Corpus was used for comparison purposes and this was also scrutinised using *Wordsmith Tools.* Results could also be compared by level with the *English Vocabulary Profile* (see below).

Learner errors were not the focus of this study (in a L1 – L2 contrastive sense), however, it is noted that all attempts at the four lexical features under examination (Multi-Word Verbs; Delexical Verbs; Collocations and Idioms) were included in the first step of the analysis regardless of errors (including all generic errors such as spelling and pronunciation). In other words, the first stage of manual coding gathered all attempts. The next stage of sorting meant making decisions about what to include and what to exclude. This was done case by case within the following principles depending on which of the four areas were under consideration:

1) Phrasal verbs[7]: All attempts were included for initial scrutiny (manual extraction). Subsequent analysis excluded those where components were incorrect (i.e. incorrect verb, preposition or adverb particle). It did not exclude instances where the components were correct but where there was syntactic error (e.g. inflection, tense or aspect). For example, all items highlighted in extract 4.1 were included even though there are clearly tense and aspect errors:

Extract 4.1 (Student <$5> A2 spoken corpus)

… Me and friends we am yeah am we are **going in** you know beside the am park in the park but ah where there is big statue… yeah there and have lunch there and **go back home** and am see something and some sleep and at night **I wake up** it was very quiet so I we only **go** one pub **to** Flannerys.

[7] The term "phrasal verb" is used as an umbrella term for are three types: Prepositional Verb, Phrasal Verb and Phrasal Prepositional Verb. This is discussed in detail in Chapter 5.

However, instances such as extract 4.2 were not included because one of the components was incorrect:

Extract 4.2 (Student <$8> C1 written Corpus)

…you go away the street until you get to the bank.

2) Delexical Verbs: All attempts were considered in the original manual extraction. These were then examined closely and if the components were correct, they were included for analysis even if there were grammatical errors (e.g. tense, aspect). For example the following item was included even though there is a tense error because the components were correct (*went + holiday)*:

Extract 4.3 (Student <$7> C1 spoken corpus)

…Last week I **go** on am **holiday**

An example of an item that was excluded is shown in extract 4.4 where the components are incorrect:

Extract 4.4 (Student <$4> A2 spoken corpus

…I made homework and made breakfast.

An example of a challenging example is found in Extract 4.5. Here student <$18> from C1 cohort uses the delexical verb *take* with *lotto.* While this is not a conventionalised pattern, it was communicatively successful in this situation and so it was included as it made sense (in the context of collecting money for a class lotto).

Extract 4.5 (Student <18> C1 spoken corpus)

… it's okay if you want **to take a lotto** I don't mind.

3) Collocations: Again all collocational attempts were considered when undertaking the first extraction manually. All items with grammatical errors were included if the collocates were correct. This included where tense or aspect error occurred but also where there were syntax errors as in Extract 4.6:

Extract 4.6 (Student <$4> A2 written corpus)

…Hi Carolina, for English class I need **write letter** to you and tell you how my is my life here in the Ireland

4) Idioms: Again all attempts at idiomatic language were considered when undertaking the first extraction manually. All items with grammatical errors were included if the idiom components were correct (e.g. tense and aspect errors). Translated idioms were also included (see chapter 8 for further discussion on this). Such items were discussed with learners to find out their intended meaning (which was logged). During class and as an introduction to the theme of idiomaticity, the teacher asked the students to give examples of idioms in their L1, the idea was to highlight to the students how knowledge of the language or at least the culture is necessary to understand the idiom, for instance:

Extract 4.7 (Student <$24> C1 spoken corpus)

…In Spanish, my language people say expression ***when a frog grows hair.*** It means roughly something is impossible, will cannot happen.

[It is roughly translated into English as something that will never happen, similar to the expression: *when pigs fly*]

Many of the items across the four areas of focus in the present study proved problematic to categorise. When this happened, the researcher consulted with his supervisors for assistance in categorisation. For example, the phrasal verb *go on holiday* occurred and the researcher had to check if this would be included in the phrasal verb chapter or the delexical (it was decided on the latter). A further example was the verb + noun collocation *ask a question* where it had to be checked whether this was an example of a delexical verb (as we discuss in chapter 6, this was included in the delexical verbs listing).

4.2.6 Writing tasks

Students from both A2 and C1 cohorts submitted written work based on the following themes (some samples from the dataset are provided in Appendix 7):

- Letter home

- Complete a job application and Curriculum Vitae
- Communication task based on a text
- Write a fairy tale that you have heard in Ireland or your home country
- Fill in form (e.g. medical card applications, library membership applications)
- Write about something that happened in the past
- Letter giving advice
- Write a narrative based on a picture story
- Predictive text activity (using folder paper, learner adds new text)
- Grammar gap fill tasks

Having described the corpus data and how it was processed, we will now focus on the methodological framework of the study.

4.3 Methodological Framework

This study is based on the multi-word lexical features used by the two cohorts of ESL learners. As a result, the data was first manually examined in order to ascertain cursorily the presence or absence of the core lexical features in the speech and writing of the learners at ACE. Following that, the corpus analytical software *Wordsmith Tools* (Scott, 2016) was then used to compile word frequency lists, concordance lines, keyword and cluster lists. Through these functions, *Wordsmith Tools* in addition to manual searches of concordances and source files, aided the analysis of learner use of the four multi-word lexical features under consideration in this study: multi-word verbs, delexical verbs, collocations and idiomatic expressions.

We will now summarise the range and role of the functionality of the corpus software used in this research: wordlists, keywords and concordances.

4.4 Wordlists

Simply put, a wordlist is a list of all of the words that appear in a corpus. These words are usually arranged alphabetically or in order of frequency and through corpus linguistic software, such as *Wordsmith Tools* (Scott, 2016), it is possible to gather information about the number of times a given word occurs in a corpus. Such wordlists are essential to identifying the core words most used in particular situations and are thus an integral part of the corpus analysis procedure.

Fig 4.1 Screenshot of Wordlist in ACE corpus, alphabetically ordered.

A	737	1.53	135	84.91
ABLE	5	0.01	1	0.63
ABOUT	138	0.29	67	42.14
ABOVE	3		1	0.63
ABROAD	2		1	0.63
ABSOLUTE	4		4	2.52
ABSOLUTELY	8	0.02	8	5.03
ACADEMY	2		1	0.63
ACCENT	172	0.36	6	3.77
ACCEPTED	4		4	2.52
ACCORDING	3		1	0.63
ACCOUNTANCY	4		4	2.52
ACROSS	3		1	0.63
ACT	2		2	1.26
ACTING	3		3	1.89
ACTION	2		1	0.63
ACTIONEER	4		4	2.52
ACTIVE	2		1	0.63
ACTOR	4		4	2.52
ACTORS	5	0.01	5	3.14
ACTUALLY	16	0.03	5	3.14
ADDITION	2		1	0.63
ADDITIVE	2		2	1.26

Wordlist frequency is generally used to compare a minimum of two corpora and such comparison enables us to compare and contrast different language varieties and types of language in terms of frequency of search items. Hunston (2006) states that frequency information is not really informative unless it is comparative. Therefore, for the purposes of this study, the wordlist from the ACE corpus are compared with the wordlist from the British National Corpus (BNC) (see Burnard 2007). The BNC is a 100 million word collection of authentic spoken (10%) and written (90%) British English

data; it was designed to represent a wide cross-section of British English from the later part of the 20th century, both spoken and written (ibid). The use of the BNC as a comparative corpus is justified because it is seen as a baseline of how English is used in a native speaking context and the target to which the learners of English aspire to. It is also a highly representative corpus, as detailed in Clear (1993) and Crowdy (1995) and the core textbooks used in the classroom are based solely on British English (New English File 2010 and Advanced Headway 2010 for example).

By way of illustration of the comparative results that can be generated using *Wordsmith Tools,* table 4.3 below presents a comparison of the most frequent 20 words of the ACE and the BNC corpora:

Table 4.3 Comparison of the top 20 most frequent words of the ACE and BNC corpora

N	**ACE**	**BNC**
1	I	THE
2	THE	AND
3	AND	I
4	AM	TO
5	YOU	YOU
6	IS	A
7	IN	IT
8	TO	THAT
9	YEAH	OF
10	IT	IN
11	LIKE	IS
12	OF	YEAH
13	HAVE	ON
14	NO	WE

15	MY	WAS
16	FOR	THEY
17	THEY	HAVE
18	SO	IT'S
19	ARE	WHAT
20	WAS	FOR

As seen in Table 4.3 above, the words listed in the ACE corpus also appear in the native speaker corpus, the BNC, however, they are ordered differently as the frequencies of occurrence of the words are different. For example, the most frequent word in the non-native ACE corpus is the personal pronoun *I*, whereas the most frequent word in the BNC is the definite article *the* (this is not surprising given the dominance of spoken data in the ACE and the reverse in the BNC). The second person personal pronoun *you* is the fifth most frequent word in both corpora. This type of comparative analysis can bring to differences in terms of forms and contextual patterns and therefore the analysis of wordlists was an important part of the methodology in this study.

4.5 Keyword lists

A further feature of *Wordsmith Tools* (Scott, 2016) in this analysis is the keyword list function. These keywords are not the 'most important words' in the corpus (Scott, 1997) instead they indicate what Phillips (1989) referred to as 'aboutness'. Keyword lists identify positive keywords (the items that occur with a higher frequency) and negative keywords (the items that occur with a significantly lower frequency) when compared to larger reference corpora. For the purposes of this study, the BNC was used as a reference corpus. The keywords of the ACE, using the BNC as the reference corpus, do not necessarily represent the most frequent words of the corpus; instead, they show the most '**unusually frequent** words' (O'Keeffe et al, 2007). Table 4.4 below shows a sample of the keywords in the

ACE corpus using the BNC as the reference corpus. Note that in this table percentages refer to coverage within respective corpora (for example *am* accounts for 7.87% of all words in ACE and 0.02% of all words in the BNC).

Table 4.4 ACE corpus Keyword List using the BNC as a reference corpus

Word	**ACE Freq.**	**ACE %**	**BNC Freq.**	**BNC%**	**Keyness**
Am	720	7.87	2,045	0.02	236.1
Laugh	164	0.49	17	0	174.5
Time	146	0.44	3,097	0.03	484.2
My	310	0.93	22,92	0.24	380.4
Here	228	0.68	16,255	0.17	291.8
New	144	0.43	6,491	0.07	288.8
Like	338	1.01	37,124	0.39	232.6
Limerick	20	0.06	204.98	42	204.3
Okay	144	0.43	10,07	0.10	188.22
Uncountable	16	0.05	0	0	181.26

The keyword list presented in Table 4.4 above highlights that the ACE corpus has linguistic features that correspond to the main characteristics of the discourse in the English classroom, for example:

- The conjugated first person singular ***am*** of the verb *be* appears as the most frequent keyword when comparing ACE with the reference BNC corpus. It appears with a normalised frequency of 720 in the ACE corpus and 2,045 in the BNC (see Fig. 4.2 below for a concordance of *am*).
- The response token *okay* occurs as a keyword in ACE. This token frequently represents student agreement in the ACE corpus.

- The adjective *uncountable* appears as the 10th keyword in the ACE corpus with a frequency of 16 and it does not appear as a keyword in the BNC. However, the recurrence of the word *uncountable* as a keyword in the ACE corpus reflects the topic of some classes as countable and uncountable nouns were taught to both groups.

Not surprisingly, the word *Limerick* has come up as a keyword in the example above as it was the location of the study.

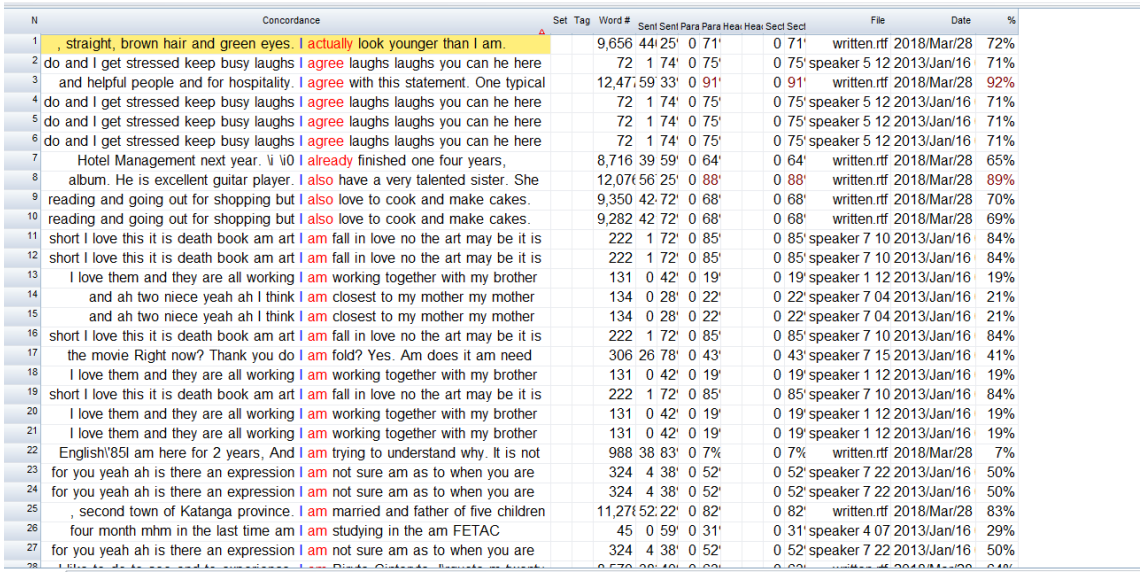
4.6 Concordance lines

Concordance lines are the results of a corpus search, using the software, for a pre-defined word or phrase. These show every occurrence of that search item in the centre of the computer screen (the *node*) and also allow for the sorting of the words that come before or after it. This can aid the identification of language patterns in the corpus. The identification of lexical patterns, such as collocations was imperative to this study.

Concordance lines enable the researcher to investigate the meaning of words but also the behaviour of items as they occur in a sentence. Such concordance lines clearly highlight the other words which a unit collocates with, but can also serve to highlight the semantics or pragmatics of an item in the context in which it occurs in the corpus. According to Kennedy (1998: 256), corpus linguistics has 'brought many contributions to the field of lexicography, for example, the uses of concordance as a source to provide authentic examples of the use of words.'

As mentioned above, concordance lines play an intrinsic role in the present study; in each chapter there are concordance lines to indicate the re-occurrence of a lexical item. The following figure. is an example of a concordance line for the most frequent word in the ACE corpus, the personal pronoun *I.*

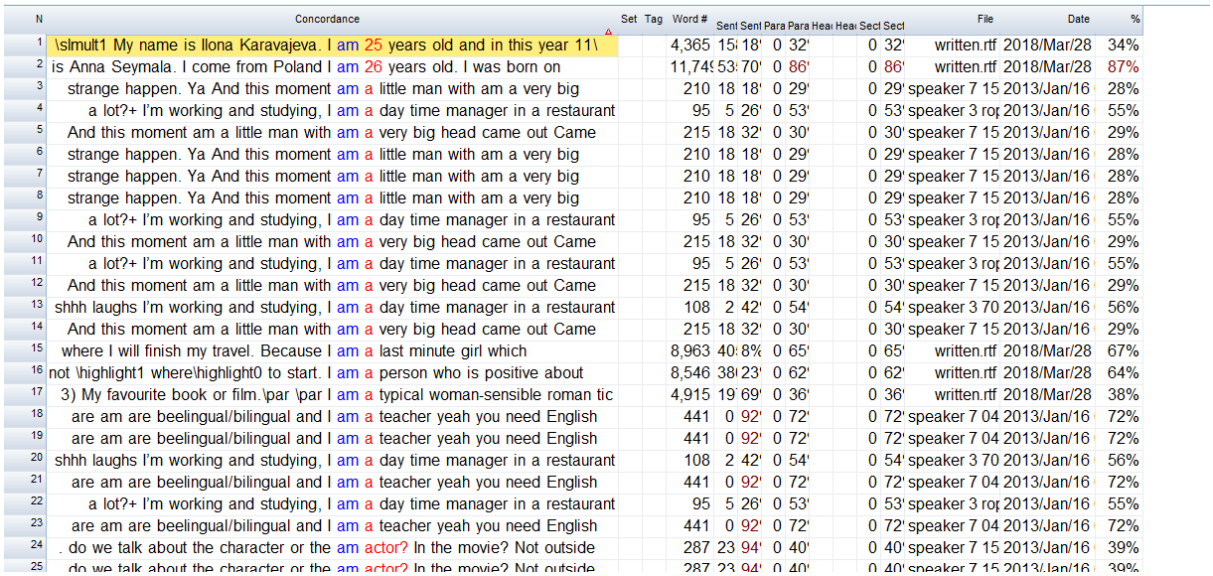
Fig. 4.2 Concordance lines for the word *I* in the ACE corpus.



N	Concordance	Set	Tag	Word #	Sent	Sent	Para	Para	Head	Head	Sect	Sect	File	Date	%
1	, straight, brown hair and green eyes. I actually look younger than I am.			9,656	44	25'	0	71'		0	71'		written.rtf	2018/Mar/28	72%
2	do and I get stressed keep busy laughs I agree laughs laughs you can he here			72	1	74'	0	75'		0	75'		speaker 5 12	2013/Jan/16	71%
3	and helpful people and for hospitality. I agree with this statement. One typical			12,47	59	33'	0	91'		0	91'		written.rtf	2018/Mar/28	92%
4	do and I get stressed keep busy laughs I agree laughs laughs you can he here			72	1	74'	0	75'		0	75'		speaker 5 12	2013/Jan/16	71%
5	do and I get stressed keep busy laughs I agree laughs laughs you can he here			72	1	74'	0	75'		0	75'		speaker 5 12	2013/Jan/16	71%
6	do and I get stressed keep busy laughs I agree laughs laughs you can he here			72	1	74'	0	75'		0	75'		speaker 5 12	2013/Jan/16	71%
7	Hotel Management next year. \i \i0 I already finished one four years,			8,716	39	59'	0	64'		0	64'		written.rtf	2018/Mar/28	65%
8	album. He is excellent guitar player. I also have a very talented sister. She			12,07	56	25'	0	88'		0	88'		written.rtf	2018/Mar/28	89%
9	reading and going out for shopping but I also love to cook and make cakes.			9,350	42	72'	0	68'		0	68'		written.rtf	2018/Mar/28	70%
10	reading and going out for shopping but I also love to cook and make cakes.			9,282	42	72'	0	68'		0	68'		written.rtf	2018/Mar/28	69%
11	short I love this it is death book am art I am fall in love no the art may be it is			222	1	72'	0	85'		0	85'		speaker 7 10	2013/Jan/16	84%
12	short I love this it is death book am art I am fall in love no the art may be it is			222	1	72'	0	85'		0	85'		speaker 7 10	2013/Jan/16	84%
13	I love them and they are all working I am working together with my brother			131	0	42'	0	19'		0	19'		speaker 1 12	2013/Jan/16	19%
14	and ah two niece yeah ah I think I am closest to my mother my mother			134	0	28'	0	22'		0	22'		speaker 7 04	2013/Jan/16	21%
15	and ah two niece yeah ah I think I am closest to my mother my mother			134	0	28'	0	22'		0	22'		speaker 7 04	2013/Jan/16	21%
16	short I love this it is death book am art I am fall in love no the art may be it is			222	1	72'	0	85'		0	85'		speaker 7 10	2013/Jan/16	84%
17	the movie Right now? Thank you do I am fold? Yes. Am does it am need			306	26	78'	0	43'		0	43'		speaker 7 15	2013/Jan/16	41%
18	I love them and they are all working I am working together with my brother			131	0	42'	0	19'		0	19'		speaker 1 12	2013/Jan/16	19%
19	short I love this it is death book am art I am fall in love no the art may be it is			222	1	72'	0	85'		0	85'		speaker 7 10	2013/Jan/16	84%
20	I love them and they are all working I am working together with my brother			131	0	42'	0	19'		0	19'		speaker 1 12	2013/Jan/16	19%
21	I love them and they are all working I am working together with my brother			131	0	42'	0	19'		0	19'		speaker 1 12	2013/Jan/16	19%
22	English\'85I am here for 2 years, And I am trying to understand why. It is not			988	38	83'	0	7%		0	7%		written.rtf	2018/Mar/28	7%
23	for you yeah ah is there an expression I am not sure am as to when you are			324	4	38'	0	52'		0	52'		speaker 7 22	2013/Jan/16	50%
24	for you yeah ah is there an expression I am not sure am as to when you are			324	4	38'	0	52'		0	52'		speaker 7 22	2013/Jan/16	50%
25	, second town of Katanga province. I am married and father of five children			11,27	52	22'	0	82'		0	82'		written.rtf	2018/Mar/28	83%
26	four month mhm in the last time am I am studying in the am FETAC			45	0	59'	0	31'		0	31'		speaker 4 07	2013/Jan/16	29%
27	for you yeah ah is there an expression I am not sure am as to when you are			324	4	38'	0	52'		0	52'		speaker 7 22	2013/Jan/16	50%

The most frequent keyword, identified in table 4.4 can also be concordanced, as illustrated in Fig. 4.3:

Fig. 4.3 Concordance lines for the top key word, *am,* the in the ACE corpus



N	Concordance	Set	Tag	Word #	Sent	Sent	Para	Para	Head	Head	Sect	Sect	File	Date	%
1	\slmult1 My name is Ilona Karavajeva. I am 25 years old and in this year 11\			4,365	15	18'	0	32'		0	32'		written.rtf	2018/Mar/28	34%
2	is Anna Seymala. I come from Poland I am 26 years old. I was born on			11,74	53	70'	0	86'		0	86'		written.rtf	2018/Mar/28	87%
3	strange happen. Ya And this moment am a little man with am a very big			210	18	18'	0	29'		0	29'		speaker 7 15	2013/Jan/16	28%
4	a lot?+ I'm working and studying, I am a day time manager in a restaurant			95	5	26'	0	53'		0	53'		speaker 3 ro	2013/Jan/16	55%
5	And this moment am a little man with am a very big head came out Came			215	18	32'	0	30'		0	30'		speaker 7 15	2013/Jan/16	29%
6	strange happen. Ya And this moment am a little man with am a very big			210	18	18'	0	29'		0	29'		speaker 7 15	2013/Jan/16	28%
7	strange happen. Ya And this moment am a little man with am a very big			210	18	18'	0	29'		0	29'		speaker 7 15	2013/Jan/16	28%
8	strange happen. Ya And this moment am a little man with am a very big			210	18	18'	0	29'		0	29'		speaker 7 15	2013/Jan/16	28%
9	a lot?+ I'm working and studying, I am a day time manager in a restaurant			95	5	26'	0	53'		0	53'		speaker 3 ro	2013/Jan/16	55%
10	And this moment am a little man with am a very big head came out Came			215	18	32'	0	30'		0	30'		speaker 7 15	2013/Jan/16	29%
11	a lot?+ I'm working and studying, I am a day time manager in a restaurant			95	5	26'	0	53'		0	53'		speaker 3 ro	2013/Jan/16	55%
12	And this moment am a little man with am a very big head came out Came			215	18	32'	0	30'		0	30'		speaker 7 15	2013/Jan/16	29%
13	shhh laughs I'm working and studying, I am a day time manager in a restaurant			108	2	42'	0	54'		0	54'		speaker 3 70	2013/Jan/16	56%
14	And this moment am a little man with am a very big head came out Came			215	18	32'	0	30'		0	30'		speaker 7 15	2013/Jan/16	29%
15	where I will finish my travel. Because I am a last minute girl which			8,963	40	8%	0	65'		0	65'		written.rtf	2018/Mar/28	67%
16	not \highlight1 where\highlight0 to start. I am a person who is positive about			8,546	38	23'	0	62'		0	62'		written.rtf	2018/Mar/28	64%
17	3) My favourite book or film.\par \par I am a typical woman-sensible roman tic			4,915	19	69'	0	36'		0	36'		written.rtf	2018/Mar/28	38%
18	are am are beelingual/bilingual and I am a teacher yeah you need English			441	0	92'	0	72'		0	72'		speaker 7 04	2013/Jan/16	72%
19	are am are beelingual/bilingual and I am a teacher yeah you need English			441	0	92'	0	72'		0	72'		speaker 7 04	2013/Jan/16	72%
20	shhh laughs I'm working and studying, I am a day time manager in a restaurant			108	2	42'	0	54'		0	54'		speaker 3 70	2013/Jan/16	56%
21	are am are beelingual/bilingual and I am a teacher yeah you need English			441	0	92'	0	72'		0	72'		speaker 7 04	2013/Jan/16	72%
22	a lot?+ I'm working and studying, I am a day time manager in a restaurant			95	5	26'	0	53'		0	53'		speaker 3 ro	2013/Jan/16	55%
23	are am are beelingual/bilingual and I am a teacher yeah you need English			441	0	92'	0	72'		0	72'		speaker 7 04	2013/Jan/16	72%
24	. do we talk about the character or the am actor? In the movie? Not outside			287	23	94'	0	40'		0	40'		speaker 7 15	2013/Jan/16	39%
25	do we talk about the character or the am actor? In the movie? Not outside			287	23	94'	0	40'		0	40'		speaker 7 15	2013/Jan/16	39%

Wordsmith Tools (Scott, 2016) also enables the researcher to return to the source file which contains the full transcription any occurrence, which in turn shows the context in which it is was used. This is crucial in disambiguating meaning and use of a given item.

4.7 Normalisation

When comparing corpora of different sizes the frequencies must be normalised so that we can compare 'like with like'. To do this, we convert frequency counts in our study corpora to a common base (e.g. per 100, 1,000, 10,000, 1,000,000); usually frequencies are compared per million words. For example, the preposition *about* occurs 98 times in the ACE corpus. If we want to compare it with other larger corpora, we must 'normalise it' to per million words. By dividing 98 by 170,000 (the total word count of the ACE corpus) and then multiplying it by 1 million we get 576.4 per million words. This allows us to compare findings with other often larger corpora such as the comparative corpus the BNC (100,000,000 words) where *about* occurs 34,674 times (346.74 times per million words).

4.8 Cluster analysis

Another important function of Wordsmith Tools (2016) that was used in this study was the 'cluster analysis' function. Biber et al (1999) refer to this language phenomenon as 'lexical bundles' whereas O'Keeffe et al (2007) refer to clusters as 'chunks'. Other terms include 'formulas, n-grams, fixed expressions, prefabricated patterns (or prefabs) (see Gray and Biber, 2015: 125). Either way, they all emphasise the same Neo-Firthian structure, a view of language accompanied by its collocations in either spoken or written genres.

To generate a cluster list using *Wordsmith Tools* (Scott, 2016), the researcher needs to first create a wordlist of the corpus. Next, the researcher needs to select *compute* and following that, select *make cluster option*. Then, the researcher must decide the number of combinations, for instance, 2, 3, 4, 5, or 6 words and any frequency cut-off point. The results will vary depending on the size of the corpus;

for example, if it is a small corpus like the ACE, it would be difficult to find any clusters exceeding five words. A further useful feature of a cluster analysis in Wordsmith Tools (Scott, 2016) is that the researcher can click on the lemma column and view the different language forms of the occurrence. An analysis of clusters provides valuable information about the distribution and behaviour of language units as it occurs in the contexts of different genres. In the case of the ACE corpus, a cluster analysis highlights the patterns of both real and authentic spoken interactions and the language features which appear in ESOL classroom discourse. By way of example of the cluster functionality of Wordsmith Tools (Scott, 2016), Table 4.5 below illustrates the top ten two-word clusters in the ACE corpus.

Table 4.5 the 10 most frequent two-word clusters in the ACE corpus (per million words)

N	**ACE**	**Frequency PMW**
1	do not	6,000
2	I am	4,305
3	of course	4,164
4	in the	3,035
5	I have	2,823
6	I think	2,682
7	it was	2,400
8	do you	2,400
9	you have	2,400
10	it is	2,320

In Table 4.5 above, the most frequent two word clusters are illustrated. It is clear from the analysis that the auxiliary verb *do* followed by the negative particle *not* is the highest occurring two word

cluster in the ACE corpus, with an occurrence of 6,000 per million words. On further inspection, it is clear that this cluster occurs mainly in a declarative negative sentence and is frequently a response to a closed question, for example (take from the ACE A2 cohort, spoken corpus):

Extract 4.8 (<$4> and <$12> A2 spoken corpus)

<$4>: do you play football?

<$12>: no. I ***do not.***

The second highest occurring cluster in the corpus is the affirmative personal pronoun followed by the conjugated verb to be *I am*; with a frequency of 4,305 per million words. Again, this is an example of a response token to a direct question asked by either the teacher or another learner. As illustrated in Table 4.5, the frequent appearance of pronouns such as *I, you, he* etc., indicate the nature of a casual conversation taking place. For example, if we look at the top ten two-word clusters, we can see that the majority contain a personal pronoun: *I am* is second, *I have* is fifth, *I think* is sixth and *you have* and *it is* are ninth and tenth correspondingly.

An interesting occurrence (which will be explained further in Chapter 7 (Collocations)) is the idiomatic collocation *of course*. This is third in the top ten clusters of the corpus, with a frequency of 4,164 per million words. The expression: *of course* occurred frequently in the corpus as a response token but also as a discourse marker. Extract 4.9 below shows a student (C1 cohort, spoken corpus) responding to the teacher's prompt, "So tell me something that you miss from you home country from the teacher which asked"

Extract 4.9 <$6> (C1 cohort, spoken corpus - oral interview)

…my family *of course*. I was in Poland in September and now I will go in the August for the holiday.

Table 4.6 illustrates the top 10 three-word clusters in the ACE corpus:

Table 4.6 10 most frequent three-word clusters in the ACE corpus (per million words)

N	ACE	Frequency
1	I would like	2,545
2	It is a	1,694
3	My name is	1,482
4	Do you have	1,376
5	There is a	1,164
6	Like to have	1,058
7	I have to	952
8	What do you	952
9	You have to	952
10	a lot of	847

Again, similar patterns occur in the three word clusters of the ACE corpus: *I would like*, *it is a, my name is* and *do you have* all occur in the top three, further highlighting the importance of pronouns in the ACE corpus. However, the high occurrence of requests such as *I would like* (2,545 per million words) and *do you have* (1,376 per million words) could, to some extent, be attributed to classroom task effect since the A2 cohort are required to demonstrate the ability to order food from a menu.

In addition, the clusters relating to spatial deixis, *there is a…* and *it is a*…, have a high occurrence with 1,694 and 1,164 per million words, respectively. A clear example of classroom language is the collocation *my name is*, which appears as the third most frequent cluster with a frequency of 1,482 per million words.

Table 4.7 illustrates the top 10 most frequent four-word clusters in the ACE corpus:

Table 4.7 Top 10 four-word clusters in the ACE corpus (per million words)

N	ACE	Frequency
1	And it's okay	4,121
2	I would like to	3,105
3	Are you pulling my	776
4	For the main course	776
5	What is your name	647
6	a little my family	517
7	a pity but we	517
8	a lot of people	517
9	Oh thanks a lot	517
10	am studying in the	517

As can be seen from Table 4.7, the four-word cluster list serves to highlight that the ACE corpus contains patterns of combinations characteristic of an ESOL classroom discourse. Firstly, once again, personal and possessive pronouns appear in the top ten clusters: *it, I, you, my* and *we,* for example. Secondly, the conjunction *and* appears at the top as a unit in the combination *and it's okay* with a frequency of 4,121 (per million words). Thirdly, the polite response token *thanks a lot* appears preceded by the exclamative *oh* with a frequency of 517 occurrences (per million words).

As already outlined, one of the core data sections of this research is learner use of idiomatic expressions. Evidence of a fixed expression appears in the form of *are you pulling my…* This is third in rank order of frequency and, after closer analysis, through concordancing, we see that it is followed by the object *leg*. These patterns will be central to this analysis.

In summary, word frequency lists, keyword lists, concordance lines and cluster lists are vital to the present study. In general, they demonstrate that the ACE corpus contains language features characteristic of authentic English language in use, both spoken and written. It has been shown that the chunks identified through a cluster analysis inform linguists, teachers and students of the patterns and behaviour of a language. As O'Keeffe et al state: 'One could reasonably posit that an overemphasis in language teaching on single words out of context may leave second language learners ill-prepared in terms of both their processing of heavily chunked input such as casual conversation, and of their own productive fluency (O'Keeffe et al, 2007: 63).

4.9 'In Search of the Unknown': The Observer's Paradox (Meyerhoff, 2006).

According to Labov (1972: 209), 'the aim of linguistic research in the community must be to find out how people talk when they are not being systematically observed; yet we can only obtain these data by systematic observation'. This contradiction, oxymoron, or 'paradox' is what Labov (1966) referred to as the **Observer's Paradox.** At the core of linguistic research is the systematic collection of authentic language, as it is really used in various facets of society. Such a gathering of data is, in most cases, impossible without the use of recording equipment. Recording devices are habitually used by language researchers to record interactions as they occur in natural and or specified interactions. It has been argued by many (Labov, 1966; 1972; Edge and Richards, 1998; Meyerhoff, 2006; Clancy, 2010, for instance), that a pre-existing knowledge of this procedure and indeed the equipment itself affects (sometimes majorly) the authenticity of the language produced by the test subjects. The observer's paradox relates to the role of both the researcher and the equipment in the linguistic interactions being examined. It raises the question as to whether the presence of the researcher and / or the equipment might cause participants to alter the language they would normally produce and therefore undermine the main goal of linguistic research, namely the collection of naturally occurring speech within a community. Labov (1966) maintained that the observer's paradox can question the validity of any sociolinguistic research as it could be argued that the language the

participants produce is not the same as the language they would produce if they were not being observed.

Furthermore, Heisenberg's (1926: 38) **Uncertainty Principle,** states '[therefore] we cannot observe something without changing it'. As mentioned throughout this present research, all ESL classes at ACE are conducted via the communicative language teaching method, with interactive activities such as roleplays and conversations at its core. On numerous occasions, the teacher (and therefore researcher) was called upon to take part in various activities in the language class. Within these classes, the observer frequently became the participant, a further paradox postulated by Sarangi (2002). The duality in the role of this current researcher is, by Sarangi, divided into two paradoxes ***1) the analyst paradox*** and ***2) the participant paradox***. Sarangi (2002) attempts to align both the analysist's and participant's perspectives so that insights from the participants can be used to inform the overall analysis. The multiple roles of this researcher as observer, researcher, teacher and also participant reduces the impact of this paradox to some extent. The researcher is a member of the hypothetical communities of practice of 'teacher', 'participant' and 'student' and, as a result, is in a position to inform the analytical process. As the researcher was the teacher of both classes the researcher had prior knowledge of the abilities of the learners. Furthermore the fact that the learners knew the teacher meant that they were not nervous during the data collection phase.

Obviously, many factors affect the language gathering and analysis stages of a study like the present research. One such element is what Bell (1976) referred to as the ***principle of formality***. Bell (1976) deduced that when the informant being recorded is aware that the recording is taking place, a context is inadvertently created, in which the speaker will pay conscious attention and make a conscious effort to the speech which he or she is contributing. This, Clancy (2010: 107) claims 'makes it difficult to observe the true vernacular'. It can be argued that the language attained by means such as recording is quite often an inauthentic representation of the language the same person would use in a non-recorded interaction. Since the primary goal of corpus linguistics, and even sociolinguistics, is to determine the real language being used by a group in a native speaker like situation, the principle of

formality creates a further paradox. Be that as it may, a lot of the language produced in the corpus was elicited and recorded during interactions in the classroom where the learners frequently took part in simulated roleplays such as going to a restaurant and asking for directions. As already acknowledged the learners' use of language outside of the classroom was impossible to measure so the writings they completed at home were highlighted as such and the various roleplays conducted in the classroom yielded a representation of how these learners use language. Without data from their daily lives, we cannot say for certain how they perform in real life situations outside of the classroom. Nonetheless, the classroom-based recordings give us some indication of what their authentic use of English in everyday life might be.

4.10 Analytical Framework

As discussed in detail in Chapter 1, the research questions addressed in this study are:

- **Main Research Question:**

  To what degree can adult learners of ESL use multi-word lexical items (multi-word verbs, delexical verbs, collocations and idiomatic/figurative language) in their speaking and writing?

  Sub-Questions:

  i) What is the difference between the lexical competence of the A2 level student and the C1 level student?

  ii) Which level (A2 or C1) use the majority of multi-word units and language strings in their speaking and writing?

  iii) Is there a progression from more high-frequency transparent strings towards low frequency opaque strings?

As this chapter has detailed, the English language students at A2 and C1 levels of competency were recorded for the 12-week course period and their written work was also collected. Classroom interactions, pair work, individual work, group work presentations along with written essays, projects and examinations were all recorded to compile the corpus. The analysis chapters of this thesis will focus on the following areas (see above for information on selection of these items):

Chapter 5: Multi-Word Verbs

The Multi-word Units (i.e. phrasal verbs, prepositional verbs, and phrasal prepositional verbs) used by the students of both levels were analysed in detail. The analysis audits whether and how students at both levels use these multi-word units.

Chapter 6: Delexical Verbs

This chapter analysed learners' use of delexical combinations across levels. These structures, where the lexical power of the verb is removed and attributing it to the object has been indicative of proficiency learner level and even native speaker like language. This chapter allows us to audit whether and to what degree students at both levels use such components.

Chapter 7: Collocations

The use of collocations, for example, *blond hair* and *strong tea,* are the focus of the analysis in Chapter 7. Collocations have been proven in the literature to be another component of high level proficiency. This chapter will investigate whether and how collocational patterns across both cohorts.

Chapter 8: Idiomatic Expressions

The use of idioms is explored in Chapter 8. Such language is also an indicator of advanced proficiency and this chapter will appraise the degree to which they are used across level the two cohorts as well as the nature of this use.

4.11 Conclusion

In this chapter, the data used in the present study, the ACE corpus was described in terms of the participants, their school and their placement in levels. It detailed the corpus design in terms of the collection of spoken (class) data and written assignments. It explained how the corpus data was transcribed and how it was subsequently manually annotated so as to extract all instances of the four lexical features under examination: multi-word verbs, delexical verbs, collocations and idiomatic expressions. A description was also provided of the processes of selection of items through close consideration and categorisation.

Following the initial description of the ACE corpus and the outline of the compilation process, I focused on techniques relevant to the analysis for example *Wordsmith Tools* (Scott, 2016) was used to create the corpus word frequency lists, keyword lists, concordance lines, and clusters analysis. Finally, the analysis chapters were outlined in terms of their focus.

In the next chapter, we move to the first analysis phase of the study and we look at multi-word verbs. We first define and explain them as language phenomenona. As previously mentioned in Chapter 3 Conceptual Framework, such verbs are considered a language feature of low level learners as they are frequently taught at A1 level, but they are proven to be a high frequency exponent in native speaker, spoken language. Thus, it will be interesting to investigate whether they are used by both the A2 and C1 cohorts of the ACE.

Chapter 5 Multi-Word Verbs

Supposing is good, but **finding out** is better
(De Voto, 1922)
Phrasal Verbs…the scourge of the learner
(Sinclair, 1996)

5.0 Introduction

This chapter is the first data analysis chapter and focuses on multi-word verbs across the three core types of: 1) phrasal verbs 2) prepositional verbs and 3) phrasal prepositional verbs (see Carter and McCarthy, 2006). In the literature, all three are frequently referred to as multi-word verbs but it is commonplace in pedagogical works to find all three types referred to as 'phrasal verbs' (McCarthy, 2007; Soars and Sayer, 2003*;* Oxenden and Koening, 1997). This chapter will analyse students' examples across the three aforementioned categories of phrasal, prepositional and phrasal prepositional verbs, spoken and written by learners of both the A2 and C1 cohorts in the ACE corpus.

5.1 Multi-Word Verbs Description

First of all, multi-word verbs are exponents of multi-word units or lexical chunks and that is why they are of interest in this study, as discussed in chapter 3. They occur as patterns comprising a lexical verb + particle(s). The particle is normally an adverb or a preposition, or in the case of phrasal prepositional verbs both. Meaning-wise, they are either transparent or opaque. It has been clearly shown that such multi-word verbs are used on a particularly frequent basis by native speakers of English, and are further evidence of native speaker like fluency (Sinclair, 1991; Timmis, 2003; McCarthy, 2000, 2007). As a result of their high frequency it could be argued that in order to communicate successfully in English, a learner needs to be able to use such verbs accurately.

Multi-word verbs are particularly common in informal registers but they do tend to occur rather frequently in Academic English also (see Cornell, 1985, for further discussion). They are found in popular culture (e.g. songs, movies, television series, magazines) as well as in newspaper headlines. They are also frequent components of everyday language in use (see McCarthy, 1998). Such verbs

are encountered early in the English as a foreign language learning process with various course materials, such as *New English File* (Oxenden and Koening, 2016) teaching them from a Pre-Intermediate (A2) level. In addition, the *English Vocabulary Profile* (EVP) (Capel, 2012) shows examples of multi-word verbs use as occurring at A1 and A2 levels (e.g. *look out* and *come back* respectively). Accurate and frequent use remains to be an indication of proficiency in the language as indicated by the English Grammar Profile (O'Keeffe and Mark, 2017) and within the International English Language Teaching Systems (IELTS) examination including them in its Band 8 and 9 (C2) descriptors. Within applied linguistics, many have illustrated the prevalence of multi-word verbs in native speaker use and in the language of expert users of English (see Moon, 1997), McCarthy, 1998, Wray, 2000; 2002; Schmitt 2004 and O'Keeffe et al, 2007). Additionally, there are, at this time, numerous dictionaries dedicated to the topic (Oxford, 2001; Cambridge, 2004; McGraw Hill, 2005) and also teaching course books such as Sinclair and Moon (1989), McCarthy and O'Dell (2004), Clandfield (1998), Barlow (2006), for example. Furthermore, the University of Cambridge Local Examinations Syndicate (UCLES) Preliminary Test's Vocabulary List lists 81 core multi-word verbs, all of which are intended to be used by test takers in a productive manner. This chapter will assess the form, meaning and use of the multi-word verbs across the two cohorts (A2 and C1) of learners across the spoken or written corpus data.

5.2 Identifying and classifying types of multi-word verbs

In terms of nomenclature, there is considerable variation. In addition to *multi-word units* and the umbrella EFL use of *phrasal verbs,* some of these include following:

- Verb-Adverb Combination – (Kennedy 1920)
- Wordsets – (Ralph 1964)
- Discontinuous Verb – (Live 1965)
- Verb-Particle Construction –(Lipka 1972)

- Verb-Particle Combination – (Fraser 1974)
- Two-Word Verb – (Meyer 1975)

In this chapter, the definition of multi-word verbs from Carter, McCarthy, Mark and O'Keeffe (2011) is used which states that multi-word verbs are verbs which consist of a verb and one or two particles or prepositions (e.g. *up, over, in, down*) and that these can be broken down into three types of multi-word verbs: phrasal verbs, prepositional verbs and phrasal-prepositional verbs (see also Courtney, 1983). When conducting my analysis, I was guided by Riordan (2012) who notes that the most simple of the three to identify is a *phrasal prepositional verb*. This is a unit or chunk which combines both 1) a verb, 2) an adverb and 3) a preposition. In order to be classified as a phrasal prepositional verb, the example there must contain a verb plus both an adverb and a preposition (e.g. *look forward to*). Differentiating between *prepositional verbs* and *phrasal verbs* can be more challenging and Riordan's (2012) insight was used as a guide: if the unit does not take an object and is therefore intransitive, then it is a phrasal verb, for example *go away*. However, if it is transitive, and has an object, one must look at its ability to separate. If it separates, then it is a phrasal verb, for example *put on your coat*, versus *put your coat on*, but if it is inseparable, then it is a prepositional verb, for example *we went out last night* versus **we went last night out*.

5.3 Corpus analysis of multi-word verbs: the BNC as a reference corpus

As discussed in chapter 4, the use of corpora as a linguistic tool to analyse language makes it easier to identify the core forms and functions of features such as multi-word verbs within a given dataset. However, when analysing such results it is important to have a means of comparison. As detailed in chapter 4, the British National Corpus is used in this study as the baseline or reference corpus. A collection of 100 million words, the BNC contains written and spoken language gathered in the late 20th century (Burnard, 2007). In order to understand first, how frequently multi-word verbs occur in English, I conducted an investigation on the BNC to see what multi-word verbs occurred and at what frequency. Table 5.1 below shows the top 20 most frequently occurring multi-word verbs in the

BNC. Column one shows the rank from one to twenty, column two gives the lexical verb, column three the particle which creates the multi-word construction, column four gives the raw frequency result and finally column five gives the normalised result per million words (PMW).

Table 5.1 Raw and PMW frequencies of Top 20 Multi-Word Verbs in the BNC

Rank	Verb	Particle	Frequency	PMW
1	go	On	14,903	149.03
2	carry	Out	10,798	107.98
3	set	Up	10,360	103.60
4	pick	Up	9,037	90.37
5	go	Back	8,065	80.65
6	come	Back	8,029	80.29
7	go	Out	7,688	76.88
8	point	Out	6,984	69.84
9	find	Out	6,619	66.19
10	come	Up	5,523	55.23
11	make	Up	5,469	54.69
12	take	over	5,420	54.02
13	come	Out	5,022	50.22
14	come	On	4,830	48.03
15	come	In	4,814	48.14
16	go	down	4,781	47.81
17	work	Out	4,703	47.03
18	set	Out	4,633	46.33
19	take	Up	4,608	46.08

20	get	back	4,552	45.52

(based on Gardner and Davies 2007)

Table 5.1 above clearly shows the most frequently occurring multi-word verb in the BNC corpus as *go on* with 14,903 (149 PMW) occurrences, while the twentieth most frequent in the corpus is *get back* with 4,552 instances (45.52 PMW). We will now look at the ACE data and how it was analysed.

5.4 Analysis: Multi-Word Verbs in the ACE Corpus

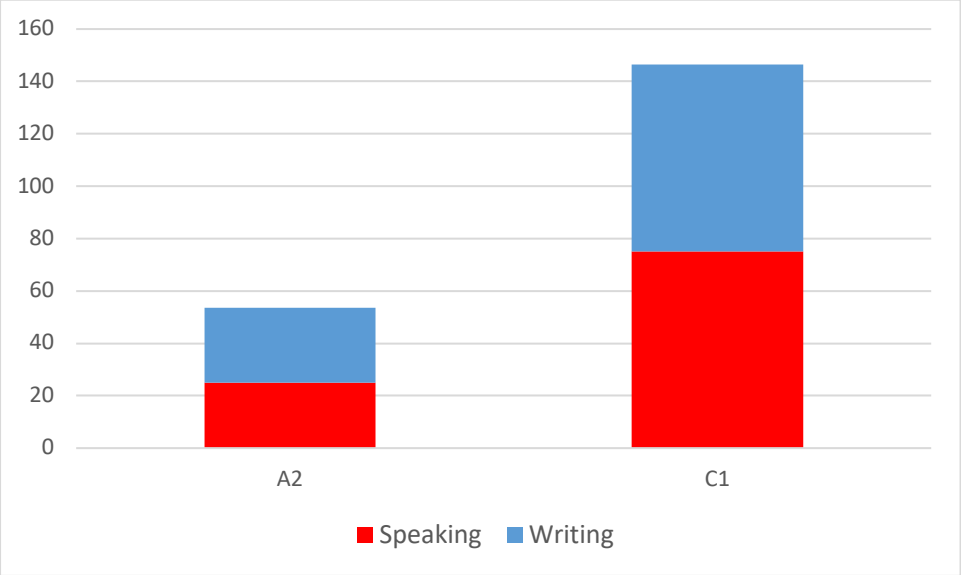
The main focus of this research is to identify what, if any, multi-word verbs the learners of the ACE corpus use in speaking and writing and to what degree. In order to do this, the researcher manually identified and coded the multi-word verbs (MWVs) in the corpus and then categorised them across the three types (discussed above). Table 5.2 shows the overall results in percentage of use across the cohort of participants:

Table 5.2 The number of individual learners using multi-word verbs in the ACE corpus

	Speaking		**Writing**	
	Raw	**%**	**Raw**	**%**
A2 (9 participants)	5	55	6	67
C1 (15 participants)	15	100	15	100
Total	20		21	

From Table 5.2, we can see that all C1 learners use MWs in both speaking and writing while more than half of the A2 participants use them (55% in speaking and 67% in writing). Fig. 5.1 below illustrates the distribution of multi-word verbs across both cohorts of the ACE corpus as a whole.

Fig. 5.1 Multi-word verbs use by individual learners in the ACE Corpus across A2 and C1 learners by percentage, across spoken and written sub-corpora.



When we examine the data more finely, we see that MWV are used more in speaking than in writing by both cohorts (Table 5.3):

Table 5.3: Breakdown of total number of occurrences of MWVs across levels and spoken versus written data (raw and PMW)

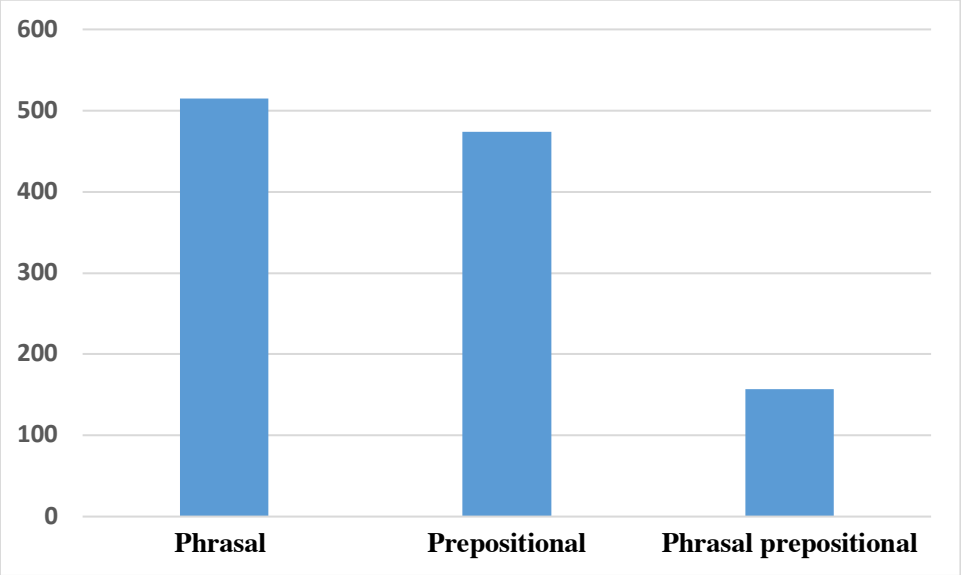
Level	**Spoken corpus**	**PMW**	**Written corpus**	**PMW occurrences**
A2	324	5,063	45	6,429
C1	522	6,070	255	19,615
Total	**846**		**300**	

Interestingly, A2 learners use MWVs 5,063 times PMWs in speaking but this rises to 6429PMWs in writing while C1 learners use them 6,070 PMWs in speaking and 19,615 times PMW in writing. In both cases, we see increase in the usage in writing but this can perhaps be explained by greater time to prepare for language output in a writing task (thus allowing more time to retrieve and display

lexical items). We also note a stark difference between the frequency of use of MWVs in C1 learner writing compared with speaking: learners are found to use MWVs over three times more frequently in writing. This frequency aligns with the English Grammar Profile for C1 learners which points to the ability to use a wide range of MWVs at C1 and C2 levels.

Figure 5.2 offers a breakdown of MWVs by type within ACE. Phrasal and prepositional verbs substantially outnumber phrasal prepositional verbs, which is not unsurprising.

Fig. 5.2 Breakdown of all uses of MWVs across three types



When we break these results down further, by level, we also see variation in use across level (Table 5.4).

Table 5.4: MWVs broken down by type and by level

Type	A2	%	C1	%	Total
Phrasal	194	39.03	321	49.46	**515**
Prepositional	232	46.68	242	37.29	**474**

Phrasal prepositional	71	14.29	86	13.25	**157**
Total	**497**	**100.00**	**649**	**100.00**	**1146**

Interestingly, the distribution of the types of MWVs is reasonably similar across the levels by percentage, especially in relation to phrasal prepositional verbs, which account for 14.29% of A2 MWVs and 13.25% of C1 uses. In terms of forms, Fig. 5.3 shows that the following five forms are used most by learners at both levels in the ACE corpus, by percentage**:**

Fig. 5.3 The five most frequent multi-word verbs in the ACE corpus (%)

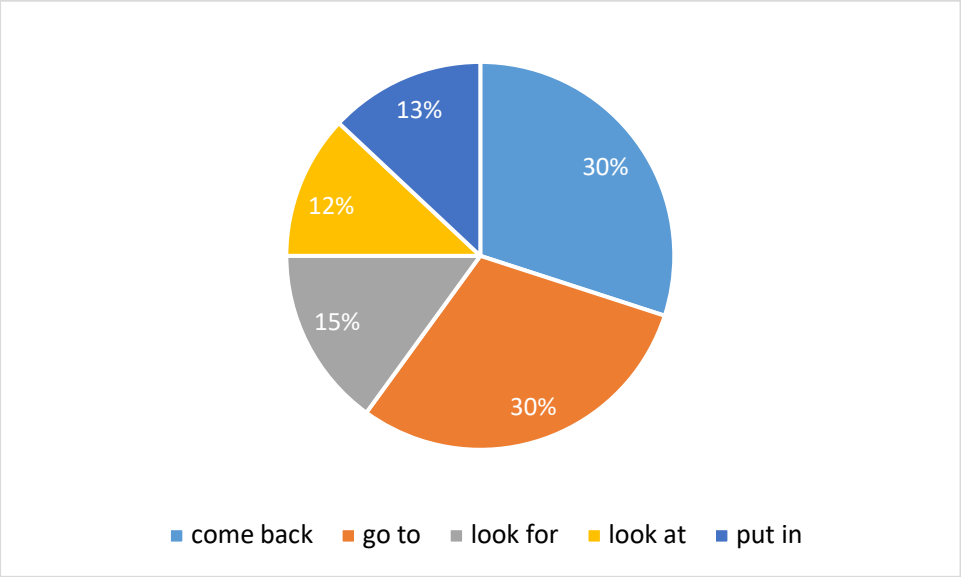


Table 5.5 provides a breakdown of these by raw frequency, by form. Here we can see some degree of variation in how they are distributed. For example *go to* is the most frequent form at A2 level while come back is the most frequent form at C1:

Table 5.5: Top five most frequent MWV forms across levels (raw frequencies).

Multi-word verb	**A2**	**C1**	**Total**
come back	264	501	765
go to	322	443	765
look for	116	266	382
look at	163	142	305
put in	123	205	328

Taking both cohorts together, there are 765 utterances of the expression: *come back* and the expression *go to*. The A2 cohort used the multi-word verb *come back* a total of 264 and the C1 group 501 times. In relation to *go to*, the A2 cohort articulated 322 examples while the C1 cohort gave 443**.** The multi-word verb *look for*, as the second most frequent multi-word verb in the ACE corpus overall with 382 occurrences, this breaks down as 116 examples in the A2 cohort and 266 in C1. The multi-word verb *look at* occurred 163 times in the A2 corpus and actually 142 times in the C1 group. Finally, there were 328 examples of *put in* in the total corpus: 205 in the C1 cohort and 123 in the A2.

The next step was to ascertain at what level on the CEFR scale these structures occurred so as to have a greater sense of what is normally expected in terms of the profile of use of these MWVs. In order to do so, the researcher searched the *English Vocabulary Profile* (Capel, 2012) for the multi-word verbs below[8]. The fourth column in the table outlines at what level in the CEFR and EVP they appear.

[8] See Chapter 4 Methodology for a full description of the English Vocabulary Profile (Capel, 2012).

Table 5.6 Occurrence of multi-word verbs in ACE, plus the CEFR level at which they appear in the English Vocabulary Profile (Capel, 2012).

Multi-Word Verb	Type	Occurrence (raw)	EVP CEFR
Come back	Phrasal	152	A2
Go to	Prepositional	152	A2
Look at	Prepositional	100	B2
Put on	Phrasal	100	A1
Put in	Phrasal	94	A1
Look forward to	Phrasal Prepositional	88	B1
Put down	Phrasal	47	A1
Wait for	Prepositional	47	A1
Take after	Prepositional	41	B2
Come in	Prepositional	41	A2
Go on up to	Phrasal Prepositional	41	B1
Come to	Prepositional	35	B2
Go on	Phrasal	35	B1
Give to	Prepositional	35	A2
Come from	Phrasal	23	A1
Come into	Prepositional	23	C2
Go straight ahead	Phrasal Prepositional	23	B1
Give back	Phrasal	11	A2
Come with	Phrasal	11	B2
Take leave	Phrasal	11	C1
Get Dressed	Phrasal	11	A2

Went off	Phrasal	5	B2
Go home	Phrasal	5	A1
Went home	Phrasal	5	A1
Got there	Phrasal	5	A1
Get out of	Phrasal Prepositional	5	B2

After manually coding and sorting all multi-word verbs found in the ACE corpus, the next step was to identify the twenty most frequent verbs in the spoken and written corpora, respectively. Table 5.6 below outlines the twenty normalised most frequent verbs in the ACE spoken corpus and shows at what level each occurred. From this table, we can see a general trend that a wider range of forms is used at C1 level.

Table 5.7 the top twenty most frequent multi-word verbs in the ACE spoken corpus

Lexical Verb	**Particle**	**RAW**	**Occurrence (PMW)**	**A2**	**C1**
come	back	89	524	37	52
go	to	89	524	44	45
look	for	76	447	36	40
look	at	74	435	0	74
put	on	72	424	0	72
put	in	70	412	20	50
put	down	67	394	0	67
wait	for	58	341	8	50
take	after	54	318	0	54
go	in	52	306	52	0

come	to	52	306	6	46
go	on	51	300	51	0
give	to	51	300	21	30
get	to	51	300	35	16
come	from	46	271	12	34
ask	about	46	271	37	0
depend	on	37	218	0	37
take	leave	37	218	0	37
check	in	37	218	9	28
go	off	37	218	0	37
Percentage				**60%**	**90%**

At this point, based on the quantitative summaries thus far, we can say that MWVs are used by both cohorts to varying degrees.

- A2 learners show use of MWVs of all types and this exceeds expectations based on the English Grammar Profile which profiles A2 learners as using limited ranges of phrasal and prepositional verbs and says that *look forward to* is the only phrasal prepositional verb that we can expect A2 learners to use competently (obviously in a routinised manner, typically at the end of a letter or email).
- C1 learners are profiled in the EGP as being able to use a wide range of MWVs and this is found to be the case in the ACE C1 data also.
- Of the top 20 most frequent forms in the ACE data, A2 learners show that they are using 60% of them while C1 learners use 90%. The result here for A2 exceeds expectations.
- We note that there is greater use of MWVs in spoken language and this relates to greater informality and also the varying requirements of written tasks (see below).

- Though there seems to be a higher degree of use of MWVs by A2 level learners than expected, we note that not all participants in the A2 cohort are using these forms (see Table 5.2).

We will now take a more qualitative look at the learners' use of MWVs in ACE so as to get more insight.

5.4.1 Come back

Come back occurred as the most frequent form in both cohorts and this in not unexpected as we see from table 5.6 where it is seen as typical of A2 learners in the EVP. The following concordance list, generated by *Wordsmith Tools* (Scott, 2016), illustrates the occurrences of the phrasal verb *come back* in the A2 sub-corpus. This structure accounted for 30% of all MWVs in the corpus. As can be seen from the concordance list below, all spoken examples are phrasal verbs as they are all intransitive.

Fig. 5.4 Concordance screenshot sample of *come back* (A2)

N	Concordance	Set	Tag	Word #	Sen	Sen	Para	Para	Hea	Hea	Sec	Sec	File	Date	%
1	it eh watchid/watched the parade and come back to am home a few friends			16	0	23	0	7%			0	7%	speaker 10 2	2013/Jan/22	8%
2	want I go to Dublin on Thursday and come back Saturday and next			194	1	74	0	43			0	43	speaker 7 54	2013/Jan/22	39%
3	it eh watchid/watched the parade and come back to am home a few friends			16	0	23	0	7%			0	7%	speaker 10 2	2013/Jan/22	8%
4	come back to am home a few friends come back and so some drink no			24	0	34	0	10			0	10	speaker 10 2	2013/Jan/22	10%
5	come back to am home a few friends come back and so some drink no			24	0	34	0	10			0	10	speaker 10 2	2013/Jan/22	10%
6	eh but he went to see am when he come back eh he eh and I forget that			132	0	43	0	27			0	27	speaker 18 0	2013/Jan/22	25%
7	train ah leave leave okay and when I come back to Limerick what time is			456	2	62	0	92			0	92	speaker 16 1	2013/Jan/22	91%
8	train ah leave leave okay and when I come back to Limerick what time is			456	2	62	0	92			0	92	speaker 16 1	2013/Jan/22	91%
9	a rich man eh he had to make Isabelle come back after five years and marry			91	0	29	0	19			0	19	speaker 18 0	2013/Jan/22	18%
10	my mam was really afraid ah I wouldn't come back from here so yeah yeah ah			78	2	69	0	21			0	21	speaker 5 83	2013/Jan/16	22%
11	my mam was really afraid ah I wouldn't come back from here so yeah yeah ah			78	2	69	0	21			0	21	speaker 5 83	2013/Jan/16	22%
12	can I ah can I do this I would like to come back ah am ah the next day			346	4	72	0	83			0	83	speaker 21 0	2013/Jan/22	83%
13	for this evening ah I would like to come back ah this evening am can			416	0	97	0	84			0	84	speaker 16 1	2013/Jan/22	83%
14	am oh what time is the last train to come back from Galway? yes this			483	3	80	0	89			0	89	speaker 7 04	2013/Jan/22	87%
15	am oh what time is the last train to come back from Galway? yes this			483	3	80	0	89			0	89	speaker 7 04	2013/Jan/22	87%
16	for this evening ah I would like to come back ah this evening am can			416	0	97	0	84			0	84	speaker 16 1	2013/Jan/22	83%
17	laughs yeah sure am no sorry I have to come back tomorrow so I take a day			487	4	54	0	98			0	98	speaker 16 1	2013/Jan/22	98%
18	laughs yeah sure am no sorry I have to come back tomorrow so I take a day			487	4	54	0	98			0	98	speaker 16 1	2013/Jan/22	98%
19	can I ah can I do this I would like to come back ah am ah the next day			346	4	72	0	83			0	83	speaker 21 0	2013/Jan/22	83%
20	ah I go to Spain ehmm but it's hard to come back am five days I ah I ah I			24	0	41	0	6%			0	6%	speaker 18 1	2013/Jan/22	5%
21	rich man am the day before that he will come back eh but he went to see am			121	0	39	0	25			0	25	speaker 18 0	2013/Jan/22	23%

Moreover, Table 5.8 below highlights the occurrences of *come back* across both cohorts in both spoken and written modes.

Table 5.8 *Come back* across both cohorts

Exponent	Type	Level	Medium
go to Dublin and **come back**.	Phrasal verb	A2	Spoken
watched the parade and **come back**.	Phrasal verb	A2	Spoken
a few friends **come back** and so some drinks	Phrasal verb	C1	Written
when I **come back** to limerick	Phrasal verb	C1	Spoken
he **come back** tomorrow	Phrasal verb	A2	Spoken
I will go to spain but it's hard to **come back**.	Phrasal verb	C1	Spoken

Though there are errors in tense and aspect in many of the above examples, all exponents are used with the intended meaning of a physical direction signalling a return. In the spoken instances above,

the multi-word verb *come back* reflects the informal nature of the conversation taking place and ties in with the quantitative findings that more MWVs are used in spoken contexts because of informality. It must be noted that were the MWV to be replaced with the regular lexical verb: *return,* then the conversation would appear less relaxed or informal and arguably stilted. We would therefore point to the importance of MWVs in spoken language among these ESL users in terms of maintaining casual friendly relations both between themselves and their teacher. Extract 5.1 below is an example drawn from an interaction between an A2 speaker and the teacher. Student <$7> is taking part in the midterm oral language examination. Here we can see that the A2 level student, even in this formal and high-stakes interaction, is able to use the informal MWV *come back* rather than *return* in the role play and this aligns with the Specific Learning Outcomes (SLO's) of the Further Education and Training Awards Council (FETAC) (see Appendix 5). The task is to roleplay the situation of purchasing a ticket at a train station. The researcher / teacher <T> plays the part of the teller:

Extract 5.1 (Student <$7> A2 spoken corpus)

<T> good afternoon…how can I help you?

<$7> how am good evening how are you? Ah am I would like to buy a ticket to Galway

<T> ok would you like a single or a return ticket?

< $7> yes I want a day return ticket … yes I would like to to return today for this evening I would like to **come back** ah this evening am

<T> of course…that is no problem. So a day return ticket to Galway that will be €7.45

<$7> ehmm am am what time is the next train?

<T> the next train leaves at 12.15 so in ten minutes

<$7> okay am oh what time is the last train to **come back** from Galway?

<T> yes this evening at 20.45.

<$7> ok thank you very much

<T> you are welcome have a nice trip

Here, the student takes the *return* prompt from the teacher's turn but on her second contribution, substituted the one word equivalent with the phrasal verb *come back*. It is important to note that the teacher never used the phrasal verb *come back.* In fact it is the student who opted to use the phrasal verb. The student is showing an ability to connect with the teacher/teller's use of *return* and then to relexicalize it as the more informal *come back* from thereafter. This aided in making the exchange less formal and more friendly resulting in the speaker sounding more native like and establishing social identity (McCarthy and O'Dell, 2004). This short exchange clearly illustrates that A2 level Student <$7> understands the meaning of the structure *come back,* and furthermore the lexical equivalent *return*.

In comparison, students from the C1 group were required to produce their first piece of writing in the form of an essay on the topic of 'my favourite film'. Student <$6> chose to describe the plot of the movie, *The Lion King*. (Note spelling and grammatical errors are included in this transcription).

Extract 5.2 (Student <$6> C1written corpus)

…Inded, Scar takes the power and makes the hyena live with the lions. But, the hyena became to powerfull and the eat everything without **care about** the cercal of life. Evry one is starving at this point and evrything is dying. The fact Nalla explains what is going on to simba makes him react, so they **come back** to their territory.

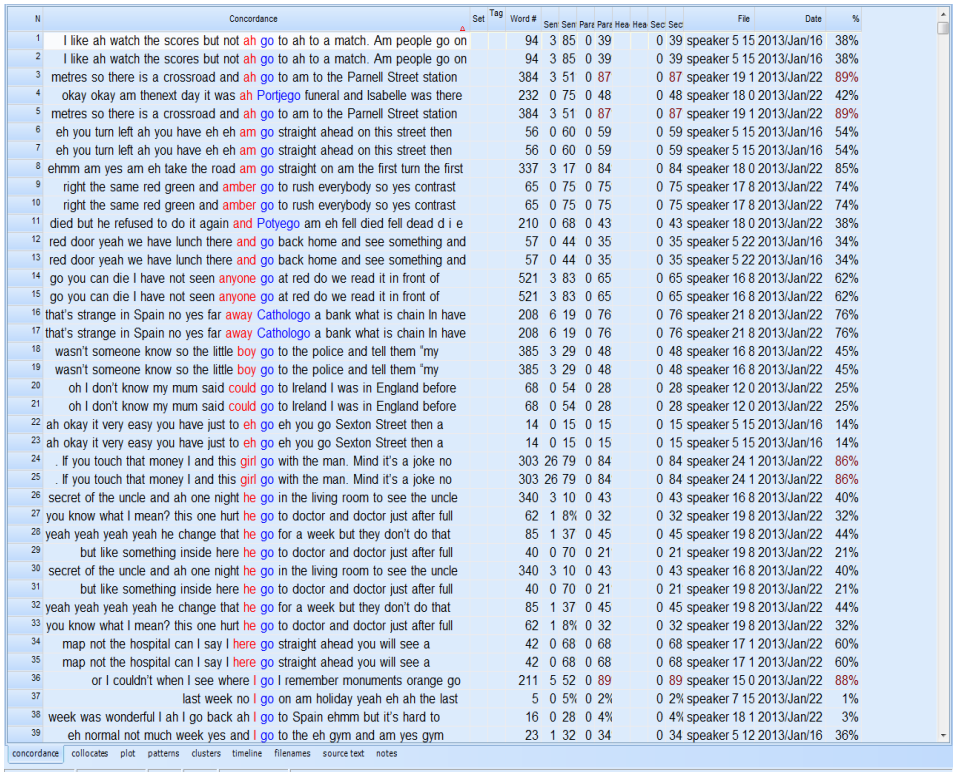
In this extract, there is clearly one prepositional verb *care about* and one phrasal verb *come back.* The intended meaning of *care about* cannot be conveyed correctly in any other way than through the use of this structure. The second multi-word verb in this text is the same verb used by Student <$7> in Extract 5.1 above, i.e. *come back*. Here the verb *come* and the adverb *back* are combined to mean a

physical action of returning to a familiar place (albeit without the correct tense marking). *Come back* functions here to make the writing more native and more informal. It could be argued that the writer opted for the more informal MWV form *come back* as opposed to the single word equivalent *return* even though this was a written task because it is not a formal academic writing task and that this shows a register-awareness on the part of the learner.

5.4.2 Go and its complements

We will now take a look at the high-frequency verb *go* and its complements that form many of the MWV items. As we can see from Table 5.7, the lexical verb *go* collocates with numerous particles in the overall ACE corpus. The phrasal verb *go to,* similar to the phrasal verb *come back,* accounts for 30% of all MWV occurrences, with 524 PMW uses in the spoken data. Fig. 5.4 below is a concordance list of the verb *go* in the corpus and gives an indication of its many patterns.

Fig. 5.5 Concordance screenshot of sample *go* + particle in C1 spoken sub-corpus



N	Concordance	Set	Tag	Word #	Sen	Sen	Para	Para	Hea	Hea	Sec	Sec	File	Date	%
1	I like ah watch the scores but not ah go to ah to a match. Am people go on			94	3	85	0	39			0	39	speaker 5 15	2013/Jan/16	38%
2	I like ah watch the scores but not ah go to ah to a match. Am people go on			94	3	85	0	39			0	39	speaker 5 15	2013/Jan/16	38%
3	metres so there is a crossroad and ah go to am to the Parnell Street station			384	3	51	0	87			0	87	speaker 19 1	2013/Jan/22	89%
4	okay okay am thenext day it was ah Portjego funeral and Isabelle was there			232	0	75	0	48			0	48	speaker 18 0	2013/Jan/22	42%
5	metres so there is a crossroad and ah go to am to the Parnell Street station			384	3	51	0	87			0	87	speaker 19 1	2013/Jan/22	89%
6	eh you turn left ah you have eh eh am go straight ahead on this street then			56	0	60	0	59			0	59	speaker 5 15	2013/Jan/16	54%
7	eh you turn left ah you have eh eh am go straight ahead on this street then			56	0	60	0	59			0	59	speaker 5 15	2013/Jan/16	54%
8	ehmm am yes am eh take the road am go straight on am the first turn the first			337	3	17	0	84			0	84	speaker 18 0	2013/Jan/22	85%
9	right the same red green and amber go to rush everybody so yes contrast			65	0	75	0	75			0	75	speaker 17 8	2013/Jan/22	74%
10	right the same red green and amber go to rush everybody so yes contrast			65	0	75	0	75			0	75	speaker 17 8	2013/Jan/22	74%
11	died but he refused to do it again and Potyego am eh fell died fell dead d i e			210	0	68	0	43			0	43	speaker 18 0	2013/Jan/22	38%
12	red door yeah we have lunch there and go back home and see something and			57	0	44	0	35			0	35	speaker 5 22	2013/Jan/16	34%
13	red door yeah we have lunch there and go back home and see something and			57	0	44	0	35			0	35	speaker 5 22	2013/Jan/16	34%
14	go you can die I have not seen anyone go at red do we read it in front of			521	3	83	0	65			0	65	speaker 16 8	2013/Jan/22	62%
15	go you can die I have not seen anyone go at red do we read it in front of			521	3	83	0	65			0	65	speaker 16 8	2013/Jan/22	62%
16	that's strange in Spain no yes far away Cathologo a bank what is chain In have			208	6	19	0	76			0	76	speaker 21 8	2013/Jan/22	76%
17	that's strange in Spain no yes far away Cathologo a bank what is chain In have			208	6	19	0	76			0	76	speaker 21 8	2013/Jan/22	76%
18	wasn't someone know so the little boy go to the police and tell them "my			385	3	29	0	48			0	48	speaker 16 8	2013/Jan/22	45%
19	wasn't someone know so the little boy go to the police and tell them "my			385	3	29	0	48			0	48	speaker 16 8	2013/Jan/22	45%
20	oh I don't know my mum said could go to Ireland I was in England before			68	0	54	0	28			0	28	speaker 12 0	2013/Jan/22	25%
21	oh I don't know my mum said could go to Ireland I was in England before			68	0	54	0	28			0	28	speaker 12 0	2013/Jan/22	25%
22	ah okay it very easy you have just to eh go eh you go Sexton Street then a			14	0	15	0	15			0	15	speaker 5 15	2013/Jan/16	14%
23	ah okay it very easy you have just to eh go eh you go Sexton Street then a			14	0	15	0	15			0	15	speaker 5 15	2013/Jan/16	14%
24	If you touch that money I and this girl go with the man. Mind it's a joke no			303	26	79	0	84			0	84	speaker 24 1	2013/Jan/22	86%
25	If you touch that money I and this girl go with the man. Mind it's a joke no			303	26	79	0	84			0	84	speaker 24 1	2013/Jan/22	86%
26	secret of the uncle and ah one night he go in the living room to see the uncle			340	3	10	0	43			0	43	speaker 16 8	2013/Jan/22	40%
27	you know what I mean? this one hurt he go to doctor and doctor just after full			62	1	8%	0	32			0	32	speaker 19 8	2013/Jan/22	32%
28	yeah yeah yeah yeah he change that he go for a week but they don't do that			85	1	37	0	45			0	45	speaker 19 8	2013/Jan/22	44%
29	but like something inside here he go to doctor and doctor just after full			40	0	70	0	21			0	21	speaker 19 8	2013/Jan/22	21%
30	secret of the uncle and ah one night he go in the living room to see the uncle			340	3	10	0	43			0	43	speaker 16 8	2013/Jan/22	40%
31	but like something inside here he go to doctor and doctor just after full			40	0	70	0	21			0	21	speaker 19 8	2013/Jan/22	21%
32	yeah yeah yeah yeah he change that he go for a week but they don't do that			85	1	37	0	45			0	45	speaker 19 8	2013/Jan/22	44%
33	you know what I mean? this one hurt he go to doctor and doctor just after full			62	1	8%	0	32			0	32	speaker 19 8	2013/Jan/22	32%
34	map not the hospital can I say I here go straight ahead you will see a			42	0	68	0	68			0	68	speaker 17 1	2013/Jan/22	60%
35	map not the hospital can I say I here go straight ahead you will see a			42	0	68	0	68			0	68	speaker 17 1	2013/Jan/22	60%
36	or I couldn't when I see where I go I remember monuments orange go			211	5	52	0	89			0	89	speaker 15 0	2013/Jan/22	88%
37	last week no I go on am holiday yeah eh ah the last			5	0	5%	0	2%			0	2%	speaker 7 15	2013/Jan/22	1%
38	week was wonderful I ah I go back ah I go to Spain ehmm but it's hard to			16	0	28	0	4%			0	4%	speaker 18 1	2013/Jan/22	3%
39	eh normal not much week yes and I go to the eh gym and am yes gym			23	1	32	0	34			0	34	speaker 5 12	2013/Jan/16	36%

concordance collocates plot patterns clusters timeline filenames source text notes

As Fig. 5.4 illustrates, even in this short span of a concordance screenshot, C1 learners meet the expectation that they can use a wide range of MWV in relation to this one pattern *go* + particle. However, we also see a range of patterns with *go* + particle at A2 as Extract 5.3 illustrates. In this extract, Student 18 (A2) is taking part in the oral examination task. In the task the learners are required to give the examiner directions to a pre-designated location on a map of the local city.

Extract 5.3 (Student <$18> A2 spoken corpus)

…am yes am eh take the road am **go straight** on am the first turn the first on the left and turn the first on the right in Sexton Street am and turn again **go to the first on** the left **go on up to** the eh first street on the turn right

In this roleplay, Student <$18> on numerous occasions used the lexical verb *go* followed by the adverb *straight* in a literal manner to indicate the perpendicular direction of destination. It must be noted that the words *perpendicular* or *vertical* would not fit the instruction of to move forward as they sound somewhat artificial. Furthermore, Student <$18> opted for the more figurative, phrasal prepositional verb: *go on up to* instead of choosing the single verb *continue*. Again, these structures used to give direction, functioned in making the exchange more relaxed and less face threatening. Certainly, the single lexical verb *continue* could have been substituted instead of *go on up* to but this would somehow detract from the casual nature of the interaction being simulated (within a high-stakes situation). Therefore, we see that even at A2, in a formal examination, the learner is able to opt for the more informal verb + particle pattern that more typically reflects language in use in their environment as an ESL learner. The EGP does not set this level of expectation in the use of MWVs at A2 level. For example, the only phrasal prepositional verb that A2 learner can typically use according to the EGP is *look forward to* (as discussed above).

In a further example of the unexpected use of phrasal prepositional verbs by A2 learners in the ACE, along with a range of other MWVs, we see in a recording from an A2 class (10/2/11). Student

<$5> is telling the teacher and class what she did at the weekend (Extract 5.4). [The group had previously learned the idiomatic phrasal prepositional verb *get up to* as a casual way of describing activities they did in the past]. As an ice-breaker, the teacher asked the group: 'what did you get up to at the weekend?' Student <$5> responds:

Extract 5.4 (Student <$5> A2 spoken corpus)

…I did not **get up to** much. Me and friends we am yeah am we are **going in** you know beside the am park in the park but ah where there is big statue… yeah there and have lunch there and **go back home** and am see something and some sleep and at night **I wake up** it was very quiet so I we only **go** one pub **to** Flannerys.

A2 learner <$5> introduced the topic by repeating the recently acquired phrasal prepositional verb *get up to,* based on the teacher's prompt. This expression is an example of a multi-word verb which functions to make language appear more casual and ergo native speaker like. Moreover, Student <$5> chose to use the prepositional verb *go in* in the present progressive instead of opting for the single lexical verb *enter*. *Enter* would again change the tone of the description and make it more formal. Furthermore, on finishing the description of the activities, the student states that he [*went*] *back home,* ergo combining the preposition *back* with the verb *go* (albeit it with no past tense marking). If the speaker had used the one word equivalent *return,* the language expressed would not appear as relaxed and would have failed to reflect the casual nature of the interaction and the informality of the register. In addition, Student <$5> combined the irregular verb *wake* with the preposition *up* to form the semi-transparent phrasal verb *wake up* instead of the more formal *awoke*. We note that the verb *awake* occurs 325 times (PMW) in the BNC, while there are 716 occurrences of the multi-word verb alternative *wake up* (PMW). Therefore the A2 learners use of this phrasal verb *wake up* is in line with the native speaker preference.

During an introductory speaking activity, the learners of the C1 cohort had to randomly describe pictures which the teacher gave them (see Appendix 7). These pictures were taken from *Reward Intermediate Resource Pack* (Kay, 2000) and each individual student received a different picture. They were set three minutes to prepare what they would say but they were not allowed to write anything and would therefore be encouraged to speak freely and at length. The activity finished with the teacher asking each student a number of questions. Student <$1> gave the following explanation of a story based on his picture.

Extract 5.5 (Student <$1> C1 spoken corpus)

…So basically I have am the same picture as the lady in the black. Her the black hat. So it starts yesterday and the lady got brush her teeth wash herself and then she **got dressed.** She was wearing ah black t-shirt tight slim jeans and ordinary boots. And the husband **came along** and he brought her a letter the was written that she her mortgage was accepted so she was kind of very happy. Then she **went to** the bank. And am they signed the deed for the mortgage so and now like in the same minute she getting <$=>she getting</$=> her house.

From extract 5.5, we can see that Student <$1> tended to overuse the verb *get* to some extent; such an overuse led to a number of grammatical errors. However, he used three different examples of multi-word verbs. He used the delexical and semi-transparent expression *get dressed.* This irregular lexical verb followed by the adjective *dressed,* functions in making his speaking in English sound more natural, as opposed to if he were to opt for the single lexical verb to *don*. As a matter of fact, *to don ones clothes* does not phonetically or stylistically fit well here, since it would sound completely unnatural. Moreover, the phrasal verb *get dressed* sounds more native speaker like and in effect fluent. In addition, the same could be said for the following two multi-word verbs: Student <$1> used the irregular lexical verb of *go* in the past tense combined with the preposition *to,* to mean an action of moving from one place to another. Once again this multi-word verb sounds more

native like than the formal *proceed to.* The same could be said of his description of the husband who *came along* instead of *arrived at.* Here the use of the irregular verb *come* in the past tense combined with the preposition *along* makes the action of the husband more inclusive in the sense that he accompanied his wife. On the other hand, if Student <$1> had stated *arrive at* instead, the entire meaning of the sentence would change as the husband would have therefore arrived alone and would not have accompanied the wife. Again, these multi-word verbs function and aid in making the student's speech sound more informal and native like.

5.4.3 Look and its collocations

The data clearly highlights *look* as one of the most frequent lexical verbs occurring in the corpus (see chapter 4 Methodology). Overall, *look at* represents 17% of the multi-word verbs in the corpus and *look for* equates to 21% of the entire multi-word verb occurrences in the ACE corpus. The lexical meaning of *look at* is evident in many expressions such as the prepositional verbs look *at the weather* and *I didn't want to look at the map* and *also look at the picture* where the meaning of *look* in all three is literal. Again here we see that A2 learners can show a far from limited use of MWVs contrary to what is expected at this level based on the EGP. For instance, Student <$15> from the A2 group shows the use of two patterns in one short fragment (Extract 5.6) and once of these is *look for* in the sense of *seek* (the other is *live in*):

Extract 5.6 (Student <$15> A2 spoken corpus)

…*I live in* a city in a shared house am eh I **look for** work.

While one might expect to find *look for* used from A1 level (based on the EVP) in a literal sense of searching for an object, its figurative use here (*look for work*) is less expected of A1 and A2 learners who are profiled to have a limited use of MWVs.

Extract 5.7 (C1 spoken sub-corpus) offers an interesting example of multiple uses of MWVs with *look* + particle. Student <$7> is describing a picture that she has been given by the teacher. Notice the density of MWV use in this fragment from one utterance:

Extract 5.7 (Student <$>7 C1 spoken corpus)

…And they decide not to shop. But they don't mind they **talking about** the things that they shout. He <$=>shout shout shout</$=> and finally eh my uncle **look at** him to beg him to am **shut up** but he look at the mother and told her "mam mam **look** there is" ah I have to **look at** the name. "There is ah a flying saucer in the other side walk". And they don't they **laugh at** him they don't **look at** the other side because some missing ya. But when they **looked at** the other side there were am was a very big and round and bright flying saucer. Ya and they emm they were amazing. Everybody was **looking at** the **at** the same place. Ya. And they were am **waiting for** something strange happen Ya.

Here, the learner uses the literal prepositional verb *to talk about*. This structure of verb followed by preposition is of neutral register. It is used across all genres, while at the same time maintaining the similar intended denotation. Student <$7> also combined the irregular verb *shut* with the preposition of place *up* to form the opaque *shut up*. This colloquially informal and often offensive construction is used to silence a speaker. In sharp contrast, the formal counterpart: *silence,* does not possess the same abruptness and power as the multi-word verb. Student <$7>, in this short piece of speech, used the literal prepositional verb of *look at* on five occasions. The structure of the regular verb *look* followed by the preposition *at* is a much more naturally occurring construction than its counterpart *view*. Student <$7> finally finishes her picture description with the combination of the regular verb *wait* followed by the preposition *for*. Again this multi-word verb construction sounds more native like and neutral than its more formal counterpart *await*.

5.4.4 Put and Care

As previously stated, the EGP profiles A2 learners are only being able to use a limited range of MWVs but there are many instances where the ESL learners contradict this. Two items which illustrate this qualitatively in extract 5.8 are *put someone up* and *care about.* Both of these forms are used in one utterance by an A2 learner and yet they are profiled as forms that we would expect *B2*

and *B1* (respectively) to use according to the EVP. While the learner's use has morphological errors, it is clear that the learner understands these forms and can confidently use them to communicate. Looking at this student in context of her learner profile, she tends to attempt to speak and interact as much as possible. That is, she is very communicative but her grammatical and lexical accuracy is often quite poor. Her written competency is much weaker than her ability to communicate in speaking. This is not untypical of an A2 ESL learner. Because they interact often in spoken situations in the workplace or socially through English, they can often gain confidence in oral language use that is beyond their written competency (and often, they may not have even studied English formally). The important point to note is that this A2 learner who struggles in written communication is using two MWVs that are typical of B2 (*put someone up*) and B1 (*care about*) level learners. In this extract, she is describing a charity that she used to work for when living in her home country of Poland.

Extract 5.8 (Student <$16> A2 level spoken corpus)

…so they ah took a woman and then they have a like **put her up** and she ah woman what do you say someone who **care about** everything like am **care.**

The speaker chose to use the phrasal verb *put up* rather than *accommodate.* This idiomatic multi-word structure of the irregular verb followed by a preposition aids her speaking in sounding more native like than the one word alternative; *accommodate*. Student <$16> accurately used this separable transitive prepositional verb along with *care about*, where again the structure of the verb *care* followed by the adverb *about* aids the speech in sounding more fluent.

Student <$24> from the C1 group uses the following literal multi-word verbs during his oral examination (4/5/11). Extract 5.9 includes a literal use of *put* + particle but more interestingly, the extract displays the ease with which the speaker draws on a number of MWVs in this high-stakes examination situation. These examples are taken from his response to the question *where do you work?* He stated:

Extract 5.9 (Student <$24> C1 spoken corpus)

…I'm a <$=>I'm a</$=> forklift driver I ***work in*** a warehouse warehouse what we do is just receive goods from am the United States and then ah we ***put it in*** the warehouse ah in the warehouse we have eh am ah four drivers that would be receive a order we generate company that ***deals with*** technology parts and the parts ***come from*** United States so am the order with the computer and then we ***sell*** it ***to*** those who yeah

In this short exchange Student <$24> gave the following literal prepositional verb: to *work in*. This construction of a regular verb followed by the preposition *in* resonates a relaxed and familiar tone between the speakers. Such a relaxed, informal and neutral in formality tone would be lost if the speaker used the verb *employ* as in *I am employed in* or *at*. Student <$24> also used the literal multi-word verbs *put in* and *sell to* and the literal prepositional verb *deal with* in describing the duties of his job.

Having looked at a number of qualitative examples, we can say that there is ample evidence of use of MWVs at both levels. Most surprisingly, there is use of MWVs at A2 that are not expected at this level when we compare with the EGP and the EVP. However, we note that these are often error prone in terms of tense and agreement for example. Nonetheless, in all cases, these errors did not impede comprehension. It seems that high frequency MWVs that are heard by both A2 and C1 learners are frequently used in their speech. Let us now look at the written data.

Table 5.9 below outlines the top twenty most frequent multi-word verbs in the ACE written corpus. Column one shows the lexical verb, column two the particle, column three the occurrence normalised PMWs and finally column four; the cohort in which it occurs.

Table 5.9 The twenty most frequent multi-word verbs in the ACE written corpus

Lexical Verb	**Particle**	**Occurrence (PMW)**	**A2**	**C1**	**EVP***
Talk	about	411		√	C2
Write	about	411	√	√	-
Look	for	347	√	√	A1
Live	in	276		√	B1
Go	out	276	√		A1
Work	in	241	√	√	A1
Decide	to	205		√	A2
Come	in	205		√	A1
Talk	to	250	√	√	A1
Find	out	135	√		A2
Go	on	135	√	√	A1
Listen	to	132	√		A1
communicate	with	132		√	B1
Get	on	130		√	B1
make	up	100	√	√	A2
Wait	to	100	√		A1
Need	to	35		√	A1
Save	for	35		√	A2
Search	for	35		√	B1
Show	up	22		√	B1

Percentage			50%	80%	

*The earliest level at which items appear on the EVP are noted. Where there is no result, this item is not listed in the EVP.

Table 5.9 shows an interesting range of MWVs in the top 20 and it also shows that they are used to differing degrees. Some are used by both cohorts though, not surprisingly, to a lesser extent by the A2 cohort: 50% of the top 20 items are used by A2 learners while C1 learners use 80% of them. Counter to expectations, there are some items that are only used by A2 learners, namely: *wait to, listen to, go out.* These are all items that are expected at A2 level (based on the EVP).

As can be seen from Table 5.9 above *talk about* and *write about* are the most frequent multi-word verbs used in the ACE written corpus with 411 results (PMW) each. Extracts 5.10 and 5.11

Student <$11> from the C1 cohort was tasked with an essay on the differences between their home country and Ireland wrote:

Extract 5.10 (Student <$11> C1 spoken corpus)

The quality of life here is very good but on the other hand we have a few cultural and deep facts to ***talk about***. I have to highlight the fantastic 9 to 5 working hours as a big shock for me.

The A2 group's first written homework was to write 50 words about themselves. Student <$10> wrote the following:

Extract 5.11 (Student <$10> A2 written corpus)

…I ***talk about*** myself is the simple and easy, so my myself. My name is …., I'm twenty-seven years old.

Student <$1> from the C1 cohort wrote the following as an introduction when asked to write about his favourite film. In this introduction Student <$1> also uses the phrasal verb *based on*:

Extract 5.12 (Student <$1> C1written corpus)

…There are two films about Iron Man, but my favourite one is second film, so I will ***write about*** it. It is a 2008 American superhero film ***based on*** the Marvel Comics.

The prepositional verb *live in* also occurs frequently in the written corpus with 276 normalised results. Student <$8> wrote the following:

Extract 5.13 (Student <$6> A2 written corpus)

…But my favourite hobby is spend time with my family. I love my family. When I was in Madrid I used to **live in** my father's house.

Extract 5.14 illustrates a C1 level student <$20> using *live in*:

Extract 5.14 (Student <$20> C1written corpus)

Beijing is the capital of China it's a big city and a lot of people ***live in*** Beijing.

In summary, looking at the range of uses of MWVs in the ACE relative to the EVP (Table 5.9), we can say that while there is ample spread of forms. Examining them qualitatively, we can see learners using them in their writing. However, overall, we can conclude that the types of MWVs used in the learners' writing are mostly forms that they have acquired at A1 and A2 levels as detailed in Table 5.10 (note one of the top 20 items is not listed in the EVP):

Table 5.10: Summary of the EVP profiles of the MWVs used by learners in their writing

EVP level	**No. of MWVs**	**%**
A1	9	47.37
A2	4	21.05
B1	5	26.32
C2	1	5.26
Total	**19**	**100**

As we can see from Table 5.10, close to 50% of the MWVs that are used in the written corpus are items that learners in general have acquired by A1 level. This adds weight to the point above in relation to ESL learners' ability to use more MWVs, across a wider range, in their speech and that

this is not always reflected in their writing. As we speculated, this may be because ESL learners are exposed to many high frequency MWVs in everyday casual conversation but writing genres can pose more complex contexts of use and more formal registers where learners are not taking risks with the use of less frequent MWVs generally. We will now explore the MWVs and register.

5.5 Register and one word substitution

As mentioned above, MWVs in speech, suggest an informal and familiar tone between the interlocutors. In order to express a similar meaning in a more formal situation, language users frequently choose to use a pre-existing one word synonym (and this partly explains the lower frequency of MWV use as discussed above). Most MWVs can be substituted by a single-word equivalent in more formal registers, especially in writing. Table 5.11 below focuses on fifteen MWVs. These are items not found in the top 20 most frequent items listed in Tables 5.7 and 5.9. They were chosen manually form the ACE data to illustrate the effect on register when one substitutes them with a one-word equivalent. Table 5.11 serves to illustrate that if we were to replace the MWVs below with their synonyms, the tone and interpretation of the utterance / sentence would change, so much so that in certain cases the language could appear to sound incorrect. For example, compare: *I come from Ireland* to *I originate from Ireland.* While both of these sentences are grammatically correct, the second one seems inappropriate in its register in most genres.

Table 5.11 Multi-word verbs uttered in ACE corpus and their possible one word equivalent

Multi-word Verb	**A2**	**C1**	**Formal One-Word Synonym**
put up with		√	*Tolerate*
come up with		√	*create or invent*
get on with		√	*continue or relate*

pick up	√		*collect or acquire*
count on	√		*Depend*
step on it		√	*Hurry*
bring up	√	√	*raise, mention*
go away	√	√	*Leave*
shut up	√	√	*Silence*
go forward		√	*Continue*
come from	√	√	*Originate*
look for	√	√	*Search*
come back	√	√	*Return*
fill in		√	*Complete*
wake up	√	√	*Awaken*

We note that not all multi-word verbs are synonymous with one unit. For example, verbs such as *wear out* and *knock out* cannot be replaced with any single unit. Their intended meaning cannot be authentically conveyed without the use of such multi-word constructions. In addition, the phrasal verb *put on* is, according to MacMillan (2005), 20 times more frequent than the more literary, single verb *don*. *Protrud*e is, also according to the corpus, only about half as frequent as *stick out,* and tends to occur predominantly in medical scenarios. The work of Fletcher (2005) and MacMillan (2005) shows that MWVs are found in various text genres and registers and they note that they are essential for expressing different opinions in English.

In summary, while one could argue that the use of high-frequency MWVs in the ACE data could perhaps point to a lack of more formal vocabulary (especially in writing), it is argued that there is not always a one word equivalent (Fletcher, 2005) but more importantly, in relation to spoken language, the use of MWVs is an important marker of informality that is crucial at a relational level.

Finally, we also note that, in accordance with the work of Buttery and Caines (2012), lack of use of language exponents can also be explained by lack of 'opportunity of use'. As already mentioned there is no explicit learner outcomes (Appendices 1 and 2) for any of the lexical strings, the teacher did conduct a class on a sample of phrasal verbs taken from *New Headway Advanced* and giving and asking for directions, both lessons resulted in the occurrence of MWVs such as *go to, go on up to, take after, take over, look forward to* and *come up* with for example. Each learner therefore had the same opportunity to produce multi-word verbs in the classroom.

5.6 Conclusion

Let us return to the main focus of this chapter: whether and to what degree learners at A2 and C1 in the sample cohorts use MWVs. We can clearly state that from our analysis here that both A2 and C1 learners appear to productively use a range of phrasal verbs, prepositional verbs and phrasal prepositional verbs in both their spoken and written English. We can summarise overall that, especially in relation to A2 learners, this finding is unexpected relative to profile frameworks EVP and EGP as well as CEFR can do statements, IELTS and FCE examination syllabi. It is argued that their use mirrors what they frequently hear in their ESL context. So while A2 learners might not be expected to know certain MWVs in general, when they live and work in an English speaking environment, they frequently encounter these items and, as discussed in Chapter 1, there is most likely a subconscious acquisition process in train and this aligns with the Usage Based acquisition model (Ellis and Larsen-Freeman, 2006) as discussed in chapter 1. It points to the lack of accuracy of generic profiles based on EFL general English learners when dealing with ESL contexts. On many occasions in spoken contexts in this chapter, we saw instances of learners using MWVs in their speech. Often these instances had errors but these did not impede conversations. It showed that even at A2, learners were adequately familiar with and confident to try out MWVs that are deemed 'beyond' their level of competency.

In the context of the written data, we noted that while there was frequent use of MWVs, the nature of the forms in terms of competency levels (i.e. within the EVP and EGP) was lower than in the spoken classroom data. We acknowledge that the ability to accurately measure learner performance outside of the language classroom remains a limitation to the study but that the written samples are almost all from writing homework (i.e. they are evidence of written student performance outside of the classroom also). Furthermore, the debate remains as to whether productive use indicates understanding. Without clinical experimentation, we cannot prove this, of course.

With reference to the sub question 'Which cohort used more multi-word verbs?', it was established that not all A2 learners used MWVs (55% in spoken data and 67% in written data at A2) while all learners in the C1 used MWVs to some degree. We noted that for PMWs there are more MWVs used in the written data than in the spoken data overall (see Table 5.3). However, while the PMW use is relatively comparable in speaking (324 PMW in A2 and 522 in C1 data) it is starkly different in the written data, with MWVs 6,428 PMW in the A2 written data compared with 19,615 PMWs in the C1 writing samples. However, it is interesting to see that across both cohorts the competency range of the MWVs used was lower than in the spoken sample as a while with learners using many forms that are notionally acquired by A1 level according to the EVP.

Interestingly, as table 5.2 indicates, 5 A2 and 15 C1 learners used MWVs. When the total uses of the forms is normalised per speaker, it shows that per student, the A2 learners used more MWVs (9940 per 100 words used at A2 compared with 4327 at C1). While these figures do not account for the nature of the use, whether correct or incorrect, whether repeated (see below), and so on, they do point to a confidence in many A2 learners in the use of what are seen as 'difficult' vocabulary items and it underscores the differing learning context of ESL students. Items that are high frequency in their linguistic environment seem to be acquired and used more than one would expect at this level (when compared with competency frameworks based on EFL learners). This also points again to the inadequacy of such frameworks for ESL contexts.

In terms of the corpus compilation process and as discussed earlier, opportunity of use did in some way affect the students' performance. As evident from the corpus, both the A2 and C1 cohorts prepared for, and completed an oral examination in the target language. One component of the examination was to give and ask for directions on a map of the local city. The recordings from this task gave rise to numerous collocations with the verbs *take* and *go*. In particular, expressions such as *take a right*, *take off*, *go to*, *go up to, go straight on* occur frequently in the corpus; (with respect to *go to* occurring 76 times in the corpus and thus averaging at 25%). Furthermore, it is imperative to highlight that the students opted to use such structures in opposition to any given possible one word equivalent. This resulted in their speaking sounding and appearing more fluent and establishing social solidarity with the other interlocutor.

As illustrated earlier, multi-word verbs function in conversations in a similar way to idiomatic or figurative expressions. They are used to imitate native speaker discourse and such imitation functions to reflect the cultural solidarity and inclusion expressed by a native speaker. Therefore, the multi-word verb acts as an inclusive device that aids such unity, so much so that the cultural solidarity of a speech community is normally reflected in the language they use on a daily basis. In fact, Sapir (1956: 104) contends that 'every cultural pattern and every single act of social behaviour involves communication in either an explicit or implicit sense. The tool for this communication is language'. Clearly, and as shown here, language and the meaning thereof, is embedded in the culture in which it is expressed. Register and the language used to express such register is an essential component of any language function. Structures such as the MWVs, in this chapter, reflect a more relaxed and informal register as they tend not to occur so frequently in academia and other formal situations.

To summarise, it is clear that the students at both A2 and C1 levels in the ACE data use MWVs in their interaction inside the classroom and also in their sample writings. Overall, such occurrences function in making their spoken and written language clear, relaxed and informal and it is posited that this is an important discourse function for ESL learners in their choice to use these items.

In the next chapter, we will focus on a further lexical collocation: delexical verbs. Delexical verbs are another type of multi word unit and the chapter will both identify and analyse these across A2 and C1 cohorts in the ACE data across speaking and writing.

Chapter 6 Delexical Verbs

She took him to the cleaners. (Susie Bamworth, 2004).
It is the interaction between words and sentence structure which actually convey sense. (David Crystal, 2005: 190).

6.0 Introduction

This chapter focuses, in detail, on delexical verbs. Delexical or 'weakened' verbs have become the focus of some linguistic investigation (see Leech, Rayson and Wilson, 2001 and McCarthy, 2014) and are shown to be a prominent feature in English native speaker discourse. In this chapter, we look at delexical verb use is ESL learners using the ACE corpus. Firstly, the chapter will examine existing definitions. Secondly, it will identify and discuss the verbs most frequently used delexically in the datasets and compare their frequency of use with the BNC as a baseline. Thirdly, the chapter will explore some syntactic and semantic features of these verbs before focusing in depth on the forms and functions of the delexical verbs drawn from the data of the ACE corpus. In addition to the BNC, the EVP will also be used for comparison with the ACE findings. The BNC will give us accurate frequency of occurrence in a large representative native speaker corpus while the EVP will show us the level at which these delexical verbs are typically known by English language students. The overall aim of the chapter is to ascertain the degree to which the learners at both language levels, A2 and C1, of ACE use delexical verbs in their spoken and written classroom interactions.

6.1 Defining Delexical Verbs

Carter and McCarthy (2006: 784) note that delexical expressions using high frequency verbs such as *do, get, give, make, take* enable a verb-type meaning to be expressed in the noun object that follows the verb. They use the example of *give* in the following patterns to show how the meaning of the verb is drawn from its noun collocate. Carter and McCarthy (2006) also note that

syntactically, the preference is for Indirect Object (IO) + Direct Object (DO) rather than DO + Prepositional Complement (PC):

Gave + song

I gave them a song

(*IO +DO* rather than DO +PC – I gave a song to them)

As McCarthy, O'Keeffe and Walsh (2010: 32) put it, the meaning of delexical verbs seems 'to change depending on what they collocate with'. Essentially, the verbs are termed 'delexical' because their lexical meanings are not very definite and they pick up their meaning from their collocates. O'Keeffe, McCarthy and Carter (2007: 38) describe delexical verbs as having 'low lexical content and their meanings *rest* on their collocates'. They point out that one problem associated with the high frequency of the delexical verbs is the fact that their low lexical content has to be 'complemented by the lexical content of the words they combine with, and these collocating words may often be a relatively low frequency beyond the core [vocabulary]'. (O'Keeffe et al, 2007: 38).

Syntactically, delexical verbs occur in five different forms (Carter and McCarthy 2006):

Table 6.1: Five patterns with delexical verbs (Carter and McCarthy 2006)

Type	**Example**	**Feature**
Verb + (determiner) noun	*have an argument* *have lunch*	Rather than carrying the lexical meaning of the verb (in this case *have* = *possess*), the meaning is drawn from the noun collocate.
Verb + adjective	*go crazy* *go red*	Typically idiomatic, the verb draws its meaning from the adjective.

Verb + preposition	*go on* *get by*	Idiomatic and quite similar to the phrasal verb, the verb's main meaning lies with the preposition.
Verb + gerund (*ing*):	*go shopping* *go camping*	Formulaic in nature, the meaning of the verb in these patterns is drawn from the *ing* for that follows.
Verb + adverb	*Sleep well* *Take care*	Often idiomatic and transparent semantically, the verb in these patterns relies on the adverb for its meaning.

Definitions here will guide the analysis that follows. We note that there is some degree of overlap here in terms of other analysis chapters (e.g. phrasal verbs in MWVs, chapter 5 and idioms in chapter 8). However, the focus here is on the delexical nature of the patterns. Not all verbs in MWV patterns nor idioms are delexical hence a separate treatment here (albeit with some overlap).

6.2 General features of delexical verbs

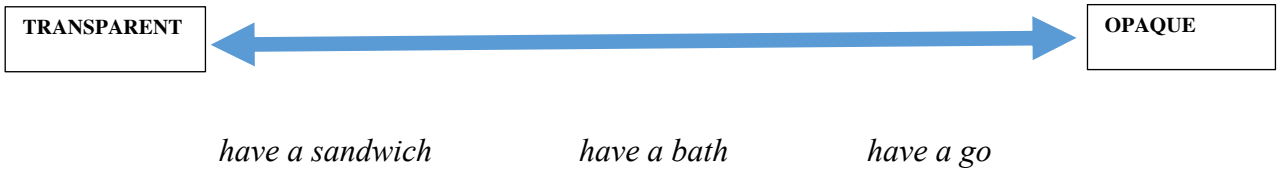
6.2.1 Transparency/opacity versus Idiomaticity.

Degree of semantic transparency and idiomaticity are important features of delexical verbs. Sinclair's (1991) ***Idiom and Open Choice Principles*** (See Chapter 8 Idiomatic Expressions for a full discussion) views utterances as either a) a carefully selected choice in grammatical and syntactical meaning, where the intended meaning of the speaker is self-explanatory or b) a language unit or chunk, where the meaning is not derived from individual units (for example *of course,* although two separate words, it has a single meaning and operates as one idiomatic unit). Sinclair's (1991) principles also apply to delexicality as these verbs similarly exist on a cline in meaning from transparent to idiomatic. Delexical verbs such as *have an opinion* or *take a break* are reasonably

transparent in meaning. In contrast, other delexical patterns can be far less transparent in meaning such as: *he got off; take after somebody; to go ape.*

As Fig. 6.1 (based on Biber et al 1999) illustrates, we can say that there is a transparency cline in terms of the semantics of delexical patterns:

Fig 6.1 The Cline of Idiomaticity of Delexical Structures (Biber et al, 1999)



Our expectation is that opaque delexical patterns will be more challenging for learners and we would expect lower level learners not to have acquired many of any of these patterns. We shall explore this further below.

6.2.2 Register, single word equivalence and substitution

In many cases, a delexical pattern can be replaced by a single lexical verb which has a similar meaning (Sinclair 1990: 147; Carter and McCarthy 2006; McCarthy, 2014). An important related point relates to register: Carter and McCarthy (2006) note that the use of delexical verb patterns can offer greater informality than the single word equivalent. Compare: *I gave them a song* versus *I sang.* This brings about the expectation that a learner will need to be aware of the register effect of their choice of language. For example, to choose to say or write *take a holiday* over *to holiday* amounts to a register choice

Informal: *I took a holiday in France last year.*

Formal: *I holidayed in France last year.*

For a learner, the single word option can sometimes be the easier item to understand and use (and perhaps equates in some way to the form in their first language). However, while a single word

equivalent such as exemplified here is not incorrect, it sounds somewhat more formal in register and arguably far less natural in most contexts. Further exemplifying Carter and McCarthy's point about register, we can compare other frequent delexical items and their single-word equivalents in Table 6.2:

Table 6.2 Examples of delexical verbs and their single-word equivalent

Delexical Verb	**Single-verb equivalent**
give an interview	*Interview*
give a kiss	*Kiss*
give a laugh	*Laugh*
go on holidays	*Holiday*
have an argument	*Argue*
have a date	*Date*
have a look	*Look*
have respect for	*Respect*
have a shave	*Shave*
have a shower	*Shower*
make a decision	*Decide*
make a promise	*Promise*
take a glance	*Glance*

In summary, a speaker's choice to opt for the delexical pattern tends to make the language sound less formal and this may be an important factor in sounding more native-like and fluent, especially in informal situations. This is a point that we will return to when looking at the ACE data.

6.3 Corpus analysis of delexical verbs and the BNC as a reference corpus

6.3.1 Method of analysis

For a teacher, the British Council (2018) offers a list of the following nine high frequency verbs as being the most frequent to occur delexically:

do make take have give go get come bring

For a corpus linguist, these items can be viewed through a representative corpus such as the BNC to appraise their actual frequency PMWs. A quick survey of the BNC clearly confirms that these verbs are high frequency (Table 6.3) (Leech, Rayson and Wilson 2001). We can also collate the PMW frequencies of these verbs in the ACE (Table 6.3):

Table 6.3 High frequency verbs in the BNC and ACE corpora

Verb	**BNC (raw)**	**(PMW)**	**ACE (raw)**	**(PMW)**
do	161,730	1617	494	2905
have	74,767	747	320	1882
get	42,256	422	64	376
go	30,350	303	310	1823
come	21,027	210	86	505
take	15,227	152	28	164
give	11,766	117	44	258
make	11,421	114	20	117
bring	2893	289	8	47

We can see from Table 6.3 that there are some similarities and differences in the PMW frequencies across the BNC and the ACE data. However, for the purposes of the present study, we note that these results are undifferentiated so many of the verbs here may not have been used delexically (e.g.

do and *have* as auxiliaries or *get* used literally to mean fetch, and so on). This points to a challenge in finding delexical verbs in the data. In other words, it is not an automatic process. The approach used in the present study was to:

1) manually identify all uses and attempted uses of delexical verb patterns in the ACE, and

2) compare the normalised frequencies of ACE patterns with the BNC, across level and spoken versus written modes;

3) each item was also compared with the EVP level (i.e. the level of proficiency at which this pattern could be expected in learner language in general).

Apart from summarizing these comparisons in terms of frequencies and EVP levels across the A2 and C1 spoken and written data, this chapter will also looks qualitatively at the learners' use and attempted use of these patterns. This gives us the opportunity to look at how learners experiment with delexical patterns. Even though they may sometimes produce erroneous patterns, it shows an extension of their use of delexical patterns as a concept. For example, as we discuss below, learners can show multiple delexical patterns with a particular verb *take a left, take the bus* and can extend this use creatively to say *take a lotto* (see below for a further discussion on this).

Therefore in the analysis, all delexical patterns were identified, whether correct or incorrect. Incorrect items that had correct patterning were counted within the totals (e.g. *Last week, I go on holiday* was included in counts because it had correct VERB + COLLOCATE patterning though it had a tense error). However, incorrectly patterned uses were only looked at qualitatively – in other words, a use like *take the lotto,* as discussed above, could not be compared quantitatively with the BNC and did not have an EVP entry yet it offered an interesting view into attempted use.

6.3.2 Overall results

Table 6.4 summarises the overall use of delexical verbs across the 24 learners in this study (9 A2 learners and 15 C1 learners):

Table 6.4 Number and percentage of students using delexical verbs across both cohorts

	Speaking		Writing	
	Raw	**%**	**Raw**	**%**
A2 (9 participants)	4	44	9	100
C1 (15 participants)	15	100	15	100
Total	19		24	

As we see from Table 6.4, all students use delexical verbs in writing. In speaking, 44% of A2 learners used them while they were identified in all C1 level learners' data.

Table 6.5 summarises the uses of delexical verbs across spoken versus written data by level:

Table 6.5 Frequencies (raw and normalised) of delexical verbs in A2 and C2 data, across spoken and written data.

Level	**Spoken (raw)**	**PMW**	**Written (raw)**	**PMW**
A2	86	1344	50	7143
C1	108	1256	90	6923
Total	**194**		**140**	

Table 6.6 summarises, there is a total of 334 (raw) / (1965 PMWs) delexical verbs across the whole of ACE.

Table 6.6 Delexical verbs total frequencies (raw and PMWs) across speaking and writing in ACE

	Total (Raw)	**PMW**
Total speaking	194	1,293
Total writing	140	7,000
Overall	334	

Taking the Tables 6.5 and 6.6 together, we can say that delexical verbs are substantially more frequent in the written components of ACE. Looking at raw results, it seems that A2 learners use of delexical verbs in speaking accounts for 38% of all uses, with C1 learners using 62%. In writing, A2 learners appear to use just 15% of all delexical verb patterns while C1 learners use 85%. However, when the figures are normalised to PMWs based on the spoken and written data sizes, we see that the A2 use is slightly higher than that of C1 learners, with a ratio of A2 : C1 = 1,344 : 1,256. In writing, we see A2 : C1 ratio of delexical pattern use is 7,143 : 6,576. Therefore, in writing, A2 learners seem to use these patterns more (pro-rata).

As expected, we find delexical patterns at C1 level (Carter, 1997; McCarthy and O'Dell, 2004). What is striking about these results it that contrary to descriptors in the CEFR and IELTS, among others (see Appendix 5), this result clearly shows that the A2 learner cohort is using delexical verbs in both their speech and writing. the higher level cohort also use this type of verb in their classroom communication.

Let us now consider the actual forms and patterns of delexical verbs that were found in the ACE corpus through manual sifting and compare these in greater detail with the EVP and the BNC. First we look at the spoken component of the ACE.

6.3.3 Delexical verbs in the spoken component of ACE

Table 6.7 below details the most frequent delexical verb patterns (with a cut off of > 4 occurrences) in the spoken component of ACE, by level. It also identifies the expected level of occurrence of each item the EVP:

Table 6.7 Most frequent (> 4 occurrences) of delexical verbs patterns in the Spoken ACE, by level compared with the EVP expected level.

Delexical	Raw Results	(PMW) Results	A2	C1	EVP Level
have a look	42	247	√	√	B1
ask a question[9]	28	164	√	√	A1
have a holiday	19	111	√		A1
speak English	13	76	√	√	A1
save money	12	70		√	A2
have work	8	47	√	√	B1
have a pint	8	47	√	√	B2
have coffee	8	47	√		A1
improve my English	8	47		√	A2
make mistakes	6	35	√		A2
answer a question	6	35		√	A2
go to pub	6	35	√	√	A2

[9] *Ask a question* is possibly a debatable categorisation. Here it is included as a delexical pattern on the basis that while the single lexical verb *ask* (meaning to question somebody, i.e. interrogate/interview) is not synonymous with 'ask a question' as is the case with 'ask a favour', it does have a degree of delexicality and we argue that it is an example of delexicality being a matter of degree rather than an absolute.

do something	5	29		√	A1
have a talk	4	23		√	A2
do homework	4	23	√	√	A1
have lunch	4	23	√		A2
do a course	4	23	√	√	B1
have a party	4	23	√	√	A2
have a problem	4	23	√		A2

Table 6.7 shows a wide and interesting range across the most frequent items, which we will discuss qualitatively below. In summary, by way of comparison with the EVP, we can say that C1 learners' use of delexical verbs in speaking shows forms that are expected to be acquired by A1, A2, B1 and B2 levels. In contrast, we see that A2 learners both forms that they are expected to typically have acquired by A2, plus five patterns that are not expected at their level.

Let us now look at some learner uses. As can be seen in Table 6.7 above, the most frequent delexical verb in the spoken corpus is **to have a look** with 247 (PMW). While discussing an experience with a plumber, Student <$7> (a C1 student) stated the following:

Extract 6.1 (Student <$7> C1 spoken corpus)

Student <$7> yeah but no but he come **to have look** and say I have to call ah another people.

It is clear that main source of meaning lies with the noun phrase *look* rather than the verb *have*. This example also indicates the structure of *verb followed by noun* as the most frequent delexical structure in the spoken corpus. We note that *to have a look* in this usage is opaque in meaning and rather than literal. Often, a single word substitution for this phrase would simply be *look*, but in this utterance, it would not convey the intended meaning being expressed. A further example was uttered by an A2 student speaker Student <$4>, in its literal sense, who stated:

Extract 6.2 (Student <$4> A2 spoken corpus)

…I **had a look** in the fridge

As can be seen from Table 6.7, *have a look* appears in both the A2 and C1 spoken datasets despite the fact that the EVP highlights *have a look* as an exponent used by speakers at B1 level and above. The second most frequent delexical verb in the ACE spoken corpus is the regular verb pattern is *ask a question* (see footnote 9 above). It appears 164 (PMW) and is used by both A2 and C1 students and is expected to have been acquired by A1. Extract 6.3 shows A2 level Student <$6> using it:

Extract 6.3 (Student <$6> A2 spoken corpus)

I **asked** my friends **a question**

This pattern is used in a literal meaning and it is not surprising to find it used in high frequency in a classroom situation and this validates the EVP's placement of it as a pattern that is expected at A1 level.

Another high frequency and literal pattern that is expected by A1 in the EVP is *have a holiday.* This occurs 111 times (PMW) in ACE and again this is expected within the context of classroom materials and talk about holidays. One example that illustrates this is typical use was uttered by Student <$15> (A2).

Extract 6.4 (Student <$6>A2 spoken corpus)

…it is about the family who wanted **to have a holiday.**

Tying in with our earlier discussion about register, this example could be replaced with the single verb *holiday,* but this would sound unusually formal and even unnatural compared to its delexical counterpart. Also worth mentioning is the prepositional verb *go on holiday* which appears in the spoken corpus 62 (PMW) times. *Go on holiday* was discussed earlier in Chapter 5, Multi-Word Verbs (as discussed there is some overlap in the categories).

Another point of note from Table 6.7 are that there are clusters of uses around certain verbs. For example, there are nine 9 (> 4) occurrences of patterns with *have:*

have a look; have a holiday; have work; have a pint; have coffee; have a talk; have lunch; have a party; have a problem

Linking to earlier reference to usage-based models of language learning (Ellis and Larsen-Freeman, 2006) (see chapter 1), we can posit that learners are intuiting semantic prototypes linked to patterns. In this case, we can see HAVE + food/drink: *have a pint; have coffee; have lunch* and HAVE + event *have work; have a party.*

We now briefly consider some single item occurrences that are not listed on Table 6.7 as they did not meet the > 4 occurrences threshold but they are interesting nonetheless.

Table 6.8 Examples of delexical patterns with single occurrence in Spoken ACE, by level and EVP

Delexical	**Raw Results**	**(PMW) Results**	**A2**	**C1**	**EVP Level**
have a vision	1	6	√		C1
make a decision	1	6		√	B1
do the shopping	1	6		√	-
give a kiss	1	6		√	-
get the opportunity	1	6	√		-
bring something	1	6		√	-

Notably, an instance of the pattern *have a vision,* which appears at C1 in the EVP is found in the A2 data. We note that the use is also marked correctly for tense:

Extract 6.5 (Student <$4> A2 spoken corpus)

…she had she had she **had some vision** of the am ghost

Another one we note is C1 Student $4's use of *do shopping* during a speaking activity. The task was to create a narrative based on a picture taken from *Reward* (Kay, 2009). It forms the pattern of verb + *ing* noun or gerund (see above). Here, the verb behaves in a similar manner to an auxiliary verb. There is no physical effect on the gerund of *shopping* and furthermore, it does not emphasise the typical meaning of the verb *do.* It is also interesting to note that while *do shopping* does not appear in the EVP at all, *go shopping* is placed at A2 level.

Extract 6.6 (Student <$7> C1 spoken corpus)

My picture is am ah story of my am uncle Adusa. Ya? And am he is with my cousin Paul. Ya. They want **to do shopping** like all the parents

As detailed in Table 6.8, the delexical verb followed by noun *make a decision* occurs once in the spoken corpus. For example, Student <$5> said:

Extract 6.7 (Student <$5> C1 spoken corpus).

Student 5: we can't delay much longer we have to **make a decision** soon.

While *make a decision* occurs from B1 upwards, according to the EVP, it is still noteworthy that the student opted for this pattern rather than the shorter, single-word substitute: the lexical verb *decide.* Finally, we note that while the verb *take,* which, as previously discussed is seen as very prone to delexical use, does not occur in Table 6.7. However, a pattern with *take* was found in the form of *take a lotto* was found in the spoken data when used by a C1 learner who was trying to arrange a class lottery collection. While this is a non-standard pattern, we can again point to the usage-based

theory where this may indicate an interesting extension of the prototypical meaning of *take* on the part of the student. Most notably, the class that had preceded this informal exchange had focused on delexical *take* in relation to the topic of looking for and giving directions (e.g. *take a left, take the bus, take a turn*).

Extract 6.12 (Student <$18> C1 spoken corpus):

no I don't but it's okay if you want **to take a lotto** I don't mind.

Thus far, by way of brief summary and nature on the structures of delexical verbs in the spoken data:

- we can clearly see that the verb + noun pattern is the most frequent delexical structure in the corpus. This is arguably the easiest to use delexical structure and this could explain why it was the most frequent example.
- we also note that, the students were using more delexical verbs with transparent or clear meaning rather than the idiomatic.

We can also say that when Tables 6.4 and 6.5 are aggregated that while only 44% of A2 students use delexical verb patterns compared with 100% of C1 students, the A2 students use them more per individual student. When we divide the total A2 uses by the number of A2 learners who used the patterns (4), we get 21.5 uses per student compared with a rate of 7.2 per student at C1 level (i.e. 108 patterns divided by 15 students at C1 who used them).

Let us now compare the delexical verb patterns in Spoken ACE listed in Table 6.7 above (i.e. the most frequent items with a cut off of > 4 occurrences) with their frequencies in the BNC. Table 6.9 adds the BNC raw and PMW frequencies (see shaded cells) to the results presented in Table 6.7.

Table 6.9 Comparison of the most frequent delexical patterns (> 4) in Spoken ACE with BNC

Delexical pattern	Freq. (raw)	Freq. (PMW)	A2 cohort	C1 cohort	EVP level	BNC (raw)	BNC (PMW)
have a look	42	247	√	√	B1	1,909	19.09
ask a question	28	164	√	√	A1	36	0.36
have a holiday	19	111	√		A1	35	0.35
speak English	13	76	√	√	A1	172	1.72
save money	12	70		√	A2	242	2.42
have work	8	47	√	√	B1	51	0.51
have a pint	8	47	√	√	B2	26	0.26
have coffee	8	47	√		A1	41	0.41
improve my English	8	47		√	A2	1	0.01
make mistakes	6	35	√		A2	138	1.38
answer a question	6	35		√	A2	36	0.36
go to pub	6	35	√	√	A2	30	0.3
do something	5	29		√	A1	2,268	22.68
have a talk	4	23		√	A2	57	0.57
do homework	4	23	√	√	A1	12	0.12
have lunch	4	23	√		A2	123	1.23
do a course	4	23	√	√	B1	13	0.13
have a party	4	23	√	√	A2	40	0.4
have a problem	4	23	√		A2	277	2.77

Table 6.9 clearly shows that the frequencies differ. This is not surprising given that we are comparing classroom data with a baseline for all spoken language (in the BNC). What we notice is that the PMW frequencies are far higher in the ACE data and this may well relate to the classroom context where spoken activities are promoting focused uses of language around specific topics whereas the BNC represents many genres of spoken language much of which may be lexically less dense e.g. casual conversation (see McCarthy 1998). For instance, *do something* is the most frequent BNC pattern with a PMW frequency of 22.68 PMW compared with 29 times PMWs in ACE but the most frequent item in ACE (*have a look*) has a PMW frequency of 247 times PMW (compared with 19.09 times PMW in the BNC spoken data). While the BNC comparison shows up a differing pattern, it is nevertheless useful because it puts the learner uses in perspective not least of all when we see items that have a low rate in the BNC that are frequent even at A2 level in the ACE data. A case in point is *have a problem,* which this A2 learner used (marked correctly for tense)*:*

Extract 6.13 (Student <$4> A2 spoken corpus):

…We **had problems** finding a house.

In addition, and as expected due to the nature of the classes, verb followed by adjective pattern *speak English* occurs at a higher frequency in the ACE corpus with 76 occurrences (PMW) compared to the BNC's 1.72. This delexical verb is benchmarked at A1 level in the EVP as a pattern that learners are expected to know (even though it is not a high frequency pattern in the native speaker baseline of the BNC). Additionally, the pattern *improve my English* occurs 47 times (PMW) in the spoken ACE corpus and only once in the native speaker BNC and *do homework* also occurs higher in the ACE corpus with 23 occurrences (PMW) while it occurs only 0.12 times (PMW) in the BNC. When we look at the patterns listed in Table 6.9 in the EVP, we see that they are mostly considered A1 and A2 level items. Therefore, they are seen as items that a learner is

expected to know yet this 'coreness', in most cases, this is not mirrored in frequencies that we find in the BNC.

6.3.4 Delexical verbs in the written component of ACE

We now look at the 20,000 word written sub-corpus of the ACE, which is made up of student essays and homework (see Chapter 4). Manual analysis of the corpus identified the following most frequent delexical verb patterns (with a cut off of > 4 occurrences) and compares these with their expected level in the EVP (see Table 6.10).

Table 6.10 Most frequent (> 4 occurrences) of delexical verbs patterns in the Written ACE, by level compared with the EVP expected level.

Delexical pattern	Freq. (raw)	Freq. (PMW)	A2 cohort	C1 cohort	EVP level
have a party	12	70	√		A2
take a break	12	70	√		B1
get a job	10	58	√		A1
go swimming	10	58	√		A2
make money	10	58		√	B1
get married	9	52		√	A2
have an opportunity	9	52		√	B2
have a problem	8	47	√		A2
give advice	8	47	√	√	A2
go shopping	7	41		√	A1
take a shower	7	41	√		A2

write a letter	7	41	√		A2
save money	6	35	√		A2
take risks	6	35		√	B2
face a problem	5	29		√	C1
get the joke	4	28	√		C1

In terms of comparison with the EVP, from Table 6.10, we can summarise that 8 of the patterns used by A2 learners are at or below A2 level while two patterns are at B1 and C1, well above the expected level. For C1 learners, the patterns they used are assumed to have been acquired at lower levels, bar one (*face a problem)* which is expected to be acquired at C1. In terms of structure we note that the most frequent form in the written corpus is also the verb followed by noun structure. In terms of the actual range of patterns, we can see that the most frequent delexical verbs in the written ACE corpus are *have a party* and *take a break* which both have the same frequency of occurrence, 70 (PMW) in the A2 and C1 data. Both examples appeared in the A2 dataset despite the EVP stating *take a break* appears at B1 level.

Give advice is an example of a pattern that occurred in both the A2 and C1 datasets, it is linked with A2 level on the EVP. This is an example of an item therefore that one would expect to see in the data but we do find some challenges in its correct use among A2 learners as Extract 6.14 shows below. As we can see, Student <$4> shows an ability to use the collocation pattern in a way that suggests understanding of its meaning but is not successful in using the right form of the verb within the pattern.

Extract 6.14 (Student <$4> A2 written corpus)

…She go to people and **giving** them some **advice**.

Some items that occurred below the > 4 cut off for occurrences in Table 6.10 merit brief consideration. *Have a chance* occurs three times (17 PMW) among C1 learners as does the idiomatic delexical verb pattern *build a relationship*: At C1, Student 15 writes:

Extract 6.15 (Student <$15> C1 written corpus)

…as a result, we can **build relationships** fill our lives.

C1 students responding to a picture prompt from *Grammar Activities* (Ransaw, 2010) had to write a text of 150 words. On the one hand, Student 23 wrote:

Extract 6.16 (Student <$23> C1 written corpus)

…The boy was shy and embarrassed he **goes red** when the girl looked at him

Despite the error in the tense of the delexical verb, the idiomatic verb followed by adjective is correct. Furthermore, this class were taught and practised the verb *blush* in a previous class so it is interesting to note that this student chose the delexical alternative. While on the other hand, Student <$17> said that the picture reminded her of a movie she had seen and she finished describing the plot with the following sentence:

Extract 6.16 (Student <$17> C1 written corpus)

…I advise you to take many tissues because you will **shed tears**

The EVP places *go red* at B1 level while *shed a tear* appears at C2.

The researcher found it interesting that these students accurately used the metaphorical delexical verb patterns in their writing and both examples were found during the same class and actually the same task.

The next stage of analysis is to compare the findings of the written data with that of the delexical verbs in the BNC in order to compare frequencies of the learner use with native speaker writing. Table 6.11 compares the most frequent verb patterns from Table 6.10 (along with their EVP level) with their raw and PMW occurrences in the BNC.

Table 6.11 Comparison of the most frequent delexical patterns (> 4) in Written ACE with BNC

Delexical pattern	**Freq. (raw)**	**Freq. (PMW)**	**A2 cohort**	**C1 cohort**	**EVP level**	**BNC Written (Raw)**	**BNC Written (PMW)**
have a party	12	70	√		A2	40	0.4
take a break	12	70	√		B1	93	0.9
get a job	10	58	√		A1	293	2.93
go swimming	10	58	√		A2	173	1.73
make money	10	58		√	B1	300	3
get married	9	52		√	A2	485	4.85
have an opportunity	9	52		√	B2	423	4.23
have a problem	8	47	√		A2	277	2.77
give advice	8	47	√	√	A2	193	1.93

go shopping	7	41		√	A1	78	0.78
take a shower	7	41	√		A2	22	0.22
write a letter	7	41	√		A2	106	1.66
save money	6	35	√		A2	242	2.42
take risks	6	35		√	B2	29	0.29
face a problem	5	29		√	C1	3	0.03
get the joke	4	28	√		C1	7	0.07

Table 6.11 points to the following observations when we look at the learner data, the EVP levels and the BNC frequencies:

1) the rank order of frequency of items is quite different in the BNC compared with ACE. The top five items in ACE are: *have a party, take a break, get a job, go swimming, make money* while the top five in the BNC are: *get married, have an opportunity, get a job, have a problem, save money.* This shows up the real-life nature of the BNC which is reflected in these items and it also reflects the more classroom nature of the patterns in ACE.

2) the patterns are used far less frequently overall in the BNC. For example, the most frequent pattern in the BNC, *get married*, is used 4.85 times PMW while the most frequent items in the ACE (*have a party, take a break*) occur 70 times PMW. Again this points to the classroom being a place where certain types of language patterns occur intensively (*take a break* is a good example of this).

3) the items in the learner data though they seem not to reflect the frequency of the BNC, they are generally in line with what is expected of learners between A2 – C1 levels when we look at the EVP. This again suggests that there is a classroom or syllabus factor at play.

6.4 A closer examination of students' delexical verb use

In this section, we will examine some of the spoken and written utterances made by the students. We will also discuss their formations and their functions, and also focus on the students' attempt at delexicality. Opportunity of use (Buttery and Caines, 2012) and classroom task effect will be taken into consideration to ascertain if the students' use of such lexical features was influenced by the various language learning topics covered during classes. All results are (PMW) to make them comparable with the BNC.

The following shows the main delexical verbs drawn from the lower level mini corpus. It is important to note that these students had the same opportunity of use as all students, in that they also completed classes on the theme of directions, phrasal verbs, and idiomaticity. At the time of research, they were part of the FETAC level 3 group, roughly A2 in the CEFR. Some were of a fairly low level of English, but still they contributed examples of delexicality. This could be accredited to their length of time submersed in the English speaking culture of Ireland.

Within the A2 cohort, Student <$15> uttered the most examples of delexical verbs. He used phrases such as:

- ***Get advice***-verb followed by noun form, transparent in meaning, due to the noun *advice.*
- ***Make a decision***-verb followed by noun structure, transparent in meaning and could be substituted with the verb *decide*.
- ***Go on holiday***: this delexical example contains the formation of verb followed by preposition followed by noun and is also fairly transparent in meaning and could be substituted (grammatically but not lexically) with the single unit *holiday.*

Student <$4> used the following phrase and chose to use the American English noun of *vacation* rather than the British English noun *holiday* in:

- ***Went on vacation****:* again this phrase takes the formation of the past tense of the irregular verb *go* and noun *vacation* joined with the preposition *on* this is a delexical example as one

does not physically go to a place named *holiday* or in this case, go on something called *a vacation*.

Student <$5> (A2 cohort) articulated the following delexical verbs:

- ***Have lunch:*** a typical example of the delexical function of the irregular verb *have,* this verb is followed by the concrete noun *lunch*, again one does not physically possess an object titled *lunch,* but the meaning is nevertheless transparent as not much inference is needed in order to comprehend the statement. This phrase could also be substituted with the verb *to lunch.*
- ***Get stressed****:* This phrase follows the formation of irregular verb *get* succeeded by the adjective *stressed*. The verb *get* here is almost completely devoid of meaning in that it does not convey any meaning on its own. It is idiomatic in meaning as one does not actually receive an item that is *stressed*.
- ***Have a question****:* An example of a delexical verb with the noun *question* preceded by the verb *have*. This is a typical delexical collocation as one does not physically have or possess an item called a *question*. It is transparent in meaning and could be replaced with the lexical verb *to question*.

So far we have analysed the main delexical occurrences in the ACE corpus by focusing on the lower level, A2 cohort. Next, we will look at the occurrences in the higher level, C1 cohort. Expectedly, the higher cohort contributed more examples with a ratio of 244 (PMW) to 168 (PMW). This is ascribed to their level of fluency and length of time of residence in Ireland. Student <$18> used two very idiomatic delexical phases:

Extract 6.18 (Student 18 C1 written corpus)

ah to ah to to am **make sure** that this person is the same in your stereotypes what you think that can be changed

This phrase contains the formation of verb followed by adverb. This is a very idiomatic example as without the adverb *sure* the verb *make* is vacant of meaning, and for a second language speaker, there is a degree of interpretation required in order to understand the statement. It could also be substituted with the verb *to ensure.*

Student <$18> used the pattern: ***get lost.*** This phrase was stated by the speaker during the directions class and again is an example of a delexical verb; although transparent in meaning it is delexical in that one does not get or receive something named *lost*.

Student <$7> used the following two delexical items:

- ***Have a talk***
- ***Get dark***

The formation of delexicality here is that of a verb followed by an adjective, it is delexical in that a noun titled *dark* does not exist and receiving something called *dark* is not possible. It could be replaced by the verb *to darken*.

Student <$24> contributed the following delexical verb:

- ***Make an offer***

The frequent recurring structure of verb followed by noun appears again here. The example is idiomatic in meaning as an *offer* is not physically made or constructed, it could also be replaced with the single lexical verb *to offer*.

The above findings clearly evidence the use of delexicality by the ESL students at the college. Where single word substitution was possible, it was frequently ignored and the speakers opted for the delexical option instead. These aid in making the speakers' language sound more fluent and native like. Not so surprising is the fact that the higher level C1 students used delexicality more than the lower A2 cohort. All students had the same opportunity of use in that both groups were taught the same topics, directions, phrasal verbs, and idiomaticity, but their use could be attributed to their immersion in the language. Students also made delexical attempts but made either a

grammatical or lexical error. Table 6.12 shows the students who made an attempt at delexicality in their speech but made some minor errors in the utterance.

Table 6.12 Student attempts and errors at delexicality

Speaker	Level	Utterance	Correction
<$1>	C1	Go dinner	Go to dinner
<$1>	C1	When she **do** something	When she does something
<$5>	A2	We hav**e** lunch	We had lunch
<$7>	C1	They wanted to do shopping	They wanted to do the shopping
<$7>	C1	Last week I **go** on am holiday	Last week I went on holiday
<$13>	A2	Okay I lost I I try to go **to** walking	Okay I lost I I try to go walking

Let us now look more closely at some individual learners. In this section, I have chosen four students' use of delexical verbs to analyse in further detail. Two are taken from the lower A2 cohort and two are taken from the higher cohort C1 group and they were chosen randomly from the 24 students. Note that all student profile information is contained in Appendix 3. It is important to note, that students from the lower cohort did not take part in every recorded class.

The first student analysed from the lower cohort was student <$15>, a twenty-two year old female from Saudi Arabia.

Table 6.13 Student <$15> (A2 cohort)

Date	Delexical verb	Form	Meaning	How many utterances by student (Raw)	How many utterances by student (PMW)
8/3/11	go on holiday	Verb + Noun	Transparent	7	41
22/3/11	make a decision	Verb + Noun	Transparent	7	41
7/2/11	give me address	Verb + Noun	Transparent	3	17.6
8/3/11	get advice	Verb + Noun	Transparent	2	11.7

On the 8/3/11 this student used the delexical *get advice* which takes the regular formation of verb followed by noun, she also used on that day the delexical verb *go on holiday,* which again has the structure of verb and noun and could be substituted with the single verb *to holida*y. On the 22/3/11 she used the following utterance *make a decision,* which again has the formation of verb followed by noun and could be replaced with the lexical verb *to decide*. None of this student's delexical uses were idiomatic, instead they were all transparent in meaning. However, it must be noted that this student is a member of the lower cohort and still managed to use delexicality accurately. It is also interesting to note that this student was part of the class given on asking for and giving directions but she did not use any of the phrases associated with that theme.

The next student chosen was from the A2 cohort. Student <$17>, a female, forty seven years old from Pakistan.

Table 6.14 Student <$17> (A2 cohort)

Date	Delexical Verb	Form	Meaning	How many utterances by student (Raw)	How many utterances by student ((PMW))
5/4/11	have a question	Verb + Noun	Transparent	8	47
15/3/11	make mistakes	Verb + Noun	Transparent	6	35
15/3/11	go crazy	Verb + Adjective	Idiomatic	2	11.7
5/4/11	I've got charm	Verb + Noun	Idiomatic	1	5.8

During a class on directions, this student uttered two delexical verbs associated with the theme and both examples contain the lexical verb *go.* They are *go straight ahead* which has the formation of verb and adverb and *go back* which has the same formation, but could be replaced with the lexical verb *return.* It was predicted beforehand that such classroom task effect would elicit such examples. During the same class, she uttered the phrase *have to pick one*, which is formed by using verb followed by determiner and could be replaced with the lexical verb *choose*. *Pick one* occurs at a frequency of 0.69 (PMW) in the BNC and appears at B1 level in the EVP. However, it is an example of an idiomatic meaning where the meaning is not clear, especially to an ESL/EFL student. And, while taking part in a class based on idioms, she contributed the phrase *I've got charm*, which is indeed highly idiomatic and is structured with verb and verb and noun.

The final student from the lower cohort to be analysed here is Student <$4>, a twenty-eight-year-old female from Italy. This student only contributed one example of a delexical verb.

Table 6.15 Student <$4> (A2 cohort)

Date	Delexical Verb	Form	Meaning	Raw Results	(PMW) Results
15/2/10	went on vacation	Verb+Noun	Transparent	1	5.8

She also uttered the Americanism *went on vacation* as opposed to the standard British *went on holiday.* Again, this student took part in both the directions and the idioms classes, yet she did not utter any of the phrases that said seminars should have elicited from her. Again, it must be noted that all students had the same opportunity to do so.

Having looked at three students from the lower cohort this chapter will now focus on learners from the higher cohort.

Student <$5> is a twenty two year old female student from Spain. She used the following delexical verbs throughout the course.

Table 6.16 Student <$5> (C1 cohort)

Date	**Delexical Verb**	**Form**	**Meaning**	**How many utterances by student (Raw)**	**ow many utterances by student (PMW)**
22/3/11	have lunch	Verb + Noun	Transparent	7	41
12/4/11	get stressed	Verb + Adjective	Idiomatic	5	29

15/3/11	make a decision	Verb + Noun	Idiomatic	4	22
7/2/11	make an offer	Verb + Noun	Idiomatic	3	17.6

Clearly, this student used a large number of delexical verbs and she was able to construct these collocations accurately. *Make a decision* is structured with verb followed by noun and could be substituted with the single lexical verb *to decide*. It is interesting that she constructed this phrase perfectly, as through L1 interference many students get this incorrect and instead use *do*. Her earlier uses of *have an offer* and *went home* on the 8/3/11 are both structured with verb followed by noun and could both be replaced with *offer* and *leave* respectively. Again, it is interesting to note that this student was present for the directions class but did not seem to use any of the learned and practised phrases. Be that as it may, she used three delexical structures: *went to town* and *have lunch* which are formed by verb followed by noun and *go back home,* which is formed by the combination of verb and adverb followed by noun and could be replaced with *return*. Furthermore, she contributed the phrase *take after* during a class based on phrasal verb usage. This phrase is structured with verb followed by adverb, is extremely idiomatic, and could be substituted with the verb *to resemble.* Finally, during a discussion on the upcoming examinations, this student states that she *gets stressed,* which has a transparent meaning and is structured with verb followed by adjective.

Table 6.17 Student <$24> (C1 cohort)

Date	Delexical verb	Form	Meaning	How many utterances by student (Raw)	How many utterances by student (PMW)
8/3/11	ask a question	Verb + Noun	Transparent	9	52
15/3/11	go on holiday	Verb + Noun	Transparent	7	41
15/3/11	make a decision	Verb+ Noun	Transparent	7	41
6/4/11	do a course	Verb + Noun	Idiomatic	4	23
20/4/11	go crazy	Verb + Adjective	Idiomatic	2	11.7

Student <$24> from Saudi Arabia was one of the more advanced students in the group. He scored high on both oral and written tests. Speaker <$24> clarified *can I ask a question?* He repeated this structure eight more times during the same lesson. The verb *ask* followed by the noun *question* could be viewed as an example of a delexical verb it is of course a matter of degree and it is not absolute *as ask a question* could also be seen as a common collocation. *Ask a question* has a frequency of occurrence in the BNC at 1.1 times (PMW) and appears at Beginner level, A1 of the EVP. Then Student <$24> said the following *go on holiday* and *make a decision,* both clear examples of delexical verbs. Overall, he said *go on holiday* a total of seven times during this class

and *make a decision* a total of four times. *Go on holiday* appears at a frequency of 1.2 times (PMW) in the BNC while *make a decision* has an occurrence of 2.47 times. *Go on holiday* is given as an example of A1 proficiency and make a decision of B1 on the EVP. On examination of Student 24's writing, the following two delexical verbs appear: *do a course* and *go crazy*. In the BNC *do a course* appears 0.13 times while *go crazy* was found to have a frequency of 0.4 (PMW). *Do a course* is given at A1 level while *go crazy* is cited at B1 level of the EVP.

6.5 Conclusion

The analysis carried out in this chapter identifies the frequency and forms of delexical verbs used by two cohorts of ESOL students of the ACE corpus. Both spoken and written contributions of the A2 and C1 learners were analysed. It was considered vital by the researcher to, firstly identify the number of delexical verbs being used in the English language. In order to successfully do this, the British National Corpus (BNC), a 100 million word collection of native speaker written and spoken language, was analysed with a result of 19,400 delexical verbs. Such a high result in examples serves to confirm the theory that these multi-word units are an important feature of native speaker English. The analysis showed ample evidence of learners, at both A2 and C1 level, using these items, with a total frequency of 334 occurrences. As discussed above, more delexical verb patterns were found in the spoken data overall and within the data, A2 learners used them far more in speaking than in writing. On the whole learners were using forms that one would expect them to use based on the EVP, with some exceptions, for example, where A2 learners used items that are not normally used until higher levels.

Where we found greatest difference when we compared the frequencies of the delexical verb patterns found in the learner data with their PMW frequencies in the BNC. Apart from the spoken item *have a look*, which was the most frequent item in both the learner spoken data and in the Spoken component of the BNC by comparison, the rank order of items by frequency differed greatly. As discussed above, it showed up a difference in the types of language items that are used

in the real-world environment across all genres in the BNC when compared with the classroom environment. However, crucially, the EVP showed the learners' use to be generally in line with expectations and this points to a classroom or syllable effect in the language that learners are frequently experiencing (e.g. *take a break, have a look, ask a question,* etc.) and the patterns that they are acquiring and using. As discussed above and in Chapter 1, this ties in with the usage-based model for language acquisition.

Related to this might be the possibility of a limitation created for opportunity of use – where it could be argued that learners do not get the chance to use all of the range of items that they might know or that they might be using more of certain types of patterns because texts or lessons might be focussing on them. In this chapter, we did attempt to ascertain whether there was a classroom task effect that would explain the high frequency of occurrence of delexical verbs. In accordance with Buttery and Caines (2010), the opportunity of use for the subjects of the language being assessed in a corpus is fundamental in identifying the ratio and number of occurrence. The research for this thesis was gathered during live three hour EFL classes over the course of a 12 week semester. Naturally, the language topics and learner objectives covered during the course had a substantial effect on the language the learners produced and task effect did indeed occur. However, in contrast to some prior assumptions, the task effect did not produce a lot of delexical usage by all students. For example, on the 4/3/11 the researcher held a session on the theme of directions, but not all students used delexical verbs during this seminar. Such a class should have elicited phrases such as *take a left* and *go straight ahead* and so on. The same could be said for the class on phrasal verbs on the 6/3/11, where a class was taken from *Headway Intermediate* and the idioms class on the 5/4/11. This shows that a classroom/task effect was not always successful in activating delexical schemata since both cohorts had the same opportunity to use delexical verbs. Returning to the usage-based model, this suggests that overt teaching of these patterns (focus on form(s)) did not seem to have a major effect and points again to learners using patterns that they must have encountered frequently (inside or outside the classroom) rather than those that have been overtly

taught to them within the 12-week study period. While this was not empirically tested, we clearly know what the overt lesson objects and materials were for each lesson or the nature of each writing task.

As expected more learners in the higher C1 cohort used more delexical verbs compared with the fewer A2 learners who used fewer delexical verb items overall. The verb followed by noun appeared as the most popular structure across all corpora as was the transparent meaning rather than idiomatic/opaque. Analysis of the available literature identified the meaning of such utterances appearing on a cline from transparent or literal in meaning to idiomatic or opaque, as discussed above. The analysis showed that, the literal structures occurred most in the learner corpus suggesting learners' preference for communicating literally rather than metaphorically. Though this could also be explained by the fact that most delexical patterns occurred in speaking and it may suggest that learners did not want to risk misunderstanding in face-to-face real-time conversation by using opaque patterns.

In summary, this data analysis chapter not only answers the question as to whether the learners of the ACE corpus use delexical verbs in their speaking and writing, but, also identifies, despite the prevailing abundance of literature to the contrary, the lower level A2 learners at ACE are also able to use delexical verbs and in some cases this is at a higher level than expected. The next chapter will focus on the language feature of collocations and ascertain the use of collocations by the ESOL learners of the ACE corpus.

Chapter 7 Collocations

Words hunt in packs
(Thornbury, 1988: 8)

The choice of one word conditions the choice of the next, and of the next again. The item and the environment are ultimately not separable, or certainly not separable by present techniques' (Sinclair, 2004)

7.0 Introduction:

This chapter examines a further multi-word lexical item, namely the phenomenon of collocation. The term was coined by J. R. Firth in 1951 when he stated 'I propose to bring forward as a technical term, meaning by collocation and apply the test of collocabilty' (1957: 194). Firth identified how the English language is composed of words and units that can both combine and occur frequently together. These combinations are termed ***collocations***. Firth (1951) first introduced the term *collocation* in his 1951 paper *Modes of Meaning*. He explained collocabilty with the unit '*ass*'. He stated that one of its main and frequently occurring meanings arises from its habitual occurrence with the phrase *you silly*…. A frequent example used to explain collocation is *blonde hair* (McCarthy, 1990: 12), which can be juxtaposed with examples like *yellow hair*, or *blonde T-shirt*, for instance, to show the patterned and habitual nature of the first pattern in contrast to the second and third examples. The adjective, *blonde,* is labelled the 'node' whereas the noun, *hair,* is referred to as the 'collocate' (the unit that occurs in specified environment of the node). The importance of such patterns in language was further emphasised by the Neo-Firthians Halliday (1966), Sinclair (1996) and Aisenstadt (1981) who used corpora as a tool for analysing, with ease, such co-occurrences in a body of texts.

Similar to the topics of the three previous chapters, collocations are a common feature of native English language in use. McCarthy and O'Dell (2006) demarcate collocations as 'a pair of words that are often used together. These combinations sound natural to native speakers but students of English need to make a special effort to learn them because they are often difficult to guess' (2006: 1). Based on McCarthy and O'Dell's assertion, it was considered important to identify what, if any collocations the ESL learners of the ACE corpus use in their speaking and writing. This chapter will

analyse the contributions of both cohorts of the ACE for collocations. The chapter will first define and give an overview of the topic of collocations before presenting the analysis of the ACE spoken and written corpora.

7.1 Four definitions of Collocation

This section will provide four differing definitions of collocation. It is important to highlight the various definitions of collocation because the four definitions shown here contribute to both a native and non-native English speaker's overall understanding of the topic.

1) O'Keeffe, McCarthy and Carter (2007: 59) state that 'collocations are not absolute or deterministic, but are probabilistic events, resulting from repeated combinations used and encountered by the speaker of a language'. On explaining collocation, O'Keeffe et al (2007) give the example of the verb *bark's* high probability of occurrences with the noun *dog* and the impossible combination with the noun *cat*. This definition of collocation will be the central definition of this research.

2) *The Oxford Advanced Learner's Dictionary* defines a collocation as 'a combination of words in language that happen very often and more frequently than would happen by chance' (Hornby, 2005: 293). Hornby (2005) shows collocations as prefabricated units, stored together in the mental lexicon, retrievable when necessary. It has been proven that they are not mere *ad hoc* occurrences and this constitutes the main reason for focusing on this definition.

3) Sinclair (1991) noted collocation as 'the co-occurrence of two or more words within a short space of each other in a text where the natural measure of proximity is a maximum of four words' (1991: 170). Sinclair was the first linguist to emphasise the crucial aspect of distance between the node and collocate in a collocation.

4) Benson, Benson and Ilson (1997) define collocations as specified, identifiable and non-idiomatic, recurrent combinations.

For the purposes of this chapter, the broad definition, based on repeated combinations, of O'Keeffe, McCarthy and Carter (2007) will be used.

7.2 Collocation Overview

As mentioned, it was J.R. Firth who first coined the term 'collocation'. Regarded as 'the father of collocation' (Carter and McCarthy, 2014), Firth (1951) strived to differentiate between habitual collocations and other mere co-occurring lexical items. But Firth's definition and explanation of the phenomenon of collocabilty has been criticised by many scholars, who considered his description to be 'too vague' (see Benson, Benson and Ilson (1997) and Leon (2005)). The mastery of such exponents by native and non-native speakers is seen as a central aspect of communicative competence. Henriksen (2011) states that collocations enable the native speaker to process language 'both fluently and idiomatically, fulfilling basic communicative needs' (2011: 1).

Barfield and Gyllstad (2009b), Nation, (2001), Schmitt (2004), Wood (2010) and Wray (2002), all identify collocations and language chunks as an integral component of native speaker discourse that influences greatly the communicative competence of the speaker. According to Bashi (2004), collocational competence in an L2 is 'part and parcel' of overall language competence. As learners add collocations to their language repertoire, their ability to communicate increases rapidly.

In attempting to differentiate collocations from other lexical co-occurrences, Cowie and Howarth (1996) claim that collocations are characterised as

1. **Institutionalised**: that is collocations are accepted and used within the culture and language they are expressed.
2. **Memorized:** they are stored in and retrievable from the mental lexicon of the language user.
3. **Restricted:** collocations are fixed and therefore restricted. The combinations cannot be altered.

4. **Semantically opaque units**: the meaning of collocations is for the most part opaque and need to be interpreted correctly rather than transparent.

Cowie and Howarth (1996) give the following examples of collocations:

a) ***Major catastrophe*** - it is fixed and restricted as the adjective cannot be substituted with a synonym such as *big.*

b) ***Blonde hair*** - the adjective cannot be substituted with the near synonym *yellow*.

c) ***Strong coffee*** - the adjective cannot be replaced with the synonym *powerful.*

7.3 The types of Collocations

There are many differing types of collocations. The main four in the literature are:

- **<u>Adjective and Noun Collocation</u>**

Nouns tend to have typical adjectives which they combine with, for example, *the real thing* versus *the genuine thing* and *the genuine article* versus *the real article, a broad summary* and *not a wide summary, great detail* and *not big detail.*

- **<u>Verb and Adverb Collocation</u>**

For example: *she drives too quickly*, *let's move quickly*, *she drives too fast*, *let's move fast let's move swiftly.*

- **Adverb and Adjective Collocation**

Certain adverbs typically modify particular adjectives, for example, the adverb of degree *utterly. Utterly* tends to occur most frequently with negative adjectives for example: *utterly depressed*, *utterly ridiculous* and *utterly naïve*.

- **Verb and Object Collocation**

Verbs and their objects often form various and different collocations, such as in the expressions *raise your hand*, *raise a family*, *visit*, *go to or check out a website*.

Benson et al (1997) viewed collocation as a feature of both grammar and lexis. They identified two main groups, grammatical and lexical. *Grammatical collocations* combine or associate a lexical unit with a grammatical unit, for example, *good at* or *afraid of*. A *lexical collocation*, on the other hand, often consists of adjectives, nouns, verbs, and adverbs: *utterly stupid* or *richly decorated*. John Sinclair's *Corpus, Concordance and Collocation* (1991) is accredited as the most widely used and accepted explanation of collocation. According to Sinclair (1991), the meaning of a word is not only dependent on the context in which it occurs but also the lexical and grammatical elements with which the word co-occurs or collocates (Sinclair 1991: 108). Furthermore, Carter and McCarthy (2006: 8) postulate:

> The notion of collocation shifts the emphasis from the single word to pairs of words as integrated chunks of meaning, and collocation has become an uncontroversial element in a good deal of language description and pedagogy.

Neo-Firthians also identified how words co-occur grammatically in what is termed *colligation* and is the syntagmatic combination of parts of speech and grammatical categories. Sinclair (1998: 15)

explained colligation with the verb *budge* and the fact that its majority of co-occurrences are preceded by a modal auxiliary verb like in the example *he will/he won't budge*.

Another important consideration in relation to *collocation* is the strength of the relationship between the two words. As noted by Carter et al (2011), collocations may be *strong* or *weak* in their links. In the case of strong collocations, the link between the two words is quite fixed and restricted. For example (based on Carter et al (2011), consider the link between *make/express/fulfil + a wish.* Because very few words can collocate with the noun *wish,* this makes *wish* a strong collocator. In comparison, weak collocations are where a word can collocate with many other words. For instance, the word *big* can collocate with hundreds of other words (*big house, big car, big dog, big man,* and so on), therefore it is a weak collocator.

7.4 Data Analysis and the ACE corpus

7.4.1 Overview of results

We now focus on the ACE corpus and explore whether or not the A2 and C1 learners use collocations in their classroom spoken interactions and learner writing. In terms of methodology, the transcriptions of the ACE spoken and written data across the A2 and C2 cohorts were exhaustively searched through manual sifting so as to identify collocations. Following that, the software *Wordsmith Tools* (Scott, 2016) was used to identify and extract frequency counts and concordance lists.

Table 7.1 shows us the distribution of students using collocations across both levels. In speaking, collocational patterns were identified across all learners while in writing 88% of A2 learners used correct collocations in their writing compared with 100% of C1 learners.

Table 7.1 Number and percentage of students using collocations across both cohorts

	Speaking		Writing	
	Raw	**%**	**Raw**	**%**
A2 (9 participants)	9	100%	8	88%
C1 (15 participants)	15	100%	15	100%
Total	24		23	

Table 7.2 is a summary of the overall frequencies of collocations across levels.

Table 7.2 Overall frequency of collocations in ACE by level (raw and PMW)

CEFR level	**Frequency (raw)**	**Frequency (PMW)**
A2	117	1648
C1	201	2030

When the frequencies in Table 7.2 are further broken down, by level and by speaking versus writing, in the ACE sub-corpora, we find that collocations are widely used by both the A2 and C1 cohorts overall, see Table 7.3. When we examine their distribution by percentage, we find that 40% and 33% of spoken and written collocations, respectively, were used by A2 learners while 60% and 67% of collocations, respectively, were used by C1 learners (Table 7.3).

Table 7.3 Frequency of collocations by level in spoken and written ACE sub-corpora

Level	Spoken (raw)	PMW	%	Written (raw)	PMW	%
A2	76	1188	40	41	5857	33
C1	116	1349	60	85	6538	67
Total	**192**			**126**		

While on one hand the higher frequency and percentage of use by C1 learners compared with A2 learners is not surprising, it is notable that A2 learners use so many collocations relative to C1 learners. It confirms the centrality of collocation in language use. Even at Elementary level (A2), learners have acquired and can use many collocations. We will explore below the nature of these uses.

7.4.2 Collocations in the spoken component of ACE

Having manually identified all collocations and then cross checked their frequencies using *Wordsmith Tools,* the EVP was then consulted to find at what level each collocation was expected occurred in learner language. In some cases, collocations used by learners, though correct in their pattern, were not listed in the EVP. In these cases, no level is provided in Table 7.4. The BNC was a useful counterpoint here also because even when a pattern did not feature in the EVP, we could still compare it with the native speaker baseline frequency PMWs. Table 7.4 below outlines the 20 most frequent collocations in the ACE spoken corpus in comparison with the EVP level and the BNC PMW frequency.

Table 7.4 The 20 most frequent collocations in the ACE spoken corpus compared with the BNC and their EVP level[10].

Collocation	A2	C1	ACE Frequency (PMW)	BNC Frequency (PMW)	EVP Level
of course		√	233	51	A1
kind of	√	√	200	27	B2
thank you	√	√	152	52	A1
spend money		√	104	2	A2
or something		√	96	24	A2
the last time	√		71	2	A2
flying saucer	√		70	0.03	-
blonde hair	√	√	62	0.2	A2
a bit		√	47	69	A2
Christmas time	√	√	47	0.5	-
summer time	√	√	47	0.2	-
I don't mind		√	35	3.5	A2
admire somebody		√	23	0.1	B2
spend time		√	23	4	B2
sounds good	√		23	0.9	A2
important information		√	23	73	A2
last year	√	√	20	0.66	A1

[10] Here and elsewhere, lemmas are used, e.g. *spend money* represents all other forms such as *spent money, spends money, spending money,* etc.

From Table 7.4 we can observe that the 20 most frequent collocations in the learners' speaking are many times less frequent (PMW) in the native speaker spoken corpus (Spoken BNC). For example, *of course* is used 233 times per million words in the learner data and this is almost five times more than in the BNC spoken data. The pattern *thank you* is used 52 times PMW in the BNC (which is the highest frequency of all 20 patterns from ACE) while the learners' speaking it is found to have a PMW frequency of three times this number. This might suggest that learners are over-using these patterns or perhaps that the classroom recordings are promoting repeated uses of these patterns. However, it is argued here and below that this is not the case. Another important point of comparison in Table 7.4 is with the EVP. For those items that were found in the EVP, all bar three are shown to be acquired by A1 or A2 level. In other words, while they may appear to be far less frequent in the BNC, most of the patterns listed in Table 7.4 are seen as collocations that learners acquire at Beginners' or Elementary level.

When we compare A2 learners' collocations with the EVP, we see that apart from *kind of*, all of the patterns are expected by A2. The C1 learners do not show use of items that are expected any later than B2 level. We will now take a qualitative look at some of the items listed in Table 7.4.

7.4.3 *Of course*

The most frequently occurring collocation in the ACE corpus is the idiomatic *of course* with 233 occurrences per million words. Sinclair (1991) used this discourse marker to explain collocation by identifying it as two individual words with one unitary meaning. *Of course* appears 233 times per million words in the ACE spoken corpus and it occurs with a frequency of 51 PMW in the Spoken BNC. Furthermore, *of course* only occurred in the C1 level which shows that it is the more advanced students which use this phrase. However, according to the EVP, learners at as low a level as A1 should be able to use the discourse marker. It was ascertained that the majority of occurrences of *of course* occurred as a response token with a number of students using it in reply to a question from the teacher.

Extract 7.1 (Student <$1> C1 spoken corpus)

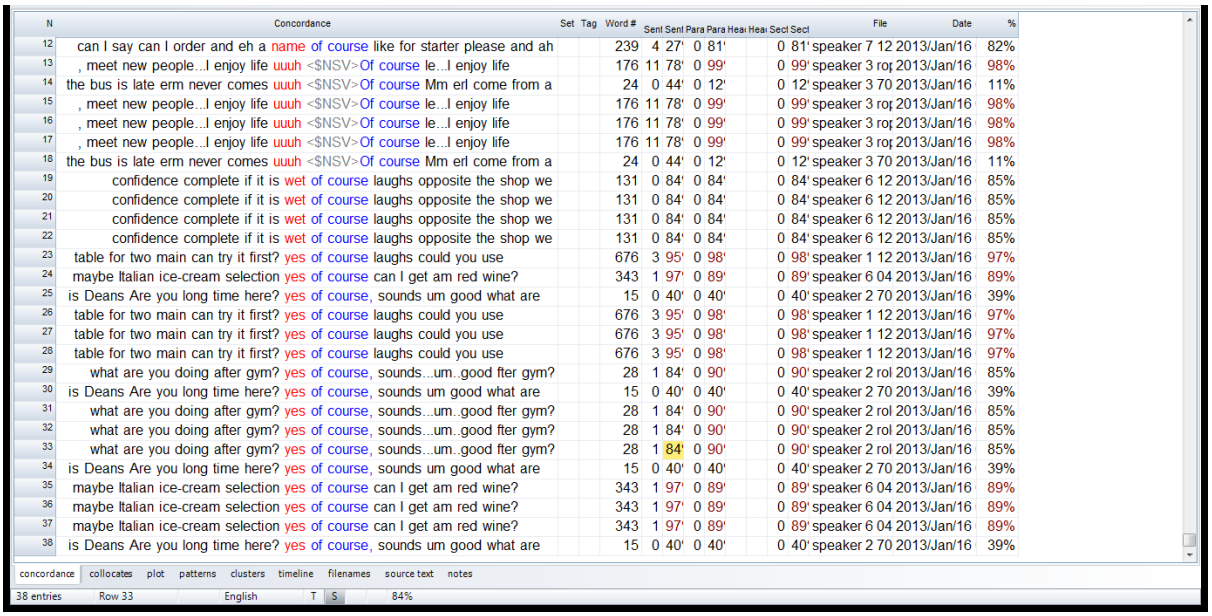
T: can you describe your life in this city?

Student <$1>: **of course**. I like living here because it's a friendly city.

This response token use in classroom interactions may explain why the item appears to be used much more in the learner data than in the BNC. The Spoken BNC sets out to represent all types of spoken interactions and so conversational interactions such as this will be represented to a lesser degree in the balanced Spoken BNC. For example, it will include monologic samples such as lectures, speeches, radio advertisements, sermons, and so on.

Fig. 7.1 below provides a concordance of *of course,* which further illustrates its role in responses.

Fig. 7.1 Concordance for *of course*



N	Concordance	Set	Tag	Word #	Sent	Sent	Para	Para	Head	Head	Sect	Sect	File	Date	%
12	can I say can I order and eh a name of course like for starter please and ah			239	4	27'	0	81'			0	81'	speaker 7 12	2013/Jan/16	82%
13	, meet new people...I enjoy life uuuh <$NSV>Of course le...I enjoy life			176	11	78'	0	99'			0	99'	speaker 3 ro	2013/Jan/16	98%
14	the bus is late erm never comes uuuh <$NSV>Of course Mm erl come from a			24	0	44'	0	12'			0	12'	speaker 3 70	2013/Jan/16	11%
15	, meet new people...I enjoy life uuuh <$NSV>Of course le...I enjoy life			176	11	78'	0	99'			0	99'	speaker 3 ro	2013/Jan/16	98%
16	, meet new people...I enjoy life uuuh <$NSV>Of course le...I enjoy life			176	11	78'	0	99'			0	99'	speaker 3 ro	2013/Jan/16	98%
17	, meet new people...I enjoy life uuuh <$NSV>Of course le...I enjoy life			176	11	78'	0	99'			0	99'	speaker 3 ro	2013/Jan/16	98%
18	the bus is late erm never comes uuuh <$NSV>Of course Mm erl come from a			24	0	44'	0	12'			0	12'	speaker 3 70	2013/Jan/16	11%
19	confidence complete if it is wet of course laughs opposite the shop we			131	0	84'	0	84'			0	84'	speaker 6 12	2013/Jan/16	85%
20	confidence complete if it is wet of course laughs opposite the shop we			131	0	84'	0	84'			0	84'	speaker 6 12	2013/Jan/16	85%
21	confidence complete if it is wet of course laughs opposite the shop we			131	0	84'	0	84'			0	84'	speaker 6 12	2013/Jan/16	85%
22	confidence complete if it is wet of course laughs opposite the shop we			131	0	84'	0	84'			0	84'	speaker 6 12	2013/Jan/16	85%
23	table for two main can try it first? yes of course laughs could you use			676	3	95'	0	98'			0	98'	speaker 1 12	2013/Jan/16	97%
24	maybe Italian ice-cream selection yes of course can I get am red wine?			343	1	97'	0	89'			0	89'	speaker 6 04	2013/Jan/16	89%
25	is Deans Are you long time here? yes of course, sounds um good what are			15	0	40'	0	40'			0	40'	speaker 2 70	2013/Jan/16	39%
26	table for two main can try it first? yes of course laughs could you use			676	3	95'	0	98'			0	98'	speaker 1 12	2013/Jan/16	97%
27	table for two main can try it first? yes of course laughs could you use			676	3	95'	0	98'			0	98'	speaker 1 12	2013/Jan/16	97%
28	table for two main can try it first? yes of course laughs could you use			676	3	95'	0	98'			0	98'	speaker 1 12	2013/Jan/16	97%
29	what are you doing after gym? yes of course, sounds...um..good fter gym?			28	1	84'	0	90'			0	90'	speaker 2 rol	2013/Jan/16	85%
30	is Deans Are you long time here? yes of course, sounds um good what are			15	0	40'	0	40'			0	40'	speaker 2 70	2013/Jan/16	39%
31	what are you doing after gym? yes of course, sounds...um..good fter gym?			28	1	84'	0	90'			0	90'	speaker 2 rol	2013/Jan/16	85%
32	what are you doing after gym? yes of course, sounds...um..good fter gym?			28	1	84'	0	90'			0	90'	speaker 2 rol	2013/Jan/16	85%
33	what are you doing after gym? yes of course, sounds...um..good fter gym?			28	1	84'	0	90'			0	90'	speaker 2 rol	2013/Jan/16	85%
34	is Deans Are you long time here? yes of course, sounds um good what are			15	0	40'	0	40'			0	40'	speaker 2 70	2013/Jan/16	39%
35	maybe Italian ice-cream selection yes of course can I get am red wine?			343	1	97'	0	89'			0	89'	speaker 6 04	2013/Jan/16	89%
36	maybe Italian ice-cream selection yes of course can I get am red wine?			343	1	97'	0	89'			0	89'	speaker 6 04	2013/Jan/16	89%
37	maybe Italian ice-cream selection yes of course can I get am red wine?			343	1	97'	0	89'			0	89'	speaker 6 04	2013/Jan/16	89%
38	is Deans Are you long time here? yes of course, sounds um good what are			15	0	40'	0	40'			0	40'	speaker 2 70	2013/Jan/16	39%

concordance collocates plot patterns clusters timeline filenames source text notes

38 entries Row 33 English T S 84%

7.4.4 *Kind of*

The vague hedge *kind of* occurred in both levels of the spoken ACE corpus with 200 results per million words and it occurred 27 times PMWs in the BNC. Cutting (2000) and Carter and McCarthy (2000) identify such vague language or hedges as an integral part of native and natural

language in use and this is typified in this example from an A2 learner. Student <$4> from the A2 cohort gave this response while communicating with another learner:

Extract 7.2 (Student <$6> and <$4> A2 spoken corpus)

Student <$6>: you like your job?

Student <$4>: it's **kind of** interesting I guess

The comparison with the BNC shows *kind of* to be 7 times less frequent in native speaker discourse compared to the ACE data. However, as discussed above, this may be explained by the broad span of registers and genres that the BNC data represents. *Kind of* occurs in the spoken corpus at both A2 and C1 level. Despite the EVP stating that it occurs at the B2 level and above, the ACE corpus clearly shows that lower level learners are using this example of a hedge. In another pair work interaction, Student <$12> of the C1 cohort used *kind of* as a part of a vagueness marker (*all kinds of things*) when he gave the following example:

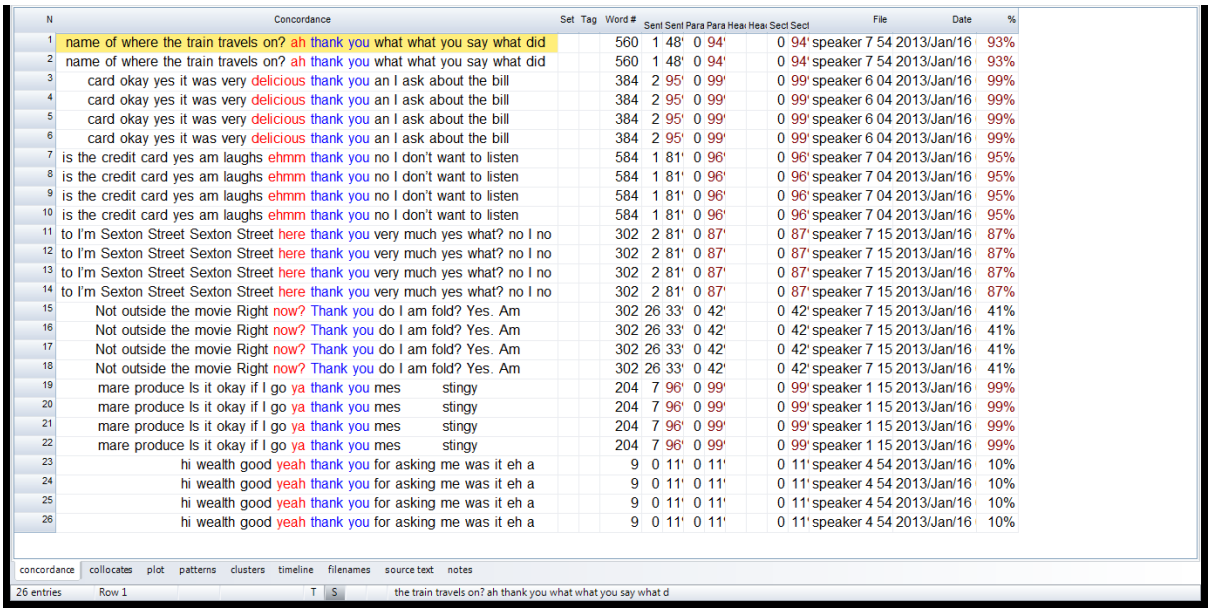
Extract 7.3 Student <$12> C1 spoken corpus)

…it was a busy weekend I had some much to do **all kinds of things.**

7.4.5 *Thank you*

Thank you occurred numerous times in the classroom as part of thanking routines, often within roleplay tasks, and as a result occurs with a high frequency in the Spoken ACE (152 PMW). Since *thank you* occurs so frequently in native speaker language, within thanking speech act routines, it was expected to have a high result in the ACE. Its frequency in the Spoken BNC is one third that of the ACE but, as discussed above, this is not surprising given the broader representation of the BNC. Within learner language, this is expected to be a high frequency pattern and universally acquired by Beginners' level (A1) according to the EVP. Fig. 7.6 below shows a concordance list for *thank you* in the corpus.

Fig. 7.2 Concordance list for *thank you*



N	Concordance	Set	Tag	Word #	Sent	Sent	Para	Para	Head	Head	Sect	Sect	File	Date	%
1	name of where the train travels on? ah thank you what what you say what did			560	1	48'	0	94'		0	94'		speaker 7 54	2013/Jan/16	93%
2	name of where the train travels on? ah thank you what what you say what did			560	1	48'	0	94'		0	94'		speaker 7 54	2013/Jan/16	93%
3	card okay yes it was very delicious thank you an I ask about the bill			384	2	95'	0	99'		0	99'		speaker 6 04	2013/Jan/16	99%
4	card okay yes it was very delicious thank you an I ask about the bill			384	2	95'	0	99'		0	99'		speaker 6 04	2013/Jan/16	99%
5	card okay yes it was very delicious thank you an I ask about the bill			384	2	95'	0	99'		0	99'		speaker 6 04	2013/Jan/16	99%
6	card okay yes it was very delicious thank you an I ask about the bill			384	2	95'	0	99'		0	99'		speaker 6 04	2013/Jan/16	99%
7	is the credit card yes am laughs ehmm thank you no I don't want to listen			584	1	81'	0	96'		0	96'		speaker 7 04	2013/Jan/16	95%
8	is the credit card yes am laughs ehmm thank you no I don't want to listen			584	1	81'	0	96'		0	96'		speaker 7 04	2013/Jan/16	95%
9	is the credit card yes am laughs ehmm thank you no I don't want to listen			584	1	81'	0	96'		0	96'		speaker 7 04	2013/Jan/16	95%
10	is the credit card yes am laughs ehmm thank you no I don't want to listen			584	1	81'	0	96'		0	96'		speaker 7 04	2013/Jan/16	95%
11	to I'm Sexton Street Sexton Street here thank you very much yes what? no I no			302	2	81'	0	87'		0	87'		speaker 7 15	2013/Jan/16	87%
12	to I'm Sexton Street Sexton Street here thank you very much yes what? no I no			302	2	81'	0	87'		0	87'		speaker 7 15	2013/Jan/16	87%
13	to I'm Sexton Street Sexton Street here thank you very much yes what? no I no			302	2	81'	0	87'		0	87'		speaker 7 15	2013/Jan/16	87%
14	to I'm Sexton Street Sexton Street here thank you very much yes what? no I no			302	2	81'	0	87'		0	87'		speaker 7 15	2013/Jan/16	87%
15	Not outside the movie Right now? Thank you do I am fold? Yes. Am			302	26	33'	0	42'		0	42'		speaker 7 15	2013/Jan/16	41%
16	Not outside the movie Right now? Thank you do I am fold? Yes. Am			302	26	33'	0	42'		0	42'		speaker 7 15	2013/Jan/16	41%
17	Not outside the movie Right now? Thank you do I am fold? Yes. Am			302	26	33'	0	42'		0	42'		speaker 7 15	2013/Jan/16	41%
18	Not outside the movie Right now? Thank you do I am fold? Yes. Am			302	26	33'	0	42'		0	42'		speaker 7 15	2013/Jan/16	41%
19	mare produce Is it okay if I go ya thank you mes stingy			204	7	96'	0	99'		0	99'		speaker 1 15	2013/Jan/16	99%
20	mare produce Is it okay if I go ya thank you mes stingy			204	7	96'	0	99'		0	99'		speaker 1 15	2013/Jan/16	99%
21	mare produce Is it okay if I go ya thank you mes stingy			204	7	96'	0	99'		0	99'		speaker 1 15	2013/Jan/16	99%
22	mare produce Is it okay if I go ya thank you mes stingy			204	7	96'	0	99'		0	99'		speaker 1 15	2013/Jan/16	99%
23	hi wealth good yeah thank you for asking me was it eh a			9	0	11'	0	11'		0	11'		speaker 4 54	2013/Jan/16	10%
24	hi wealth good yeah thank you for asking me was it eh a			9	0	11'	0	11'		0	11'		speaker 4 54	2013/Jan/16	10%
25	hi wealth good yeah thank you for asking me was it eh a			9	0	11'	0	11'		0	11'		speaker 4 54	2013/Jan/16	10%
26	hi wealth good yeah thank you for asking me was it eh a			9	0	11'	0	11'		0	11'		speaker 4 54	2013/Jan/16	10%

As with many of the high frequency occurrences in the corpus, we can see from the concordance list that the majority of occurrences of *thank you* occurred during classroom tasks. Two roleplays conducted in class yielded numerous of occurrences of *thank you* in the corpus. Those roleplays were based on the topic of purchasing a ticket at the bus station and ordering a meal at a restaurant and resulted in the majority of examples of *thank you* in the ACE. On one hand, it could be argued that such tasks inflated the use of *thank you* but ACE spoken data represents a normal profile of English language teaching activities and so it is seen as representative of what one would typically expect in an English language classroom.

7.4.6 *Blonde hair*

The adjective noun *blonde hair* is cited by many in the literature as an example to explain collocation (McCarthy and O'Dell: 2005 and O'Keeffe, McCarthy and Carter: 2007). It occurs in the ACE corpus at both levels with an overall frequency of 62 occurrences PMWs compared with 0.2 occurrences in the BNC. Here again we see immense disparity between the ACE and BNC results, yet *blonde hair* is a feature of A2 level language according to the EVP. This again points to

it being an expected pattern in learner language. The following extract is taken from a pair work interaction between two students of the A2 cohort's first class:

Extract 7.4 (Student <$7> and <$4> A2 spoken corpus)

Student <$7>: am describe your wife

Student <$4>: she has long **blonde hair** and is thin yeah thin.

As this extract illustrates, being able to describe people is a typical A level competency in ELT syllabi (see Appendices 1 and 2).

7.4.7 *Spend time*

The idiomatic verb followed by noun collocation *spend time* occurs 23 times PMW at C1 level in the ACE corpus. The collocation appears 4 times in the BNC and is placed at B2 level in the EVP. The following extract was a response by a C1 student to the question what did you do at the weekend:

Extract 7.5 (Student <$24> C1 spoken corpus)

T: so, tell me, what did you do at the weekend?

<$24>: well, I work many hours, all week so at the weekend I relax and **spend time** with my wife and daughter. I think it's important to **spend such time** like that.

Again here we see a typical ELT classroom exchange where a teacher or a fellow student asks about a student's weekend.

Having looked at a range of types of individual items above from the spoken learner data, we now focus briefly on an example that illustrates the clustering of collocations. This example resulted from a picture prompt task. The extract is from Student <$7> (C1 cohort). As we will discuss, it also illustrates some degree of task effect.

Extract 7.6 (Student <$7> C1 spoken corpus)

…And in the middle of the way they ah see Maria with her friend Larie and they **start to talk.** They are **talking about** ah how all the things am are **more expensive** than **last year**. And they **decide** not **to shop**. But they **don't mind** they **talking about** the things that they shout. He <$=>shout shout shout</$=> and finally eh my uncle look at him to **beg him** to am **shut up** but he look at the mother and told her "mam mam look there is" ah I have to look at the name. "There is ah a **flying saucer** in the other **side walk**". And they don't they **laugh at** him they don't look at the other side because some missing ya. But when they looked at the other side there were am was a very big and round and bright flying saucer. Ya and they emm they were amazing and **talk about it**. Everybody was looking at the at the **same place**. Ya. And they were am **waiting for something strange happen**. Ya. And **this moment** am a little man with am a very big head came out.

In this exchange, Student <$7> used a number of collocations including the phrasal verb *talk about*. Interestingly, some of these occurrences only occur in this extract and therefore have a frequency of one count. The adverb and noun combination *last year* uttered by Student <$7>, occurs at both A2 and C1 level in the ACE corpus with an overall occurrence of 20 normalised results whereas it occurs in the BNC 0.66 times PMWs, yet, according to the EVP *last year* is a feature at A1 level. Due to the nature of this activity and a number of students being tasked with the same picture, opportunity of use and classroom task effect came into play, for example, the occurrences of adjective noun combination *flying saucer* which occurs in the ACE 70 times PMWs and only 0.03 times PMWs in the BNC. *Flying saucer* does not appear in the EVP. Another example is Student <$7>'s use of the phrasal verb *laugh at.* This is the only occurrence of this pattern the corpus (hence, 7 PMW) but when compared with a frequency of 1.01 PMWs in the BNC, we can see that it is over-represented compared to the norm of all spoken language. Despite this, *laugh at* also appears at A2 level in the EVP so again we can say that though it is a relatively low frequency pattern in English, it is expected of learners from Elementary level onwards.

7.5 Collocations in the written component of ACE

Next, this chapter will turn to the written dataset and focus on the most frequent collocations written by learners at both levels. Firstly, Table 7.3 below highlights the twenty most frequent collocations found in the written dataset. Their PMW frequencies are compared with the written component of the BNC and the EVP level where they are expected in learner use is also listed (where available).

Table 7.5 The 20 most frequent collocations in the ACE written corpus

Collocation	A2	C1	ACE Frequency (PMW)	BNC Frequency (PMW)	EVP Level
at the moment	√	√	364	26	A2
write a letter	√	√	341	2	A1
beautiful country	√	√	329	0.4	A1
a few months/days ago		√	311	2	A2
listen to music	√	√	235	1	A1
spend money	√	√	235	2	A2
wear clothes	√		235	0.6	A1
ideal job		√	223	0.04	B2
spend time		√	211	4	A2

day off	√	√	176	4	A2
rain a lot	√	√	164	0.02	-
friendly people	√	√	141	0.22	A2
food lovers		√	110	0.04	B1
soaked wet		√	92	0.5	B2
all the time		√	76	32	A2

With the exception of *wear clothes*, we can see that C1 learners use all of the items in the top 20 most frequent collocations. At A2 level, 9 of the top 20 forms are used. All of these A2 uses are expected by this level in the EVP. In the C1 cohort, all of the collocations are expected to be used by C1 learners and in fact, many would have been acquired by A1 or A2 level. The patterns *ideal job, food lover* and *soaked wet* are the only patterns that are used by C1 learners that are expected later than A2 level and these are normally acquired at B1 and B2 (see Table 7.5).

Also of note from Table 7.5 is that the profile of the collocations that are most frequent in the ACE data is different to the BNC. The most frequent collocation, *at the moment,* occurs 364 times PMW in the ACE written data and 26 times PMWs in the BNC. *Spend time* and *day off* are the next most frequent of the ACE top 20 items in the BNC, with 4 occurrences PMWs compared with 211 and 176 in the ACE written data respectively. Here again we see that the BNC results are far lower across all items and in many cases we see that collocations are used over 2 and 3 hundred times more by learners PMW (e.g. *write a letter, beautiful country, a few days/months ago, listen to music,* and so on). However, as noted above and in Table 7.5, all of these items are expected of EFL learners and most are expected to be acquired by A2. It points again to the language that relates to the classroom and to the syllabus rather than the language typical of all of written English, as represented in the BNC. As already outlined, the written sub corpus of the ACE is made up of

essays written both in class and as homework and assignments and therefore they are typical of classroom writing. Here we will take a qualitative look at some of the individual items listed in Table 7.3.

7.5.1 *At the moment*

The prepositional phrase and collocation *at the moment* is the most frequent collocation in the written corpus. It occurs at both levels with an overall occurrence of 364 (PMW). A search of the BNC yielded 26 (PMW) occurrences while the EVP stipulated *at the moment* as occurring at A2 level. The C1 group were tasked with writing an essay for homework on the topic of the importance of technology taken from *Cambridge Advanced Masterclass* (2010). Student <$21> wrote:

Extract 7.7 (student <$21> C1 cohort written corpus)

It's great for me, for example, **at the moment** it's really handy to skype or just text my family easily at home in Poland. It only takes one minute.

7.5.2 *Write a letter*

As this verb followed by noun collocation is an example of a delexical verb I chose to include in both Chapter 6 and here in Chapter 8. *Write a letter* occurs at both A2 and C1 level of the ACE with an overall occurrence of 364 PMWs compared to the BNC's 2 occurrences. Furthermore, *write a letter* occurs at A1 level of the EVP so it is not surprising that the lower level A2 cohort of the ACE use this collocation even though it is a low frequency pattern in English as a whole as represented by the BNC. The A2 cohort had to write a letter home and tell their family about their experience in Ireland, thus far. Student <$4> wrote the following and it illustrates how this typical classroom task generates this pattern in the learner data:

Extract 7.7 (Student <$4> A2 written corpus)

…Hi Carolina, for English class I need **write letter** to you and tell you how my is my life here in the Ireland.

Moreover, the C1 cohort for homework were tasked with writing an essay on the topic of communication. Student <$24> wrote the following:

Extract 7.8 (Student <$24> C1 written corpus)

... in the past, to talk to someone far away, you must **write a letter**. This way was very time consuming as it often took many many days for the letter to arrive at its destination.

In naturally occurring language, a letter writer is not likely to refer, metadiscoursally, to the task of writing the letter (as in Extract 7.7) or to explicitly demand of their addressee that they "must write a letter". Here we can say that there is both task effect and language display. However, this is typical of learner writing and this pattern is expected within the EVP at A1. (We do acknowledge that nowadays the pattern is less likely given that letters are no longer a frequent form of written communication).

7.5.3 *A few days/months ago*

The collocation *a few* occurred in various combinations in the ACE corpus. The majority of occurrences contain the adverb of time *ago* but it also occurs as a vague number. It occurs 311 (PMW) times in the ACE written corpus compared to 2 (PMW) occurrences in the BNC, it also appears at A2 level in the EVP but interestingly it does not occur in the A2 writing sub-corpus of ACE.

7.5.4 *Listen to music, spend money, wear clothes*

Listen to music occurs at both levels in the written corpus, with an overall frequency of 235 per million words which is compared to 1 occurrence PMW in the BNC. *Listen to music* is cited as A1 level in the EVP so it is no surprise that it occurs at both levels in the ACE but if we had based our expectations solely on the BNC written data, we would not have predicted it. Of course, being able

to talk about one's hobbies is a typical classroom task and it is also a typical A1 level competency (see Appendices 1 and 2). This use in Extract 7.8 illustrates this from the A2 cohort:

Extract 7.8 (Student <$4> A2 written corpus)

…I **listen to music** every day

The transparent verb followed by noun collocation *spend money* also occurs in both cohorts of the ACE and also has a frequency of 235 per million words while it occurs at a frequency of 2 (PMW) in the BNC. It is expected that learners can use this pattern by A2 level in the EVP. Extract 7.9 illustrates a typical use by learners:

Extract 7.9 (Student <$1> C1 written corpus)

We **spent** so so much **money** it was a fairly expensive holiday to be honest.

Finally, the verb followed by noun collocation *wear clothes* appears at A2 level of the ACE with 235 occurrences compared to 0.6 in the BNC. Despite this disparity, it appears at A1 level use in the EVP. Extract 7.10 shows its use by a C1 level learner. Here we see the verb in the past form and the use of a modifier before *clothes*:

Extract 7.10 (Student <$24> C1 written corpus)

She was a really enthusiastic person who had her own style and **wore** extravagant **clothes** all the time.

This is an interesting example because it shows the use of a collocation by a C1 learner that they typically would have acquire by A2 level but here they use it in a more sophisticated way by adding a modifier. The adjective *extravagant* is not expected until C2 level in the EVP.

7.5.5 *Friendly people*

The adjective followed by noun collocation *friendly people* occurs in both cohorts of the written corpus. This is attributed once again to the classroom task effect as both groups were tasked with comparing and contrasting Ireland with their home country, both groups had to write an essay on the topic. *Friendly people* has a frequency of 141 per million words in the corpus when compared to 0.22 occurrences in the BNC and it appears at A2 level in the EVP. Student 10 from the A2 group wrote the following:

Extract 7.13 (Student <$10> A2 spoken corpus)

…It's a beautiful country there have very **friendly people** here unlike Poland.

Soaked wet

The gradable adjective followed by adjective *soaked wet* appears 92 times per million words in the ACE corpus. It appears in the C1 written corpus and only appears with 0.5 occurrences in the BNC. *Soaked wet* is included in the B2 level of the EVP. Student 22 of the C1 ACE cohort wrote the following in a letter to home

Extract 7.14 (Student <$22> C1 written corpus)

…There is no place to avoid getting wet that is unless you stay home in a few minutes you can be **soaked wet**…I never found love in Spain but I found it on the chilly and soaked island of Ireland**.**

Thus far, we have looked at some of the most frequent collocations and compared their use in the learner data with the BNC and generally we have found that they are used substantially more by the learners in their writing. We offset this finding, however, by comparing their expected level in the EVP and we note that all of these items are seen as part of the typical learners' repertoire and most of the items will have been acquired by A2. It leads us to say that learner language, resulting from typical classroom writing tasks and homework reflects the target language syllabus so the EVP

seems to be a better point of comparison here. Let us now look at some individual learner case studies.

7.6 Case studies

By looking at some case studies of individual learners, we will give a more contextualised sense of the use of collocation. In some cases, we see clustered uses and we also see that the task can have an effect in the promotion of the use of collocation patterns. We also see that some learners can use collocates with specific discourse functions, such as discourse and vagueness marking.

Student <$1>

Student <$1> from Poland was a student in the advanced C1 cohort. He uttered 46 (PMW) collocational expressions in his speech within the classroom. He used some vague markers such as the preposition followed by noun: *or something,* at the end of the question:

Extract 7.15 (Student <$1> C1 spoken corpus)
will there be keywords **or something**?

This expression is used mostly in questions where *something* represents anything that is possible. This vague marker occurs, habitually, in spoken language (Carter et al, 2011). This speaker also used the adverb followed by adjective *kind of interesting* and discourse marker *I guess*. He also uses the expression *as well* (preposition followed by adverb) meaning *also*. He used the noun *time* proceeded by two frequently occurring nouns and time of year: *Christmas time* and *Summer time.* Clearly, the node *time* is proceeded by these collocates on a frequent basis in native spoken interactions. We can surmise that through frequent encounters with this pattern [period] + *time*, this learner has intuited the pattern semantically as well as syntactically. Here again we can make

connect to the usage-based model of acquisition (see Chapter 1). This learning is likely to have happened sub-consciously through experience of language through interactions.

This speaker <$1> also used the vague marker *those kinds of things* on one occasion and the more figurative collocation *from time to time*. This student correctly used the preposition followed by noun colligational structure when describing a famous Polish soccer player by stating he *is on steroids*. Again, the expression *steroids* is normally proceeded by the preposition *on*. He also used the structure of adjective followed by noun in the expressions *booking reference* and *important information*. *Important information* also occurs as a combination 0.61 times (PMW) in the BNC.

As previously mentioned, collocations are a language phenomenon that tend to occur in speakers at band 6 and above in IELTS and B2 in the CEFR. It is interesting to note that learners from the lower level A2 cohort of the ACE corpus, actually do use collocations correctly, as the following case study illustrates:

Student <$4>

Student <$4>, is at A2 level and is from Italy. She used the structure of verb followed by noun in the expression *admire people*. This speaker also uses the vague marker, *a bit* in the expression: *my lifestyle is a bit boring*. Student <$4> also uses the structure adjective followed by noun in the expressions *fast food* and *ancient castle*. She uses two time expressions that are regular and frequently occurring terminologies in English speech: *the last time* and the prepositional phrase *for a long time*. Bearing in mind the level of proficiency of this learner, it is interesting that she uttered 25 collocations (PMW) across the corpus recordings of this research.

Student <$11>

Student <$11> is a Polish female student who was part of the lower level A2 cohort. She uttered the expression *a drop of water.* This pattern is not expected until B1 in the EVP. This quantifier

was used by her in the correct manner and again she combined the correct quantifier with the correct object.

She used the delexical verb structure to *take photographs*, which is expected by A1 in the EVP. However, she was able to use it with the quantifier *many*. We also note that Student 11 responded to the teacher's greeting of *how are you* with the expression *I'm fine*. Again, a colloquial pleasantry that is used on a regular basis in spoken English. During the same class, she uttered the phrases *I didn't ring I wrote a letter*. She used the collocation verb with the exact noun to *write a letter.* When describing her experience of communicating in English, Student 11 uttered the following sentence:

Extract 7.16 (Student <$11> spoken corpus)

…because I want to **talk with people** I'm talking okay with people but I am not talking any good with people I'm *doing mistakes* and everything and I have to study next year so…

Here she correctly uses the verb followed by preposition followed by noun on three occasions: *talk with people*. However, she uses the incorrect delexical verb and thus fails to produce the correct collocation: *doing mistakes* instead of *make a mistake*. She also uses the vagueness marker *and everything*. In conveying surprise in a conversation with another learner she uttered the phrase *oh my gosh* an expression not very common in Ireland (only 7 PMW occurrences in the LCIE) but common in American English. When asked about where she acquired such a phrase, she attributed it to an American TV series and that she loves said expression. She further uses the vague marker a *little bit* twice.

Student <$23>

Student <$23> is a C1 learner from the Democratic Republic of Congo. He had been living in Ireland for two years prior to commencement of this course. He has very high receptive skills and

is quite a fluent and natural communicator but he still makes grammatical errors. When describing his life, he produced the following:

Extract 7.17 (Student <$23> C1 written corpus)

I am here nearly two years now, the people are different they have different **point of view** and the… attitude they living is very very poor. So my wife she came first and so she ah she **applied for** the refugee that was in two thousand and nine we join her okay my wife who travelled first yeah some problem she got the eh refugee <$/x> statue/status</$x>

We see his use of *point of view.* This is expected of learners by B2 and occurs 23 times PMW in the written BNC. He correctly used the expressions *apply for* and *refugee status*. The verb + preposition *apply for* is expected in the EVP by B1 level while *refugee status* is not in the EVP. This points to the gap between EFL and ESL. For so many ESL learners, 'apply for refugee status' is in itself a fixed unit that carries meaning for their lives and yet it is never part of an EFL syllabus as reflected in the EVP.

During a roleplay with a fellow classmate, student <$23> replied to a request with the response token, *of course,* three times, an expression that Sinclair (1991) used to explain idiomaticity where both units have no meaning alone but have only a unitary meaning, as discussed above. He further used the collocation to *tell a story* again he has chosen the correct verb and noun combination. So while he has much inaccuracy in his writing and in his speech, he shows fluency in terms of both communicative flow and also in terms of his ability to use collocations.

7.7 Conclusion

This chapter began by first defining collocations and identifying different collocation types. Next, the chapter surveyed the previous literature published on the topic prior to syntactically describing collocations. Following that, the chapter focused on the semantics and pragmatic functions of collocations. The chapter then identified collocations as a common feature of the native English

speaker before identifying collocations as a major problem for the language learner. Finally, the chapter analysed the different collocations found across the four datasets of the ACE corpus and compared these with two reference points 1) the BNC, and 2) the EVP.

The results suggest that collocations are acquired and used early in terms of level and become part of the learners vocabulary repertoire as units. This is confirmed when we compared items with the EVP. We generally found that the collocates in the learner data were mostly acquired by A2 level. However, when a teacher looks for guidance on teaching collocations, there is inconsistency because we find that the CEFR benchmarks collocation as a feature associated with C1 learners. IELTS places collocation as a feature of a Band 6 and above learner in the IELTS examination. Clearly this is not what we found in our analysis of the ACE data. The analysis in this chapter clearly shows that the ESL learners in ACE do indeed use various English collocations in their speaking and writing at both levels and it validates the usefulness of the EVP as part of a move to offer a profile of English language use.

Research tells us, as previously outlined in *Chapter 4 Conceptual Framework*, that accurate and productive use of collocations is a natural and available skill to the native speaker and is also an indication of advanced level proficiency in a second language speaker. However, our results show that A2 learners are widely using collocations. Also, when we look at these items in the EVP, we see that they are mostly listed as A1 and A2 level vocabulary items. The EVP has proved a useful benchmark for the ACE data here and it seems that it has a more realistic expectation of learners' use of collocation from Beginners' level upwards. Because the CEFR is a framework for languages in general, it seems to be less useful in relation to collocation in English language learning.

When we compared our ACE results with the BNC, we found major disparity in terms of PMW frequencies. It was not uncommon to see learners in the ACE data using collocations over 200 times PMWs more frequently than in the BNC. However, we noted on a number of occasions that this is not unexpected given that the BNC represents English as a whole, across genres, registers and

regions etc. Classroom ELT use of language is driven by syllabi that promote communication and in the context of ACE, the ESL learner is following a syllabus that will aid their communication in every day live. We note that this is reflected in some of the high-frequency patterns discussed in this chapter such as response tokens and vagueness markers, conversational routines as well as classroom tasks. In summary, the collocations in ACE were found to serve numerous pragmatic functions. These include markers of politeness and these are crucial to in-groupness within an ESL learners' context. Furthermore, the researcher identified collocations as embedded in the culture and context in which it is produced. The learners' immersion in the culture of Ireland equipped them with native speaker collocations and expressions related to the culture in which they are currently living. In conclusion, this chapter, using the ACE corpus, strongly indicates that despite most literature and syllabi both the A2 and C1 use ample collocations in their speech and writing and that the collates that they used are generally reflective of 1) what is expected in the EVP, 2) what is expected in ESL classroom tasks and conversations, and 3) what learners frequently hear in their ESL language experience(s).

Linking back to Chapter 1, we note again that the learners in the ACE data experience language in their daily lives, their workplaces and their ESL classroom. They experience language through interactions with native and non-native speakers. The patterns that they use and re-use evolve with their interactions and this again ties in with the usage-based model of language acquisition (Tomasello, 2003; Ellis & Larsen-Freeman, 2006; Ortega, 2013). We next turn to the final analysis chapter and this will look at idioms in the ACE data.

Chapter 8 Idiomatic Expressions

Captain Kirk: If we play our cards right, we may be able to find out when those whales are being released
Dr. Spock: How will playing cards help?
(Star Trek IV The Voyage Home: Nimoy, 1986)

8.0 Introduction

This chapter looks at idioms used by the English language learners in the two cohorts of ESL learners at the college, in both their classroom speaking and writing. A key objective is to assess similarities and or differences with native speaker use in terms of form and function since idiomatic expressions are used in everyday situations and exchanges by the native speaker of English. In native speaker discourse their occurrence frequently contributes to a figurative, yet natural mode of conversation. Research has shown that idioms are culturally bound and linked to the context of the situation in which they arise (Carter, 1987; Halliday 1987; Sinclair, 1991; McCarthy, 1998; Palma Fahey, 2005). Furthermore, the recognition and use of such phrases by native speakers often indicates a membership of a particular cultural group (see McCarthy, 1998; Palma Fahey, 2005) as the meaning of an idiom is embedded in the culture in which it is expressed. Moon (1998) discusses how, without knowledge of the culture, it is almost impossible to understand, and accurately interpret an idiom: 'Fixed expressions and idioms may be localized within certain sections of a language community, and are peculiar to certain varieties or domains' (Moon, 1998: 7).

As detailed in chapter 4 and in Appendix 3, the students participating in the present study have been living in Ireland for an extended period of time. In accordance with Moon (1998) their use of idioms could be seen as indicators of linguistic assimilation in the culture of Ireland. Therefore, it was felt that it would be of interest to analyse idiomatic usage in the ACE corpus. Also, it was anticipated that colloquial, local language may be used as the learners may have picked up expressions from daily interactions.

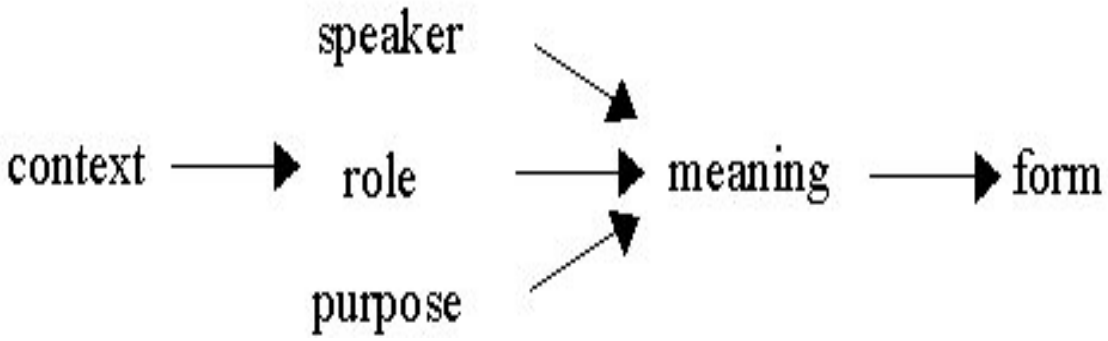
In this chapter, we will consider the form, usage, and function of the expressed idioms. We will compare and contrast the use of such expressions by students from the two different levels of competency (as detailed in Chapter 4 Methodology). What is of real interest in this chapter is to

ascertain whether the lower (A2 level) cohort use idioms in their language. Of interest to note is that in terms of the CEFR, the C1 Advanced level descriptor states: "*I can understand a wide range of idiomatic expressions and colloquialisms, appreciating shifts in style and register"* (*Can do* statements of the Council of Europe 2001). It is at this level in the CEFR that idioms first appear, which implies that levels below C1 are not expected to use idioms. Moreover, idioms are banded at a 6 (B2 level) and above in the International English Language Testing Systems (IELTS) examination. As a result, it should be interesting to investigate to what extent the lower cohort, (A2) learners, use such expressions. Reflection on the opportunity of use in relation to the college syllabus will also be examined.

8.1 An anecdote from the ACE corpus

At seminars at the University of London in the 1930's, Bronislaw Malinowksi postulated the importance of the context and the culture in which language is spoken to accurate understanding. According to Malinowski, the following is the structure of a piece of communication:

Fig 8.1 Structure of a communication according to Malinowski



In Fig. 8.1, we can see that a piece of communication begins with the context or culture in which the conversation is taking place. The next stage is the speaker himself and his role within the society and simultaneously the purpose of his communication. The next consideration is what the speaker intends

to mean by his expression and finally the form (grammatical and/or lexical). Unlike earlier ideologies, the actual form of the expression is not as important as the different components which precede it. The entire interpretation, according to the model above, depends on and begins with an understanding of the culture and context in which it exists and furthermore the role of the communicator within that culture is of paramount importance. Pawley and Syder (1983: 24) agree with Malinowski's assertion when they state that idioms are 'culturally salient', while Carter and McCarthy (1995) view the ability to ellipt and contract idioms as a 'cultural commonality' between speakers.

To demonstrate this, an introductory activity on the topic of idiomaticity was carried out. The students at the college were asked to try to translate into English some well-known idioms from their mother tongue. Without prior knowledge of the culture in which they occurred; it was extremely difficult to interpret the expressions uttered.

- ***When a frog grows hair,*** uttered by a C1 Spanish student. It is roughly translated into English as something that will never happen, similar to the expression: *when pigs fly*.

- ***You have tomatoes in your eyes,*** uttered by an A2 German student. It is roughly translated as you are not seeing the situation clearly similar to: *you are wearing rose tinted spectacles.*

- ***He was a bronze man,*** uttered by an A2 Italian student. Meaning that he was very affluent.

- ***No fighting in the church,*** articulated by a C1 student from the Democratic Republic of Congo. Could be translated as you don't argue with your own family roughly similar to: *biting the hand that feeds you.*

- ***It's raining pen knives,*** contributed by an A2 Brazilian student. Is roughly translated as it is raining heavily, similar to the English *raining cats and dogs.* .

- ***Shit for you,*** said by a C1 Polish student. Meaning the best of luck for you similar to *break a leg.*

- ***The carrots are cooked,*** contributed by a C1 French student. Is translated as not being able to change the past; similar to *don't cry over spilt milk.*

Albeit anecdotal in terms of evidence, this clearly indicates the culturally-situated nature of many fixed expressions and idiomatic phrases. We will now look more closely at the terminology and definitions associated with this area.

8.2 Fixed Expressions and Identifying an Idiomatic Phrase

The term 'Fixed Expression' was adopted from Alexander (1978 and 1979) and further developed by Carter (1987). It subsumed numerous types of language phenomena such as multi-word lexical items, phraseological units, phrasal lexemes, proverbs, and certainly idiomatic expressions. Moon (1998) highlighted how this term was unsatisfactory and arguably problematic as it is difficult to identify what is a fixed expression. Numerous publications have termed the language unit under a different title, for example, Gläser (1984), Cermak (1988), Nunberg et al (1994), Barkama (1996). Yet, despite the lack of agreement on a suitable term, each seems to describe a similar language component: 'a complex set of features that interact in various, often untidy, ways and represent a broad continuum between non-compositional and compositional groups of words' (Moon, 1998: 6). Wray (2009) claims that a consensus is 'severely limited' (Wray, 2009: 9). Many linguists have different ideas as to what can be included under the term 'fixed expression'. Corder (1973) referred to this language phenomenon as *Holophrases*, Hakuta (1974) termed them *Prefabricated Patterns*, Keller (1979) refers to them as *Gambits,* Peters (1983) identifies them as *Speech Formulae*, Pawley and Syder (1983) highlight *Lexical Stems*, Nattinger and Decarrio (1992) titles them *Lexical Phrases*, Lewis (1993) gives us *Multi-Worded Items* and in 1997 Lewis categorizes them as *Lexical Items* and *Chunks*. Finally, Williams (1998) coins the term *Prefabricated Chunks*. With such a vast number of varying types of fixed expressions, it is undoubtedly difficult to explain what exactly a fixed expression is. For the purpose of this study, fixed expressions include the following:

1. Phrasal Verbs (see Chapter 5 of this thesis)

2. Delexical Verbs and their complements (see Chapter 6 of this thesis)

3. Collocations (see Chapter 7 of this thesis)

4. Social Routines

5. Idioms and Metaphors (addressed here in Chapter 8)

What seems to have caused the most controversy is the fact that there is no exact clear difference between any of the above-mentioned language units, ergo it is difficult to identify an idiomatic expression. Bolinger (1977: 168) identifies a continuum between each type of fixed expression without any clear distinction and the New Webster's Dictionary (1993) gives the following definitions of "idiom" (emphasis in bold added):

1. The language peculiar to a people, country, class, **community**, or more rarely, an individual;

2. A construction or expression having a meaning different from the literal one or not according to the usual patterns of the language

(The New Webster's Dictionary (1993)

For the purpose of this study, the definition of idiom will be taken from McCarthy and O'Dell (2002: 6) where they state: 'idioms are expressions which have a meaning that is not obvious from the individual words…the best way to understand an idiom is to see it in context'.

8.3 Characteristics of Idioms

Simpson and Mendis (2003) see idiomaticity as a 'register specific linguistic factor' (Simpson and Mendis, 2003: 420) where it could be surmised that the longer the stretch of speech the more likely it is that an idiom will be encountered. The possibility of reoccurrence of an idiom in a stretch of speech is high, while at the same time unpredictable.

Figurative language and idioms have received considerable attention through the years (Fraser, 1970; Carter, 1987; Sinclair, 1991; Moon, 1997; McCarthy, 1998; O'Keeffe, et al 2007). Moon (1997: 43) claims that idioms are a 'form of multi-word unit', that is an item comprising two or more words where the meaning is not always interpretable from each individual component. In order to interpret the phrase, one must concentrate on the concatenated pattern and decipher the figurative or literal meaning. Stein and Su (1988: 444) use the expression *hit the roof* to illustrate idiomaticity. Here the meaning of *hit the roof* cannot be derived from the meanings of *hit* and *roof* as idiomatic expressions tend to contain at least one figurative element that is not easily predictable. For a native speaker who uses such expressions on a daily basis, it is clear that the literal meaning of physically *hitting* or *thumping* the *roof* is not the intended meaning, but, rather than a situation has caused immense anger. Grant and Bauer (2004) claim that a large number of figurative expressions can be dissected or "unpicked" to work out the meaning and these expressions are referred to as **decomposable** or **transparent** idioms. The opposite, **opaque** idioms, have little or no relation to the literal meaning. We will discuss the notion of transparency and opacity later in this chapter. Idiomatic expressions can, therefore, be characterized under the following descriptors:

1. Meaning is not completely derivable from the sum of their parts
2. They are largely rigid or fixed in structure
3. The literal meaning of the phrase tends to be less frequent but not impossible.

Furthermore, Moon also highlights three key features of a multi-word item which characterize its degree of idiomaticity (Moon, 1997):

1. **Institutionalization:** the degree to which such an expression is conventionalized and accepted in a language. It encompasses the frequency in which the string of language recurs.

2. **Fixedness**: the degree to which a multi-word unit is frozen and does not contain variables and often contains restrictions on aspect, mood and voice.

3. **Non-compositionality:** the degree to which a multi-word unit has a holistic meaning and cannot be interpreted through knowledge of each individual component but rather has a meaning in unison.

The position we take in this chapter goes back to the original meaning of 'idiomatic' as 'something peculiar to an individual language', not necessarily anything to do with semantic opacity. So the fact that English says '*I'll be there about half three'* and Swedish says '*I'll be there about half four*' for exactly the same time on the clock is a question of idiomaticity. In this chapter, we consider examples like 'fish out of water' and 'from time to time' as instances of idiomatic language though some might argue that they conflate 'different senses' of the term and should be viewed separately. Here, we hold that such a distinction is purely a construct of different linguists and that the two aspects of idiomaticity it displays (structural restriction and peculiarity of expression belonging to an individual language) are part of a single idiomatic identity of the expression. Therefore, while we refer to transparency and opacity in this chapter, we do not limit our definition of idiomaticity on the basis of semantic opacity.

8.4 How many idioms are there in English?

It is claimed that the number of idioms in English is 'uncountable' (see Quora, 2016). Be that as it may, if you google the number of idioms in the language you get a list of 3,838 expressions, while www.grammarnet.com (2019) estimates that there are at least 25,000 idiomatic expressions in British English, and approximately 80,000 in American English. To account for this unaccountability, there are many estimations as to the number of idioms in the language. Research, for example Erman and Warren (2000), estimate that around 50 percent of spoken language is composed of idioms while there are 5,000 idioms in the *Cambridge Dictionary of American Idioms* (Heacock, 2003). Table 9.1 below highlights the estimated occurrence of idiomatic expressions in English. Sorhus (1977) calculated, from her corpus of English Canadian Speech, that speakers used such items at a ratio of one in every five words. Howarth (1998) found that frequent verbs, in a social scientific/academic corpus, occurred in idiomatic expressions or restricted collocations in almost 40 percent of cases. Oppenheim (2000), while analyzing short speech, discovered that 66 percent of the language consisted of idiomatic strings. Erman and Warren (2000) calculated that 58 percent of the spoken language they investigated was formulaic. Finally, Rayson (2008) claimed that 15 percent of written text is formulaic in nature. Table 8.1 summarises the different percentages that some researchers have cited for idiom coverage in English.

Table 8.1 The percentage of Idioms in English (adapted from Schmitt, 2013).

Percentage Coverage	**Linguist**
52-58%	Erman and Warren (2000)
32%	Foster (2001)
48%-80%	Oppenheim (2000)
31%-40%	Howarth (1998)
15%	Rayson (2008)

According to Voxy (2019) the following are the ten most frequent full idioms in British English.

1) A piece of cake.
2) To cost an arm and a leg.
3) Break a leg.
4) Hit the books.
5) Let the cat out of the bag.
6) Hit the nail on the head.
7) When pigs fly.
8) You can't judge a book by its cover.
9) Bite off more than you can chew.
10) To scratch someone's back.

Looking at these expressions, it is clear that they could be used more often than the clichéd phrase of '*it's raining cats and dogs'* and also reinforces the notion that it is not so easy to realise if an idiom

is being used. Stengel (1939) reinforces this and the idea that they are very difficult to comprehend by stating:

> 'Idioms are largely responsible for specific features of language. Idiomatic speech is a kind of secret speech…[idioms] are riddles…they are the traps in a language…they are petrified jokes and their symbolism is very often incomprehensible…we feel the strange effect of idioms they force on us…pictorial thinking…while learning, we often suspect a latent original idea behind the word' (Stengel, 1939: 476-477).

8.5 Forms and Types of Idioms

As seen earlier, idioms are merely one example of fixed expressions, there are numerous in the English language. Table 8.2 identifies the types categorized by Carter (1987).

Table 8.2 Types of Idioms (Carter, 1987).

Type of Fixed Expression	**Example**
Idioms	
Irreversible Binomials	*spick and span; red tape*
Full Idioms	*to rain cats and dogs, to smell a rat*
Semi-Idioms	*dead drunk, the party kicks off at nine*
Proverbs	*a bird in the hand is worth two in the bush*
Stock Phrases	*when all is said and done*
Catchphrases	*that's another fine mess you got us into*
Allusions/Quotations	*to be or not to be that is the question*
Idiomatic Similes	*as daft as a brush; as drunk as a skunk*
Clichés	*long time no see; bottoms up*
Connectives	*to sum up; finally*

Conversational Gambits	*guess what!; I wondered if I could have a word*
Stylistic Formulae	*ladies and gentlemen; regarding my recent request*
Stereotypes	*it's not what you think!; I thought you'd never ask*

McCarthy (1998) and Palma Fahey (2005) add the following to this list:

1. **Clausal Idioms**: these are fixed expressions with the structure of verb +complement, e.g. *hit the roof, kick the bucket.*

2. **Idiomatic Phrasal Verbs:** these are lexical verbs followed by a particle, e.g. *take after somebody* (meaning to resemble a person), *get up, look after*

3. **Prepositional Idioms:** *over the moon, in two minds*.

4. **Binomials and Trinomials:** these are usually fixed and irreversible and normally contain a conjunction such as *and:* e.g. *to and fro, going and coming, hot and cold.*

5. **Frozen Similes:** a comparative expression containing the words *like* or *as: as happy as Larry, as sober as a judge, as old as the hills.*

6. **Possessive Phrases**: *the bee's knees.*

7. **Opaque Nominal Compounds***: the back of beyond, blackmail.*

8. **Cultural Allusions:** these include: slogans, proverbs, catch phrases, quotations: *my flexible friend, to be or not to be, the early bird catches the worm.*

9. **Gambits and Discourse Markers:** *how's it going? of course, well, like.*

The classifications of Carter (1987), McCarthy (1998) and Fahey Palmer (2005) will be used in the analysis below.

8.6 Data Analysis of the idiomatic expressions in ACE.

8.6.1 Summary of findings

Firstly, the 150,000 words of spoken and the 20,000 words of written data of the ACE corpus were searched manually for all figurative language and idioms based on the various examples identified above. Table 8.3 shows the breakdown in terms of percentages of learners using idiomatic expressions. In summary, all C1 learners, as we expected, use them in both speaking writing, while 66% and 77% of A2 learners use them in speaking and writing respectively.

Table 8.3 Number and percentage of students using idioms across both cohorts

	Speaking		**Writing**	
	Raw	**%**	**Raw**	**%**
A2 (9 participants)	6	66%	7	77%
C1 (15 participants)	15	100%	15	100%
Total	21		22	

When we look at the overall breakdown of the numbers of idiomatic expressions used in ACE, by level and by sub-corpus, we see that they occur with a very similar frequency in the spoken data but the A2 written data seems to show around double the number (PMWs) compared with C1 (we will discuss this further below):

Table 8.4 Breakdown of frequencies of idiomatic language in ACE speaking and writing, by level and percentage.

Level	Spoken frequency (raw)	Frequency (PMW)	%	Written frequency (raw)	Frequency (PMW)	%
A2	112	1750	42	82	11714	52
C1	152	1767	58	77	5923	48
Total	**264**			**159**		

Another observation is that A2 learners use 42% of all idiomatic items in speaking compared with 58% used by C1 learners. In writing, A2 learners use 52% of items compared with C1 learners who use 48% of all items in the sub-corpus. Across all of ACE, when we add all A2 idiomatic uses (194), they amount to 46% of all idiomatic expressions used compared with the total for C1 idiomatic use (229), which accounts for 54% of all uses. These results were not expected based on the CEFR can-do statements and other competency frameworks cited above, as we shall discuss further below.

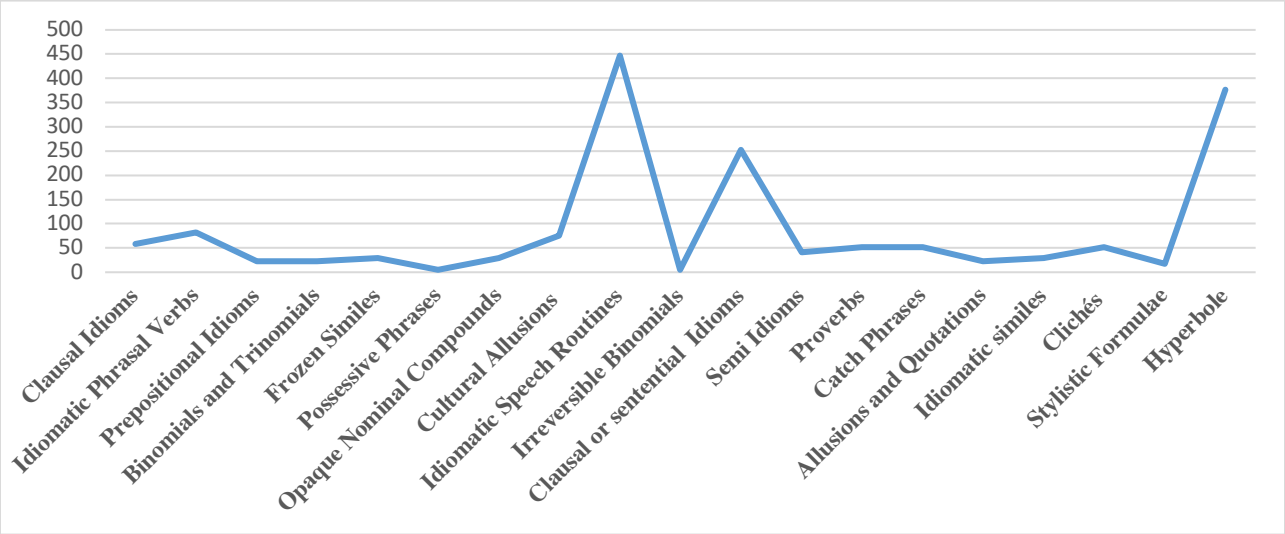
8.6.2 Overview of types of idioms used in ACE

Having manually identified all idiomatic expressions, the second step was to classify them in accordance with the categories of form as detailed above by Carter (1987) and McCarthy (1998) and also according to their function in the context of occurrence. Assessing the function of an idiom is according to McCarthy (1998) a difficult task, as the boundaries between each category of idiom are 'fuzzy' and unclear. Interpretation can therefore often prove subjective. At both stages of identification and classification of items, two independent researchers checked the results and any conflicting categorizations were resolved. Thirdly, the use and function of the expressions used were compared and contrasted across the two cohorts of learners: A2 and C1, in order to identify variations or similarities between all twenty-four participants. Finally, we determined which level used the most idiomatic expressions in the corpus.

After identifying how many idiomatic expressions were uttered in the corpus the next step was to

analyse, per category, the type of expressions uttered by the students of the corpus. Fig. 8.2 below illustrates the spread of various figurative expressions under the categories described by Carter (1987), McCarthy (1998) and Fahey Palmer (2005).

Fig 8.2 Types of idioms used in the corpus (Raw Frequency)



Here (in Fig. 8.2) it is clear that *Idiomatic Speech Routines* is the most frequently used form of figurative idiomatic language with 447 (PMW) occurrences. This is followed by *Hyperbole*, with 376 (PMW) occurrences. Table 8.5 below further breaks down the frequency results in Fig. 8.2, providing examples from ACE and showing whether items are found in spoken or written data, or both:

Table 8.5 Distribution of idiomatic expression by form, PMW frequency and sub-corpus

Type	Form	Example	Freq. (PMW)	Sub-Corpus	
				Spoken	Written
Clausal Idioms	Verb + Complement	*Pop the question, Hit rock bottom, Hit the roof, Shut your mouth*	58	√	√
Idiomatic Phrasal Verbs	Verb + Particle	*get over, throw up,*	82	√	√
Prepositional Idioms	Preposition + Complement	*Over the noise, in this case,*	23	√	
Binomials and Trinomials	A + conjunction + B	*back and forth, black and blue,*	23	√	√
Frozen Similes	Comparison using like or as	*Hungry like a wolf, as thick as two short planks, as thick as shit,*	29	√	√
Possessive Phrases	Noun + 's + head noun	*Cat's whiskers. Lion's share*	5	√	
Opaque Nominal Compounds	Noun+Noun	*Car crash, basket case*	29	√	√
Cultural Allusions	Slogans, Quotations, Catch Phrases, Proverbs	*how's it hanging, a fish out of water, once in a blue moon, a grand*	76	√	
Idiomatic Speech Routines	Gambits and discourse markers	*Sounds good, how's it going, what's the story?, how's she cutting?, how's it hanging*	447	√	
Irreversible Binomials	Structure that cannot be altered	*Ladies and gentlemen,*	5	√	
Clausal or sentential Idioms	Consist of full clause or extended sentence	*It's raining cats and dogs, are you pulling my leg, I can't get her out of my head, on the tip of my tongue, give me a hand, give*	252	√	√

		me five, once in a blue moon,			
Semi Idioms	Still has the same idiomatic meaning but the structure can be altered a little	*faced/had/encountered a problem, mind your (own) business,*	41	√	
Proverbs	A piece of language that has an extra message to the listener	*A bird in the hand is worth two in the bush, the grass is greener on the other side,*	52	√	√
Catch Phrases	Slogans that have become extra frequent in speech	*Break a leg, good luck!* *Not to worry!*	52	√	
Allusions and Quotations	A quotation that has since been institutionalized into the language	*to be or not to be, History repeats itself*	23	√	
Idiomatic similes	Comparisons using adverbs *like* or *as* but are not frozen	*Like a fish out of water, drink like a fish, sleep like a log, hungry like a wolf,*	29	√	√
Clichés		*just one life, At the end of the day, when all's said and done life,*	52	√	
Stylistic Formulae		*It gives me great pleasure … It has fallen upon me to …*	17	√	√
Hyperbole	Exaggeration and Understatement	*Dying for a drink, I could eat a horse, I'm knackered, can't get her out of my head, I'm bursting to pee*	376	√	√

From Table 8.5, we can observe that 85% of the all idiomatic expressions in ACE occur in the spoken sub-corpus while 55% of them occur in the written data. We can also break the results in Table 8.5 down further, by level as Table 8.6 illustrates;

Table 8.6 Types of idiomatic expression, by PMW frequency and level

Type	Form	Freq. PMW	Level	
			A2	C1
Clausal Idioms	Verb + Complement	58		√
Idiomatic Phrasal Verbs	Verb + Particle	82	√	√
Prepositional Idioms	Preposition + Complement	23	√	
Binomials and Trinomials	A+ conjunction +B	23		√
Frozen Similes	Comparison using *like* or *as*	29	√	√
Possessive Phrases	Noun +'s + head noun	5		√
Opaque Nominal Compounds	Noun + Noun	29	√	√
Cultural Allusions	Slogans, Quotations, Catch Phrases, Proverbs	76	√	√
Idiomatic Speech Routines	Gambits and discourse markers	447	√	√
Irreversible Binomials	Structure that cannot be altered	5	√	√
Clausal or sentential Idioms	Consist of full clause or extended sentence	252	√	√
Semi Idioms	Still has the same idiomatic meaning but the structure can be altered a little	41	√	√
Proverbs	A piece of language that has an extra message to the listener	52		√
Catch Phrases	Slogans that have become extra frequent in speech	52		√
Allusions and Quotations	A quotation that has since been institutionalized into the language	23		√
Idiomatic similes	Comparisons using adverbs like or as but are not frozen	29	√	√
Clichés		52	√	√

Stylistic Formulae		17		√
Hyperbole	Exaggeration and Understatement	376	√	√

From Table 8.6, we can see that of the 19 types of idiomatic forms, A2 learners use 12 of them (63%) while C1 learners use 18 of them (95%). Again these results are unexpected, especially in the A2 cohort. Let us now look at the results in greater details.

8.6.3 Comparing idiomatic expressions in ACE with EVP and BNC

This higher than expected use of idiomatic language in Table 8.6 can be brought more into perspective when we look at the top 20 idiomatic items in Table 8.7 and compare their frequency in the BNC (PMWs) and look at where they are expected (if at all) in the EVP.

Table 8.7 Top 20 idiomatic expressions in the ACE data, by level, compared with BNC and EVP.

Idiom	**ACE (PMW)**	**A2**	**C1**	**Spoken**	**Written**	**EVP Level**	**BNC PMW (raw)**
from time to time	82		√	√	√	B2	14.77 (1632)
how's it going	82	√	√	√		-	0.6 (14)
are you pulling my leg	71	√	√	√		-	0.34 (10)
Blackmail	71		√		√	C2	4.36 (21)
to face a problem	70	√	√	√	√	B2	0.5(3)
throw up	70	√		√		B2	5 (164)

to pass time	64		√	√	√	B1	1 (4)
give me five	59		√	√		-	0.2 (19)
on the tip of my tongue	58		√	√		-	0.2 (4)
give up	52	√	√	√	√	B1	34 (1695)
as thick as two short planks	47	√		√		-	0.01 (4)
get over	47	√		√	√	B2	6.85 (413)
on the one hand	47		√		√	B2	12.6 (1408)
on the other hand	47		√		√	B2	47 (5262)
a piece of cake	42	√		√		B2	0.2 (44)
look forward to	41	√	√	√	√	B2	21 (1000)
could eat a horse	35		√	√	√	-	0.02 (2)
give me a hand	35	√		√		-	1.4 (45)
go away	35	√	√	√	√	B1	0.6 (1244)
how's she cutting	35		√	√		-	0 (0)

Comparing the top 20 most frequently used idiomatic expressions with the EVP, Table 8.7 tells us that:

- In the A2 sub-corpus, 11 of the top 20 expressions were used (55%).
- In the C1 sub-corpus, 15 of the top 20 expressions were used (75%).
- The EVP only expects 12 of these idioms to be used by learners.
- The EVP expects the 12 items of the 20 items to be used from B1 level upwards. Therefore, all A2 uses exceed expectations of their level. All but one of the items used by C1 learners are expected to be acquired by B1 or B2 level (*blackmail* is a C2 level usage according to the EVP).

Comparing the top 20 most frequently used idiomatic expressions with the BNC (PMW) frequencies, Table 8.7 tells us that:

- Overall, the PMW frequencies of the expressions in Table 8.7 are many times higher in ACE than in the BNC. For example, *from time to time* is the most frequent item with 82 occurrences PMWs in ACE compared with 14.77 in the BNC. Expressions like *pulling [someone's] leg* occurs 71 times PMWs in ACE but only 0.34 times PMW in the BNC, and so on. Yet, as discussed above, many of these idioms are expected to be acquired by learners. The high ACE results suggest that the classroom brings more focus on and encounter with idiomatic expressions that is represented by the BNC in general English across all genres of speaking and writing.
- One of the items, *how's she cutting?*, used by both A2 and C1 learners, does not occur in the BNC. While it has no occurrence in LCIE, it is a common, informal greeting in Irish English and so it must have been acquired in Ireland by the ESL learners participating in this study.

8.6.4 Idiomatic expressions in ACE: a closer examination

Let us now take a closer qualitative look at idioms. First, we look at different idiom forms that have

occurred across both A2 and C1 cohorts (based on Table 8.6):

Idiomatic Speech Routines

In total, 447 (PMW) examples of Idiomatic Speech Routines, especially in the forms of discourse markers and gambits, were found. These account for 27% of all idiomatic expressions in ACE. Many of which were ligosemic, in that they are typical of the country of Ireland. Used as discourse markers and hedges, such expressions play a vital role in everyday conversation (see Carter and McCarthy, 1997). They frequently appear in native speaker interactions ergo, it is not surprising that they are found in the ESOL learning environment. Greetings such as *how's it going*? *What's the story*; *how's she cutting; how's it hanging?* Make up an integral part of the corpus due to the high-frequency of occurrence. *What do you do? Nice stuff; I don't mind*; *alright; you know*; and *at the moment,* examples of gambits, all function as a means of interaction and discourse. Their frequency in the ACE data again echoes the theoretical model of usage-based learning discussed in chapter 1 and elsewhere which suggests that individual language *experience*, for formal and informal triggers acquisition. The data here suggests that ESL students who live and interact in Ireland acquire many everyday speech routines regardless of their level. None of these items is expected to be acquired by A2 learners and yet they are used frequently in ACE.

Hyperbole

The second most frequent type of idiomatic expression is the figurative use of hyperbole, accounting for 23% of all expressions (376 PMWs). It was not until after the initial analysis of the corpus that the researcher decided to include hyperbole as a figurative language category. These include vernacular expressions such as:

I'm dying for a drink (A2);

I'm starving (A2, C1)

I'm exhausted (A2, C1);

I'm knackered (C1);

I'm bursting to pee (A2);

I could eat a horse (C1).

Their use highlights the exaggeration function of hyperbolic articulations. *I'm starving* (A2, C1) occurred 11 times (PMW) and *I could eat a horse* (C1) function to highlight how hungry the speaker is. *I'm exhausted* and *I'm knackered* (C1) exaggerate how tired the speaker feels. While *can't get her out of my head* (C1) and *I'm bursting to pee* (A2) imply that the speaker cannot stop thinking about this girl and the urgency and desperation of a person who needs to use the bathroom respectively. What we note here from the examples is that they communicate basic human needs and functions (feeling thirsty, hungry, tired, needing the bathroom). They are everyday informal uses of a language. However, most of them are never likely to appear in English Language Teaching materials. Learners have acquired them through interactions in Ireland. They have experienced these phrases and quickly acquired them (as evidence by the fact that A2 learners use them frequently). As such, we can speculate that they have become a core part of their repertoire. Again this links with a model of language acquisition where frequently experienced and used forms and patterns become *entrenched* chunks in the mind of the language learner (Ellis et al. 2015). *It's driving me nuts* is a Subject + Verb + Object + Object string, as Ellis et al. (2015) note but the user of this has abstracted it fully as one semantic and syntactic unit.

Clausal Idioms or Sentential Idioms

Clausal or sentential idiom account for 15% of all idiomatic expressions. This category tends to also include variable flexibility. The structure of verb followed by complement can be seen in the expressions *pop the question* (C1) meaning to propose, *to hit rock bottom* (A2) meaning to struggle and be in the worst place of all time*, hit the roof* (C1) interpreted as to become extremely angry and

shut your mouth (C1), an aggressive and informal way of telling somebody *to be quiet* and *stop talking*.

Idiomatic Phrasal Verbs

Dealt with in detail in Chapter 6, Idiomatic Phrasal Verbs account for 5% of the idiomatic expressions uttered in this corpus. In English, there are numerous expressions that combine a verb with a particle with the meaning being literal or idiomatic. Such expressions have a high-frequency in everyday spoken English. As shown above (Table 8.7), the idiomatic phrasal verb *get over* occurs at both A2 and C1 level with an occurrence of 47 (PMW) in the ACE corpus, while it occurs 6.85 times PMW in the BNC and it appears at B2 level in the EVP. *Throw up* meaning to vomit, occurs 70 times (PMW) in the C1 cohort of the ACE; it is an example of B2 level according to the EVP and has a PMW frequency of 5 in the BNC. There was one utterance of *get rid of his wife* (C1) by a C1 student when he spoke of his friend's divorce. *Get rid* of has a relatively high occurrence in the BNC of 18 PMW but it appears at C1 level in the EVP. Learners' use of such expressions is an indicator of fluency and indeed level as the idiomatic meaning often has no relationship with the literal interpretation.

Cultural Allusions and Catch Phrases

In both datasets, there are examples of cultural allusions and catch phrases, accounting for 5% and 3% of all idiomatic expressions in ACE, respectively. Among them were the following:

- Catch phrases: *okey dokey* (A2); *break a leg* (C1); *shit for you* (C1)*, to be or not to be* (C1); *once in a blue moon* (A2).

- Proverbs (including translated items): *a bird in the hand is worth two in the bush* (C1)*; the grass is greener on the other side* (C1); *bronze man* (in Italian meaning to be rich*; an elephant's dream* (C1) (in Latvian means something is impossible to happen).

As we discuss further below, many of these play an important role in casual conversations and narratives within everyday exchanges.

Clichés

There are 52 (PMW) cliché expressions in the corpus (3%). The expression *just one life* (A2) has been used at such a high level of frequency that it has now lost its originality and power. *I've had enough* (C1) and *longtime no see* (A2) are other examples. *Shit is shit* (A2) implies that nothing of the situation can be altered and *teach me the best lesson in life* was used by a C1 learner. The occurrence of *break a leg* and the Polish equivalent *shit for you* highlight a manner of wishing a person good luck, while a C1 student stated *at the end of the day.*

Prepositional Idioms

Prepositional idioms account for 1% of the idioms drawn from the ACE corpus. The preposition and complement structure can be seen in the expressions *over the noise (C1)*, which occurs 0.9 times (PMW) in the BNC. A further example in the corpus of a prepositional idiom can be seen in the utterance: *in this case.* Interestingly, *in this case* appears 27 times (PMW) in ACE and also has a frequency of 27 (PMW) in the BNC.

Binominals and Trinomials

In all, 23 (PMW) binomial/trinomial expressions were found in the ACE corpus which accounts for 1% of the idiomatic expressions in the corpus. *Back and forth* occurs 7 (PMW) times in the C1 cohort of the ACE corpus, *black and blue* appears 6 (PMW) times in both the A2 and C1 cohorts, while *tall dark and handsome* was uttered once by an A2 learner. Based on the structure of noun + conjunction + noun *spoon after dinner* said by a student from the Democratic Republic of Congo could be included as an example. This is roughly translated into English as to do something stupid or incorrect.

Frozen Similes

Overall, 29 (PMW) frozen similes were found in the ACE corpus. These 29 expressions account for 1% of the idioms found in the corpus. These idiomatic expressions make comparisons using the adverbs *like* or *as* and can often be identified by removing the first *as*. *Hungry like a wolf* was articulated by an A2 student, although not very common in native speaker British English vernacular, it was made popular by the 1997 *Duran Duran* hit. It is also apparently well-known in Polish. In identifying the stupidity of a person, a student uttered the phrases *as thick as two short planks* (A2) and *as thick as shit* (A2). Evidently, this student acquired these frequently colloquial expressions through being embedded in the local culture. There was one utterance of *sleep like a log* (C1), *drink like a fish* (A2) and *get on like a house on fire* (C1). We note that these expressions are all part of daily narratives when talking about ourselves or our friends.

Stylistic Formulae

These function as a means of distinguishing between written and spoken language, formal and informal language, monologue and dialogue and the various categories of English for Specific Purposes (ESP) and English for Academic Purposes (EAP). There are 17 occurrences of such expressions in the corpus which account for 1% of the overall occurrence of idioms. The majority of which were all uttered during oral presentations and interviews: The greetings *ladies and gentlemen* occurred 29 (PMW) times (A2, C1) and *good evening* occurred 47 (PMW) times (A2, C1) both expressions occur 2.4 and 6.2 times in the BNC (PMWs), respectively. Both expressions appeared primarily at the beginning of oral presentations. The closing of *with pleasure* (C1) occurred once and also reinforces a level of formality at which the speaker wished to be seen.

Possessive Phrases

In all, two examples of possessive phrases were found in the ACE and they are used in both the A2 and C1 sub-corpora: *the bee's knees* (A2) and *the lion's share* (C1). While as a category, these forms

are low in overall occurrence, it is of significance to note that there is no occurrence of either of these idioms in the whole of the 100 million-word BNC.

Table 8.6 above lists the most frequency idioms (top 20). We now take a look at some of these in terms of how they are used in ACE.

From time to time

The prepositional idiom *from time to time* is the most frequent idiom in the ACE corpus with 82 PMW occurrences. It is a B2 level idiom according to the EVP yet its frequency is much lower in the BNC. As detailed in Table 8.6, where we see that it has 14.77 occurrences PMW in the BNC. When we look at its occurrence, we see an example below from the spoken corpus. It was used by a C1 student and then repeated by others in the group over a number of turns. This partly explains how idioms are being used more frequently in the ACE data.

Extract 8.1 (Student <$24> C1 spoken corpus)

…From time to time I like to read in the city library.

How's it going?

Also with a PMW frequency of 82 is the Irish-English greeting: *how's it going?* This is colloquial way of saying hello in Irish English (Clancy, 2010). It has a PMW frequency of 0.6 in the BNC but in LCIE it occurs 6 times PMWs. It occurs across both the A2 and C1 levels and in both spoken and written corpora. As expected it is not listed in the EVP, we can assume that learners can only acquire this phrase in their local language context. Extract 8.2 shows an A2 level student greeting classmates and the teacher using this expression.

Extract 8.2 (<$12>A2 spoken corpus)

…hey guys, hi Justin, **how's it going?**

This is an interesting example of a very frequent idiomatic expression that the ESL learners experience on a daily basis. From the corpus, we can see that it has become part of the learner's vocabulary.

Are you pulling my leg?* and *blackmail

The archetypical, interrogative full idiom *are you pulling my leg?* and the opaque nominal compound *blackmail* are the next jointly most frequently used items. Both occur at a frequency of 71 (PMW) in the ACE corpus. *Are you pulling my leg?* occurs in both the A2 and C1 spoken corpora, does not occur in the EVP and occurs 0.34 (PMW) in the BNC,. On the other hand, *blackmail* occurs only in the C1 cohort in both the spoken and written data. It is placed at C2 level in the EVP yet it has an occurrence of 4.36 (PMW) in the BNC, which is relatively high for an idiom.

Extract 8.3 (Student <$14> C1 spoken corpus)

He then decide to **blackmail** her for cash.

Face a problem* and *throw up

The next most frequently occurring idioms in the ACE corpus are *to face a problem* and the idiomatic phrasal verb *to throw up*, both with 70 occurrences (PMW). *To face a problem* occurs at both A2 and C1 level, in both the spoken and written datasets. It is placed at B2 level in the EVP and has an occurrence of 0.5 (PMWs) in the BNC. In contrast, the idiomatic phrasal verb *throw up* only occurs at A2 level, at B2 level in the EVP and occurs 5 times (PMW) in the BNC.

Extract 8.4 (Student <$7> A2 spoken corpus)

I had many **problems to face** before I come to Ireland.

Pass time

The idiomatic expression *to pass time* occurs 64 times (PMW) in the ACE spoken corpus. It occurs at C1 level in the ACE while it occurs at B1 level in the EVP and has an occurrence of 1 (PMWs) in the BNC. Here we see an example of a C1 learner using the idiom in a natural way in speaking.

Extract 8.5 (Student <$24> C1 spoken corpus)

I do many things in my free time just **to pass the time**, really.

I could eat a horse

The hyperbolic *could eat a horse* was uttered by a C1 student and repeated by the group with an occurrence of 35 times (PMW) in the ACE. Interestingly, this expression does not appear in the EVP and only has an occurrence of 0.02 (PMW) in the BNC. As discussed above, it expresses everyday meaning.

Extract 8.6 (Student <$12> C1 spoken corpus)

I am starving **I could eat a horse**

As thick as two short planks

The colloquial and informal as *thick as two short planks* appeared with 47 PMW occurrences in the ACE corpus. It occurred at both A2 and C1 level in the spoken corpora. In comparison *as thick as two short planks* does not occur in the EVP. In the BNC, there are 0.01 PMW occurrences. Therefore, it is very notable to find it used by C1 level learners and quite exceptional that A2 level learners are using it.

Extract 8.7 (Student <$4> and <$6>A2 spoken corpus)

<$4>: bright is bright am intelligent? Can we say as thick as shit?

<$6>: or **as thick as two short planks** maybe?

It is interesting to see idioms used to *do* humour in the context of A2 learners.

Get over

The idiomatic phrasal verb is included both here and in Chapter 5: Multi Word Verbs. *Get over* occurs in the ACE corpus with a frequency of 47 occurrences (PMW) and it occurs in the spoken and written sub corpora of the A2 cohort. In comparison, *get over* occurs at B2 level in the EVP and has a frequency of 6.85 (PMW) in the BNC.

Extract 8.8 (Student <$9> A2 written corpus)

…She had many, a lot to do and she **get over** sickness quickly very.

On the one hand* and *on the other hand

The stylistic prepositional phrases on the one hand and on the other hand occur as collocations with the same frequency in the corpus. Each instance occurred together in the written section of the C1 corpus with 47 (PMW) occurrences each. In comparison, both expressions occur at B2 level in the EVP and *on the one hand* has a frequency of 12.6 (PMW) in the BNC while *on the other hand* occurs 47 (PMW) times.

Extract 8.9 (Student <$24> C1 written corpus)

On the one hand it was the best decision I ever made and **on the other hand** it was the most difficult.

Let us now look at the function of idiomatic expressions in ACE.

8.7 Analysis of the functions of idioms in the ACE corpus

At this point in the analysis, we will look at how the idioms in ACE functioned. Schmitt and Carter (2004) identified the main purposes of use of idioms:

- **Expressing a message or idea** *(a bird in the hand is worth more than two in a bush* = be happy with what you have).
- **Realizing functions** (*just browsing thanks* = declining an offer of assistance in a shop.
- **Expressing social solidarity** (*it sure is* =expressing agreement).
- **Transacting specific information in a precise and understandable way** (*He took off with her bag* = he stole her bag).
- **Signaling Discourse Organization** *(on the other hand* = conversely*).*

McCarthy (1998) argues that idioms are not mere 'quirks' of the English language chosen randomly by the speaker, but rather they function as communicative devices that hold numerous socio-interactional functions. Idioms are found in everyday jokes, stories, anecdotes, and descriptions and are according to Palma Fahey (2005), 'a communicative resource that perform a range of different functions' (2005: 191). It must also be stated that classroom task effect and opportunity of use did affect the outcome of the idioms used in the ACE corpus. For example, on the 4/11/11 students had an activity based on idioms using the course book *Headway Advanced* (Soars, Sayer: 2003), in which they were tasked with stating the English idioms they already knew and they also learned some new idioms.

The common functions performed by idioms in everyday spoken interactions also manifest themselves in the English speaking environment of the ACE corpus. They are indeed optional but they perform specific functions as they occur. An in-depth analysis of the functions of idioms in the English language is McCarthy (1998), who based his functional analysis of idioms on a large spoken corpus consisting of spoken data drawn from a number of different registers. This section will aim

to identify the functions of the idiomatic expressions used by the learners of ACE in relation to the various functions identified by McCarthy (1998). Table 8.8 summaries the functional categories found in the corpus by level across speaking and writing sub-corpora.

Table 8.8 Functions of idioms in the ACE Corpus, by level across speaking and writing

Function	A2	C1	Speaking	Writing
Idioms in everyday stories, anecdotes and jokes	√	√	√	√
Idioms in collaborative ideas	√		√	
Evaluation in oral narratives		√	√	
Negotiation of meaning and convergence		√	√	√
Discourse boundaries	√	√	√	√
Cultural solidarity	√	√	√	√

Table 8.8 shows us a spread of six functions and these are spread across levels and modes. Idioms are used for everyday narratives, anecdotes and jokes are used by both the A2 and C1 cohorts and in both the spoken and written corpora. Idioms are also used by both groups in collaborative ideas but interestingly this function is only expressed in the spoken corpus. The function evaluation in oral narratives is only expressed by the C1 cohort in their speaking. In addition, idioms are used by the C1 cohort to express the function negotiation of meaning and convergence in both the learners' speaking and writing but the A2 cohort do not use idioms to express this more complex function. Furthermore, idioms are used by both groups to highlight discourse boundaries between multiple speakers and occurs in both the spoken and written sub corpora. Finally, there are examples in both cohorts in speaking and writing of idioms used to express cultural solidarity. We will now take a

closer look at each of these functions.

Idioms in everyday stories, anecdotes and jokes

Idioms tend to appear in everyday stories and anecdotes and are, according to McCarthy (1998), far from random. They are used as an evaluation force which functions in enabling the storyteller to make the story appealing and worth listening to and they provide exciting and interesting elements of the story. Without the idiomatic use here in Extract 8.10, the story taken from the C1 spoken sub-corpus of the ACE would appear bland and boring to the listener.

Extract 8.10 (Student <$14> C1 spoken corpus)

[The speaker, Student 14, is reading an essay she wrote for homework on the movie the Lion King and introduces two new characters to the tale]

…At this point the **story takes a turn** because of two new characters Simon and Pomba: ah em a kind of African ferret and an African wild pork. Their philosophy is hakuna mata which is swailli ah phrase that is literally translate as there are no worries. Those two animals save him and they become best friends. **Far away from** his territory Simba grow up and forget fault with his new philosophy.

The change of scene in the story is compounded by the idiomatic phrasal verb *the story takes a turn.* This expression sounds more dramatic and imaginative than if the speaker used the adverb of sequence *next*. While, at the end of a tale, *Far away from* acts as a coda (Labov, 1972) which connect the story with reality. This last sentence performs a summarising role which is a known function of idioms.

Extract 8.11 (C1 cohort written corpus)

[<$6> from the C1 cohort has written the tale of a famous legend that exists in her home town of Transit, Spain. Again here, the idiomatic expressions used make the language sound more descriptive

and imaginative]

At the beginning they made quite noise but one day the noise stoped. The princess **couldn't contain herself** and then opent the door. In front of her, it was a scupture of the virgin of the Transit but it wasn't anything about the Pilgrims. It is said that these Pilgrims were two angels who stopped working when the door was opent. **For this reason** the that scupture doesn't have all fingers because of the curiosity of the princess.

The figurative language used in this piece of writing contributes to the colourful, imaginative scene of the legend. Discourse markers such as *at the beginning* (which also occurs 25 PMWs in the BNC) and *in front of her* give a spatial location to the tale. Clearly, idioms mark various junctures in stories. They are far from random and without such expressions, the tale would appear bland and boring which can be seen in the figurative *couldn't contain herself*. We note that *could/couldn't + contain herself* has total 0.4 PMW occurrences in the BNC. In the following example, we see an A2 learner using idioms in the spoken data. She is responding to a picture prompt based on which she must create a narrative story. As with all students, she is given one minute to plan what she would say:

Extract 8.12 (Student <$4> A2 spoken corpus)

And ah Mr and Ms Roosey went on vacation vacation holiday. And when they ah returned they Mr Roosey come into the room like ah white room. And all details and furniture broke. They are very scared about what's happened because they didn't know the real cause. The picture the furniture and flowers were on the floor. There was a in their house and the cup was wrong. They went of the am room and Mr listened and entered inside and all detail and all their furniture am were perfect order and all **bright. And they can sleep like a log**. But they were scared by that explain could not

In Student 4's story, we see the use of the idiomatic simile *sleep like a log* to reinforce the exhaustion of the travellers and to bring the story towards a close. We note that this idiom occurs only 0.1 times in the BNC so it is very noteworthy that it is used in an ad hoc way by the A2 student in a timed task. It suggests that this idiom is one that has become entrenched in her repertoire.

Idioms in Collaborative ideas

Collaborative ideas (McCarthy 1998) are one of the most frequent conversational activities. This is where speakers share views of the world and discuss topics they find interesting. There is normally no chronological order but a conversation based on personal observations. The idiom normally occurs in the comment stage of the discussion.

Extract 8.14 (C1 Cohort spoken corpus)

[Student <$20> from the C1 cohort is taking part in an oral examination with the teacher and is describing an event]

T: so tell me about your hobbies, am what do you do in your free time?

Student <$20>: am yes of course, I am like shopping, visiting new places eating Irish food

T: do you like the taste of Irish food?

Student <$20>: ah any food I love it Irish breakfast I love it

T: <$E> laughs<$E> me too and do you like drinking Irish alcohol?

Student <$20>: not am really any more…again

T: how come…why?

Student <$20>: when I come first time in Ireland I went to pub and got very drunk

T: oh no

Student <$20>: yes, and I remember anything

T: nothing

Student <$20>: yeah nothing I fall I think

T: oh no, did you have pain the next day?

Student <$20>: yup I was am eh am **black and blue** how you say

T: bruises

Student <$20>: yeah bruises so now I don't drink yet

T: you must be careful Hui <$E>laughs<$E>

Student <$20> uses the idiom *to be black and blue* as part of a formulation to describe his condition this idiom has a frequency of 0.1 (PMW) words in the BNC. Again we see a low frequency idiom used with ease here (preceded by the informal *yup*), suggesting that this is embedded in the speakers repertoire of lexical formulations.

Evaluation in Oral Narratives

In a spoken interaction, events are outlined first and the listener provides an evaluation of the event taking place. Idioms occur predominantly during the evaluation stage and signal interest or disinterest in the topic (McCarthy 1998). In the following extract, we focus on the evaluation being made by C1 level Student <$14> about a Polish actor. Two related idioms are used (*not very bright* and *as thick as shit*) by Student <$14>. What is interesting here is that Student <$14>, relexicalizes (McCarthy 1998) the second idiom using: *as think as two planks*. This shows an interesting example of two C1 students using idioms as a way of evaluating and converging in a spoken narrative.

Extract 8.15 (A2 Cohort spoken corpus)

[Students <$14> and <$16> from the A2 cohort are describing a well-known Polish actor

Student <$14>: oh my God eh he's awful

Student <$16>: <$E>laughs<$E> I really like him actually

Student <$14>: no way, he's not am eh **very am ah bright**

Student <$16>: bright what is bright? Stupid?

Student <$14>: ya he's **as thick as shit**

Student <$16>: wow <$E> laughs<$E> **as thick as two short planks** but I like him

This shows us a typical and frequent function of idioms. It begins with a factual observation and is then followed by an evaluative comment. McCarthy (1998) refers to this as the *observation plus comment function.* Note students had learned the phrase *thick as two short planks* in the previous class but what is important here is that we see it deployed collaboratively.

It seems that idioms tend to reflect, to an extent, a sense of informality between the speaker and the listener, at an interpersonal level. However, they also serve a purpose and function at the stages of which they occur in the communicative act. At the discourse level, it is clear that idioms are not random or unmotivated exclamations. Moon (1992) showed numerous examples of how idioms occur in text as an evaluative device. In this corpus analysis, we will show how idioms also occur as an evaluative tool in spoken interactions. Also of note from the literature is the assertion that idioms are much more likely to occur in speech when the speaker is referring to a third person or object (see Strässler, 1982). This highlights a less face threatening and evaluative function of idiom, as Extract 8.16 illustrates. Here Student 1 and Student 2 from the A2 cohort are engaged in a roleplay which they created during the first lesson. In the roleplay, they have met for the first time in a fitness club. In this extract we see the use of idiomatic catch phrases and discourse markers, illustrating how these A2 level learners have acquired an ability to deploy them in a roleplay conversation:

Extract 8.16 (A2 Cohort spoken corpus)

Student <$1>: So our dialogue is in the gym so <$E>laughs</$E>

Student <$1>: How...<$=>wow</$=>...sorry...**wow** you have big muscles...hi my name is John <$E>laughs<$E>

Student <$2>: <$E>laughs<$E> my name is Ali, you have big muscles too

Student <$1>: **Well?** How long you come here?

Student <$2>: This will be my ah third day and you?

Student <$1>: am about two weeks.

Student <$2>: **wow cool that's long enough**

Student <$1>: I'm going to the pub to have a <$=>pint</$=>...a pint of beer here in town...you want come with me?

Student <$2>: yeah sure, **that would be great** thank you

Student <$1>: ok when finish my exercise...mmm…see you after shower

(Students <$1> and <$2>: A2 spoken corpus)

In this dialogue both speakers respond to what the other says with idiomatic catch phrases and discourse markers for examples: *wow, cool, well.* *Cool* occurs 3.77 (PMW) words in the BNC. Also, the evaluative use of the idiomatic phrases occurs during the response stage of the conversation for example: *that's long enough; that would be great.* The vague discourse marker *about two weeks* makes the language sound less stilted and predictable. In Extract 8.17, we again see A2 level students apparently fluently using idiomatic expressions in a speaking task. Here Student 10 and Student 11 have been given the same speaking task where they must create a conversation with a stranger. They have decided to have met at a bus stop:

Extract 8.17 (A2 Cohort spoken corpus)

Student <$10>: ah hello how are you?

Student <$11>: I'm okay

Student <$10>: Where are you from?

Student <$11>: I'm from Lithuania

Student <$10>: I'm from Lithuania too how are you?

Student <$11>: **I'm fine**...<$E>laughs<$E>>...What are you doing in Ireland?

Student <$10>: I work in pub and you?

Student <$11>: ...**oh cool**...<$E>laughs<$E>> I don't work in...**at...moment oh hard work**

Student <$10>: Oh ok... would you like for us to go to pub sometime?

Student <$11>: **yes of course**

Again, in this conversation all of the idiomatic phrases occur in the response stage of the conversation. The evaluative stage contains discourse markers and gambits such as *I'm fine* (1.87 occurrences per million words in the BNC). In the case of the response *oh cool, yes of course,* the BNC PMW results show: *Oh cool* 0.03 (PMW); yes of course 1.09 (PMW); 268 (PMW) occurrences. When responding

to Student 11's announcement of their job Student 10 uses the semi-idiom of *hard work* which, in turn, occurs 12.08 PMWs in the BNC.

Negotiation of meaning and convergence

Another frequent feature of idioms is where the speakers are negotiating lexical meaning. The vague meanings of many idioms can help speakers in negotiating a lexical meaning of a unit. In Extract 8.19, A2 Student <$10> is describing the difficulty he has when studying philosophy. We see how he and his interlocutor negotiate meaning and converge using an idiom. For A2 level, this is not normally expected:

Extract <$8> (A2 Cohort spoken corpus)

Student <$8>: I mean it must be am so difficult

Student <$10>: oh **ya of course** for me it always a pain

Student <$8>: ye totally

Student 10: **a pain in the ass**

The idiom *pain in the ass* occurs just 0.1 times in the BNC. We can assume however that the A2 student has heard it used evaluatively in his ESL environment on a number of occasions so he can be said to have experienced it in use.

Another interesting example is found in the C1 data where a speaker uses the vague word *yoke*.

Extract 8.19 (C1 Cohort spoken corpus)

[Student <$3> from the C1 cohort is explaining a slang Irish word that she has heard]

Student <$3>: I mean I find it very useful…it can be used for anything anything at all **.. yoke**

Student <$22>: but what does it means?

Student <$3>: anything… I can say pass me that **yoke** <$E>points to phone<SE> can I have one of those **yokes** <$E> points at chair<$E> <$E>laughs<$E>

Student <$22>: so we no need learn English just one word

T: <$E>laughs<$E> but you won't be understood anywhere outside of possibly Ireland.

Yoke is a colloquial Irish expression which enables speakers to avoid precise labelling of a noun. It is used 33 times PMW in LCIE but only 1.7 times PMW in the BNC (see also Irish Times Newspaper, 2013)]. Again, here we posit that the learner has experienced this term within her everyday experience of language in Ireland and it has now become an entrenced idiomatic expression for her.

Discourse Boundaries

Speakers often choose to use idioms to change topic or to end a conversation in a polite nonthreatening manner. According to Drew and Holt, (1998) the idioms common at this closing stage are clichés, proverbs and cultural sayings. The use of such a technique ensures a rapport is maintained between the speaker and listener. The idiom aids in ensuring that the expression is non offensive or insulting. This of course ties in with functions already discussed above such as evaluation and convergence. In Extract 8.20, we see an A2 learner tactfully deploy an idiom to bring a topic to a close using the proverb *the grass is (always) greener on the other side*. Here Student 12 is letting Student 16 know of her experience in a similar situation but is being polite in doing so and trying not to offend. The proverb *the grass is (always) greener on the other side* occurs 0.08 times PMWs in the BNC but some of these uses are literal. Given that this interaction took place between two A2 level students, we see this example of the strategic use of an idiomatic expression as very noteworthy. What is also striking is that the student uses the proverb creatively in the negative.

Extract 8.20 (A2 Cohort spoken corpus)

In the following conversation from the.

Student <$12>: I have the similar experience

Student <$16>: For me I am could not wait to eh leave and then after short time I hated everything

Student <$12>: me too I learned

Student <$16>: yeah I thought everything would be great

Student <$12>: **the grass is not always greener**

Student <$16>: I know I know

In Extract 8.21, we see C1 level students (24 and 11) in a situation of conflict when discussing their differing views on gender roles. The interaction becomes more face threatening for both students but then we see Student 11 using the idiomatic expression *the end of the day* along with a formulation as a means of mitigating the point at which the disagreement and conflict could have escalated greatly. Student 24 responds with *of course* and the situation is diffused. We note that the expression *at the end of the day* occurs 6.79 PMWs in the BNC.

Extract 8.21 (C1 Cohort spoken corpus)

Student <$11>: You must have ya old views ah em when the woman stays at home ya?

Student <$24>: not always but in my country, men work women don't women ah

Student <$11>: cooks, cleans

Student <$24>: yes and takes care of children house. I think it's normal

Student <$11>: normal <$E>laughs<$E>it's not normal

Student <$24>: for me is

Student <$11>: tut tut crazy

Student <$24>: no

Student <$11>: we have different ideas **at the end of the day**

Student <$24>: **of course**

Cultural Solidarity

The final function listed in Table 8.8 above is *cultural solidarity.* Strässler (1982: 119) warns us not to use an idiom 'if you believe you are in a situation which does not allow such use. Do not use idioms if you are not sure about the present situation'. As discussed previously by Malinowski (1923) and developed further by Firth (1957) and the Neo Firthians: Sinclair (1991 and 2000) and Halliday (1985), idioms are ligosemic. Their meaning is embedded and understood in the culture and context in which they are expressed. The use of idioms allows language users to convey a social identity. Idioms contain a high degree of context, culture and common knowledge and the ACE corpus is no exception. Though the participants come from numerous and varying nationalities, they hold a commonality in that they are living, working and studying in a city in Ireland. Undoubtedly, the learners acquired or 'picked up' many idiomatic, colloquial Irish English expressions. These idioms are used as a common bonding mechanism to convey a cultural belonging of the speaker. Greetings and pleasantries were a common form of "Irishisms" and colloquialisms in general used by the students. Phrases such as *how's it hanging?* and *what's the story*? meaning *how are you*? and functioning as conversation openers, each were uttered twice by different speakers in the corpus. *Sounds good to me* was used once showing agreement between a pair in completing an activity. One speaker used the expression *give me a hand* when looking for assistance. See table 8.9 below for a further discussion on the 'Irishisms' and colloquialisms used by learners of the ACE corpus. Extract 8.22 offers an example of a very colloquial and very Irish euphemism *get the shift.* A *shift* refers to a sexual encounter. There are no examples of this use of *shift* in the BNC while it occurs 8 times PMW in LCIE. Interestingly in this exchange between A2 Student 4 and the Teacher, the student reports having heard the expression and clear the student needs to have it explained by the teacher. We note that it is interesting that the student has observed this expression as a salient item. The student is able to report it back with the correct tense also.

Extract 8.22 (A2 Cohort spoken corpus)

[Student <$4> from the A2 cohort is attempting to clarify the meaning of a frequently used Irish term]

Student <$4>: my friend ah he ah said [he] **got the ah shift**, what is this?

Teacher: <$E>laughs<$E> well the literal meaning is to move something from one place to another

Student <$4>: oh ok

Teacher: but I am I think your friend meant something else

Student <$4>: yeah

Teacher: he meant he kissed someone

Student <$4>: oh my God <$E>laughs<$E> it's no sense no means

Teacher: yes I know, it's colloquial, slang

In Extracts 8.23 and 8.24, we see C1 student using two colloquialisms though they are not exclusively to Irish English, they are certainly not found in text books so we can assume they were acquired through interaction in the L2 environment. They relate to expressions for going to the toilet: *go to the loo* and to *be bursting* (in need of going to the toilet). The latter has 0.61 (PMW) occurrences in LCIE and 0.63 in the BNC. *Bursting to pee* as a fixed expression does not occur in either LCIE or BNC.

Extract 8.23 (C1 Cohort spoken corpus)

Student 8 from the C1 cohort uses two idiomatic Irish colloquial expressions when he states

Extract 8.24 (C1 Cohort spoken corpus)

Student <$8>: can I go to the am **loo, I am bursting to pee.**

The loo is a common term in Ireland and the UK for the bathroom while the hyperbolic expression of *bursting to pee* is a frequent term used by many Irish people when they are describing the urgency in which they need to use the restroom. Anyone interacting in a casual way in the UK or Ireland would hear these expressions. This again suggests that ESL learners experience with language in their L2 environment has an impact on the language they are acquiring and using.

Other colloquial expressions which occur in the corpus can be seen in Table 8.7 below where we can also see the low occurrence in the BNC; this highlights the fact that these expressions occur more in Hiberno-English. We can also see that these considered colloquial Irish English expressions do not occur either at a higher rate in the Irish corpus; Limerick Corpus of Irish English (LCIE) (a one million word corpus of English spoken in the Republic of Ireland taken from recordings of conversations across a wide variety of settings, 82% of which are casual and informal interactions). However, where they do occur in the LCIE they clearly have a higher frequency in the Irish corpus than the BNC. Note that where an item such as *over* occurs with a high frequency in the BNC, 100 word samples were used to check for the specific meaning in Irish English. In these cases the frequency is reported in *number of occurrences / 100 concordances lines* as footnotes explain.

Table 8.9 Colloquial Irish expressions in the Spoken ACE corpus.

Irish Expression	**A2**	**C1**	**Meaning**	**BNC (PMW)**	**LCIE (PMW)**
<$1> *What's the craic*?	√		Hello, how are you?	0	1
<$22> My neighbour they always fight right and I can't sleep *over* the noise		√	Because of or due to	(0/100)[11]	(1/100)
<$18> A boy the kid		√	Hello, how are you?	0	0
<$4> he's a fine thing	√		He is very attractive	0	2
<$12> I am fed up with maths it's a pain in the ass…		√	It's so annoying	0.1	6
<$12>…I'm sick of it		√	I've had enough	0.4	6

[11] In a random search of 100 concordance lines of OVER in the BNC no instance of this usage were identified in 100 concordance lines of LCIE 1 occurrence was identified.

<$9> He isn't very bright is he? <$10> bright… is bright? ah ah intelligent? Can we say he is *as thick as shit?*	√		He is not very intelligent or clever	0.02	0
<$20> he said ah am he *got the* ah *shift*	√		He kissed someone	0	0
<$17> I can't go *I am wrecked*		√	I am exhausted	(2/100)[12]	2
<$3> *yoke* can be used for anything…can you give me that yoke	√		Thing or object	0	48
<$4> *I'm grand* how are you?	√		Ok or fine	0[13]	29
<$01> *well,* how are you?		√	Hello	0	3

What this table reinforces is that ESL learners are acquiring colloquial idiomatic expression in Ireland from their everyday interactions. These are very casual expressions and being able to use them will help an ESL learner be part of a group. For example, *A boy the kid,* is a friendly greeting only used in Limerick city. Being able to use this idiomatic routine will endear the user within a workplace or a social setting. Strikingly, these expression are not just found in the C1 data but are also in the Elementary level language.

[12] 100 concordance lines of lemma *BE + wrecked* in the BNC yielded 2 instances where it was used to mean *tired*
[13] In the 100 concordance sample from the BNC, *I'm grand* occurred twice but it was from reported speech Irish novels by Irish authors or depicting Irish characters.

8.8 Conclusion

The analysis in this chapter shows that the ESL learners within the ACE corpus do indeed use idiomatic expressions in their speaking and writing, across a range of forms and functions. Such expressions when analysed in terms of their form and function were often comparable to that of native speaker use. As seen, overall, there are 229 idiomatic expressions in the C1 corpus (54%), averaging at fifteen per student, and 194 in the A2 corpus (46%), normalising at 21 per student. Of interest is the contention that these results show, in comparison to the statistics provided by the COE and IELTS examination, levels as low as A2 learners are in fact using idiomatic expressions. Contrary to available syllabi and frameworks stipulations, it is not just above C1 and Band 6 where these expressions are being used. When we looked at the top 20 most frequently used idiomatic expressions in ACE and compared them with the EVP to check at what level, if any, learners are expected to acquire them, we find that it lists only 12 of the 20 items. All of these 12 items are placed at B1 level and upwards. Therefore, all A2 users exceed the expectations of their level.

The unexpected use of idioms (especially by A2 learners) and the use of colloquial expressions is undoubtedly a result of living and working in Ireland where they gradually acquired such cultural specific phrases. In terms of form, the five most frequent forms of figurative expressions were: 1) Idiomatic Speech Routines; 2) Hyperbole; 3) Clausal or sentential Idioms; 4) Idiomatic phrasal Verbs; and 5) Cultural Allusions. When we revisit Table 8.6 in terms of percentage, we can see that Idiomatic Speech Routines and Hyperbole account for 50% of all idiom forms and this underscores the use of idioms in everyday interactions and speech routines, using hyperbole in narratives for humorous effect, for example.

In conclusion, this section has shown that idioms are never just spontaneous, *ad hoc,* figurative alternatives to literal equivalents instead they are chosen by the speaker for a reason and that can be seen here in their functions. Idioms are used by the language learners to elaborate jokes and stories.

They focus as an evaluative force in a narrative. They aid in collaborating ideas. They are used by some as a means of negotiating meaning. They tend to mark discourse boundaries and also can convey a sense of Irish cultural solidarity. Idioms have been shown to be a communal token of a culture and frequently display an idea of this cultural and social solidarity, which is evidenced by the colloquial ("Irishisms") expressions used by the majority of learners.

Moreover, it has been highlighted how the context and culture in which these expressions occur are vital to interpreting and understanding the expression. Clearly idioms are ligosemic, and it could be argued that a strong incentive for their use here by participants is due to their length of immersion in Ireland and in Irish culture. We have seen how idioms comment on the world around them and do not merely describe it and it has been shown that idioms can appear face threatening to the listener and are thus often used when referring to a third person or object. Finally, as highlighted previously, idioms occur in a number of different structures. In our data sets we have identified clausal, phrasal, prepositional, frozen similes, binomials and trinomials, and so on. Numerous idiomatic forms have been identified across the different datasets which function in proving that language learners, despite their level of competency, do indeed use and understand fixed expressions and idioms.

In this chapter, we have also speculated that the use of idiomatic expression linked to every day routines and interactions, often very colloquial in nature, ties in with a model of language acquisition that is based on language experience, namely the usage-based model (Ellis et al. 2015). As discussed above, so many of the expressions discussed in this chapter relate to basic human needs and functions (feeling thirsty, hungry, tired, needing the bathroom). They are very often informal and colloquial uses of a language as well and do not normally appear in English Language Teaching materials. We speculate that learners have acquired them through their local ESL interactions in Ireland. In experiencing these expressions, even at Elementary level, they quickly become *entrenched* and core to the repertoire of the learners. This aligns with the usage-based model of language acquisition where

frequently experienced and used patterns become *entrenched* semantic and syntactic units in the mind of the language learner (Ellis et al. 2015).

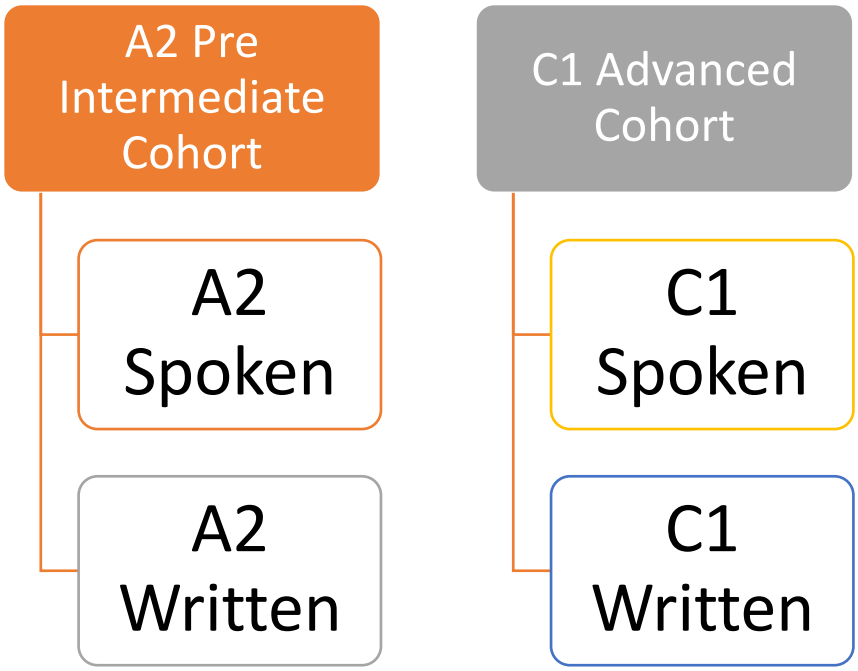
Chapter 9 Conclusion

> ‘In the end I learned a lot about language, communication and I think life. It was very enjoyable and fun and I will miss these classes’ (Student <$1> C1 cohort).

9.0 The ACE corpus and results

This study focused on the analysis of both spoken and written language drawn from four datasets of the ACE Corpus. The four datasets that make up this corpus are based on the two levels of competency of the ESOL students at the college: A2 Pre-Intermediate and C1 Upper-Intermediate/Advanced. As detailed in Chapter 4 Methodology, these students were divided into their relevant cohort based on their results in the college’s written placement test. The ESL students of the corpus were from various nationalities and L1 backgrounds (see Appendix 3 for metadata). In total, the recordings and written manuscripts of 24 students constitute the corpus; this is comprised of 9 students in the A2 cohort and 15 students in the C1 cohort. Overall, the corpus amounts to 170,000 words, which is divided into 150,000 words of spoken data and 20,000 words of written data. The corpus is then divided into four data sets, which are detailed in the Fig. below:

Fig 9.1 The division of the ACE corpus



The data was collected during the spring semester of 2010 to 2011. Recordings of students' classroom interactions, pair work, group work, individual oral presentations and the final oral examination were all transcribed to form the spoken component of the corpus. Various classroom activities, individual student's essay writing and progress examinations all combined to form the 20,000 words of the written data for the corpus. All results and findings were then normalised per million words in order to make comparisons with larger corpora (in this case the 100 million word BNC) possible. All classes were taught using the communicative language teaching method. The overall aim of the research was to establish whether the ESOL learners at the college used multi-word lexical items in their speech and writing.

9.1 The four lexical strings

Due to their high frequency of occurrence in native speaker English language, the four lexical items chosen for analysis were: 1) Multi-word Verbs 2) Delexical Verbs 3) Collocations and finally 4) idiomatic expressions. In order to give an accurate comparison between the ESL learners' use of the four features and the use of said features by native speakers of English, the findings of each multi-word lexical item in the ACE corpus were compared with the BNC. The EVP was also used as a comparison in terms of when learners are generally expected to learn these items (if at all). The results function to highlight comparisons and contrasts between ESOL speakers and native speakers but they also function to bring to life differences between ESL learners and EFL learners based on the expectations of the EVP. Let us summarise across the four main areas of analysis. When we bring the results together from the four analysis chapters (5 to 8), we find that all C1 learners use all of the four features and A2 learners use them to varying degrees as detailed in summary Table 9.1:

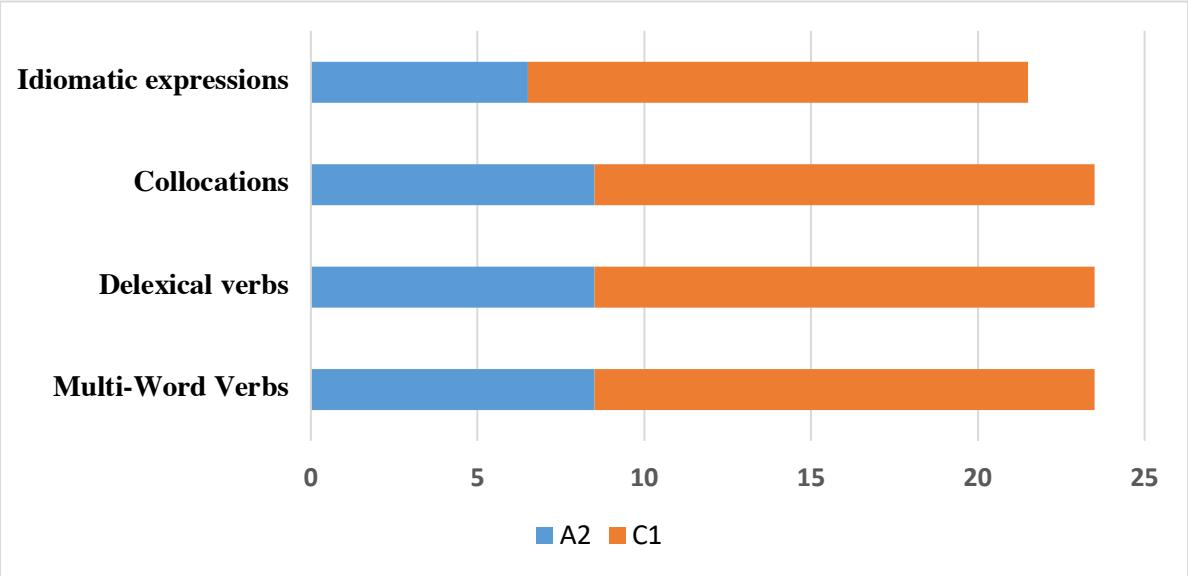
Table 9.1 Summary of percentage use of the four lexical string features

Multi-Word Verbs				
	Speaking		**Writing**	
	Raw	**%**	**Raw**	**%**
A2 (9 participants)	5	55	6	67
C1 (15 participants)	15	100	15	100
Total	20		21	
Delexical verbs				
	Speaking		**Writing**	
	Raw	**%**	**Raw**	**%**
A2 (9 participants)	4	44	9	100

C1 (15 participants)	15	100	15	100
Total	19		24	
Collocations				
	Speaking		**Writing**	
	Raw	**%**	**Raw**	**%**
A2 (9 participants)	9	100%	8	88%
C1 (15 participants)	15	100%	15	100%
Total	24		23	
Idiomatic expressions				
	Speaking		**Writing**	
	Raw	**%**	**Raw**	**%**
A2 (9 participants)	6	66%	7	77%
C1 (15 participants)	15	100%	15	100%
Total	21		22	

As Table 9.1 illustrates, there is variation in the numbers of A2 learners using the features but in all cases there is usage at this Elementary level of items that are not expected within the syllabi or EVP. Fig. 9.1, below summaries the results from Table 9.1 further into averages of numbers of students per level who use each of the four features:

Fig. 9.2 Summary of average number of students using four lexical string features, by level



While average numbers for A2 level students are lower, there is ample evidence from this of widespread use of these four features in ACE, at both levels.

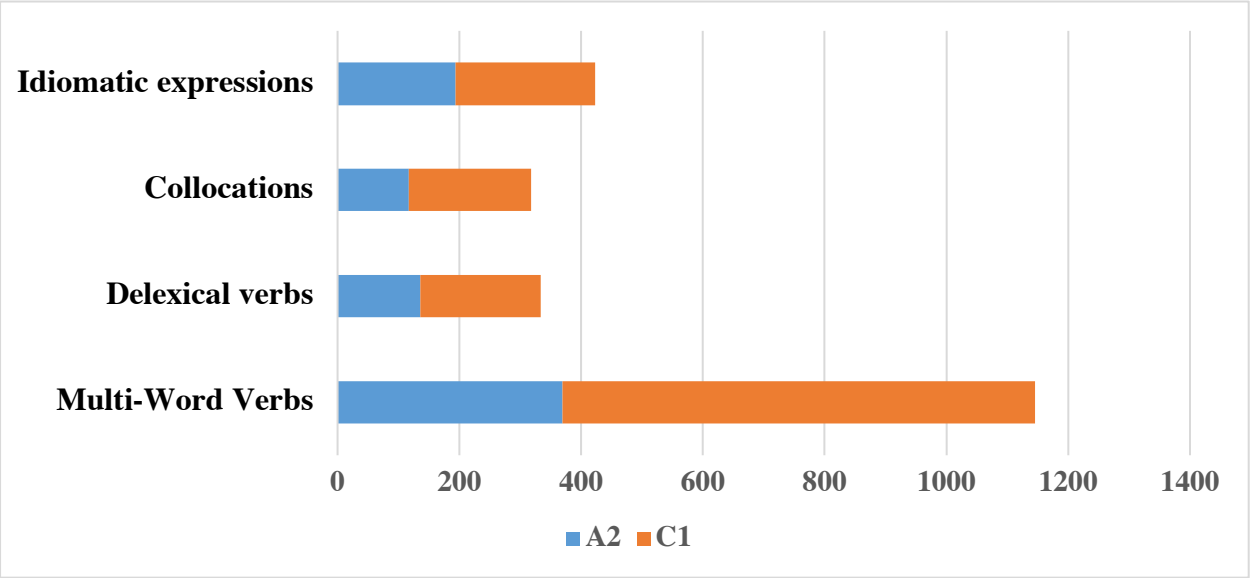
Table 9.2 and Fig. 9.2 show us summaries of the four features in terms of their frequency of use by level:

Table 9.2 Summary of total frequencies of four features, by level, sub-corpus

Multi-Word Verbs				
Level	**Spoken (raw)**	**PMW**	**Written (raw)**	**PMW**
A2	324	5063	45	6429
C1	522	6070	255	19,615
Total	**846**		**300**	
Delexical verbs				

Level	Spoken (raw)	PMW	Written (raw)	PMW
A2	86	1344	50	7143
C1	108	1256	90	6923
Total	**194**		**140**	
Collocations				
Level	**Spoken (raw)**	**PMW**	**Written (raw)**	**PMW**
A2	76	1188	41	5857
C1	116	1349	85	6538
Total	**192**		**126**	
Idiomatic expressions				
Level	**Spoken (raw)**	**PMW**	**Written (raw)**	**PMW**
A2	112	1750	82	11714
C1	152	1767	77	5923
Total	**264**		**159**	

Fig. 9.3 Summary of raw number of occurrences by feature type, by level



In broad terms, from Table 9.2 and Fig. 9.2, we can see that:

- Multi-word units are the most frequent type of lexical string that learners use in ACE. With the other three types occurring around one third as often as MWVs.
- In all cases, both A2 and C1 learners use the features but overall more items in each category occur in the C1 data, which is not surprising. In the case of MWVs, the rate of use is almost three times greater at C1 level. However, A2 learners are not generally expected to use MWVs widely, as we discuss below.

Let us now consider these findings in terms of the specific research questions of the present study.

9.2 Returning to the Research Questions

To return to the research questions as outlined in Chapter 1, the results can be summarised as follows in relation to the main research question:

1. *To what degree can adult learners of English at levels A2 and C1 use multi word lexical items in their speech and writing?*

- The findings from the ACE corpus, as summarised above, show us that the learners of both cohorts, in general, can use all four types of lexical string items under investigation, albeit to varying degrees, in both speech and writing. In the case of C1 learners, all students use all four features in their speaking and writing while some A2 learners use them in speaking and writing.
- All of the spoken occurrences of the language features were uttered by the learners in response to teacher questions, pair and mingling interactions and oral presentations. All of the written occurrences were then taken from the learners' essays, examinations and portfolios. Therefore, the students' use of these four lexical features were often spontaneous and unplanned.
- We note that there is a degree of opportunity of use at play as certain classroom activities elicited multi-word verbs and idiomatic expressions as discussed in individual chapters. However no explicit lessons on delexical verbs or collocations were given.
- The learners in the corpus had been living in Ireland for various lengths of time; some for weeks and others for years (see Appendix 3). The researcher is accrediting many of these expressions to the usage-based language acquisition that fits well with the usage-based model of learning, as we have mentioned in chapters 1, 5, 6, 7 and 8. As a result of the learners living in a host country of the target language, they acquired or 'picked up' various local expressions through experience. These items have been acquired as syntactic and semantic units and have become part of their lexical repertoire. This claim is substantiated by the high occurrence of colloquial language or 'Irishisms' found in the corpus (See chapter 8 Idiomatic Expressions).

In relation to sub question (i): *What is the difference between the lexical competence of the A2 level student and the C1 level student*?

- Integral to this research are the findings that both cohorts A2 and C1 used to an extent the four lexical features in their classroom speaking and writing. Both cohorts demonstrated the ability to use 1) Multi-Word verbs 2) Delexical Verbs 3) Collocations and 4) Idiomatic Expressions to differing degrees (see Tables 9.1 and 9.2 above). It is interesting to note that the specific learner outcomes of cohorts do not include these features so it is possible to attribute learner knowledge to the students' time in Ireland. While often findings have shown that the A2 cohort use these lexical strings with less accuracy than the C1 group, they show that they have a unitary understanding of the components in terms of their meaning. The use of MWVs in the written corpus show up the greatest difference across levels (see Table 5.3), with C1 learners using more than three times the amount of items compared with A2 learners. Fundamentally, the results that the A2 learners were attempting to use these features in their speaking and writing even though they are not expected to at this level.
- As was expected, prior to the data gathering stage of this research, the C1 learners were more communicative and interactive in the classroom. As estimated by both Nation (1990) and Milton (2011), a C1 level student should possess a lexical canon of around 7,000 to 8,000 words, while at the opposite end of the spectrum, the average A2 speaker, should know roughly 3,500. This is reflected in both the speaking and writing of both cohorts as all students involved in this study took *Nation's Vocabulary Size Test* (Nation, 1990) and the results correlated with the above estimates.

  In relation to sub question (ii): *Which level (A2 or C1) uses the majority of multi word units and language strings in speech and writing?*

This sub-question shows interesting results, as illustrated in Table 9.2 and Fig. 9.2 above and in Chapter 5, see Table 5.3. As discussed, while MWVs are by far the most frequent feature of the four that we examined (see Fig. 9.2 above and Table 9.3 below), it is here that the greatest differences appears between levels. The A2 cohort used the least multi-word verbs in their speech and writing: 369 across 5 of a total of 9 students compared with 777 items used by all 15 of the C1 cohort. As discussed in Chapter 5, within this result, there is further variation when we compare by the sub-corpora (Table 5.3).

Table 5.3: Breakdown of total number of occurrences of MWVs across levels and spoken versus written data (raw and PMW)

Level	**Spoken corpus**	**PMW**	**Written corpus**	**PMW occurrences**
A2	324	5063	45	6429
C1	522	6070	255	19,615
Total	**846**		**300**	

Here we see that PMWs, C1 learners use MWVs three times more than A2 learners in writing. The difference in PMW results in speaking across levels do not reflect this. On one hand, it shows that when C1 learners are engaged in writing tasks and have time to prepare or draw on their repertoire (and display it), their output of MWVs is three times greater than A2 students. We also noted in Chapter 5 that the EVP level of the items used at A2 is lower than that of C1.

When we look at the other three language string features, the results are somewhat, surprising in that the differences per level are not very great in terms of frequency. Table 9.3 shows us the frequencies of the four features by level (based on Fig. 9.2 above):

Table 9.3 Raw number of occurrences by feature type, by level (based on Fig. 9.3)

	A2	**C1**
Multi-Word Verbs	369	777
Delexical verbs	136	198
Collocations	117	201
Idiomatic expressions	194	229

As discussed, the discrepancy between A2 and C1 cohorts is high in relation to MWVs but this is far less the case for delexical verbs, collocations and idiomatic expressions. If we take the PMW results from Table 9.3 on their own for comparison (see Table 9.4) across levels and spoken and written sub-corpora, we see that frequencies are quite comparable and in some cases, the PMW usage is higher at A2 level. For instance:

- A2 learners use more delexical verbs in speaking PMWs than C1 learners in both speaking and writing;
- PMW, A2 learners use almost the amount of idiomatic expressions as C1 learners in speaking while in writing, A2 learners use more than double the number of idiomatic expressions PMWs than their C1 counterparts. Obviously, these results do now detail the complexities of form, meaning and use but nonetheless, they point to A2 learners frequent use of these items.

Table 9.4 PMW frequency comparisons across four lexical string features

Level	Freq. of occurrences Speaking PMW	Freq. of occurrences Writing PMW
Multi-Word Verbs		
A2	5,063	6,429
C1	6,070	19,615
Delexical verbs		
A2	1,344	7,143
C1	1,256	6,923
Collocations		
A2	1,188	5,857
C1	1,349	6,538
Idiomatic expressions		
A2	1,750	11,714
C1	1,767	5,923

It is important to remind ourselves that these items are generally neither taught or expected to be acquired at A2 level therefore these results are very striking.

In relation to sub question (iii)*: Is there a progression from more high frequency transparent strings towards low frequency opaque strings?*

Transparency rather than opacity is the key feature of the A2 cohort's use of the four lexical features. The majority of multi-word verbs, delexical verbs, collocations and even to an extent the idioms articulated by the lower cohort are transparent in meaning, particularly learners' use

of multi-word and delexical verbs. The A2 group used chunks such as *put it on the table*, *take off your shoes*, *have a party* and *go shopping*; again this could be due to their tendency as a group to avoid figurativeness. Nevertheless, certain learners in the group did use opaque idioms, such as the simile *sleep like a log* and *tears rolling down her cheek* for example. In contrast the C1 cohort used a lot more opaque low frequency strings. Students in this group used idiomatic multi-word and delexical verbs, such as *look forward to*, *put up with*, *make a decision* and *do my hair*. Ergo, it appears that as these learners transverse from A2 to C1 level, there is certainly a progression from more high frequency and transparent strings to low frequency and opaque strings. This finding aligns again with the SLA usage-based model discussed throughout in relation to the results. Essentially, this model sees the process of acquiring an additional language as a process of hearing, reading and finding meaning patterns of language over and over again and through this process, strings of words that occur together gain meaning through frequent encounter (or through overt teaching). According to Ellis (2003), these patterns develop along a similar cline from formula to low scope patterns to fully abstracted meaningful constructions or chunks.

In summary, according to this model, the acquisition process is input-driven and depends upon exposure to meaningful form-function relations. The fact that A2 learners first acquire transparent units across the four features ties in with this notion.

9.3 Relating the findings to theory and practice

The results of the research questions of this thesis successfully contribute to theory in the following ways:

- The thesis gives frequency of occurrence of the four lexical features in the speaking and writing of both the A2 and the C1 learner at the college. It identifies the degree to which the learners use multi-word verbs, delexical verbs, collocations and idioms. These results inform teachers, syllabus and materials designers in that it shows that it

is not just the higher level group that uses such exponents. Despite stipulations of published research such as the CEFR and IELTS, the corpus analysis clearly shows that the lower level group actually do use the lexical strings.

- The answers to the research questions also contribute to the area of SLA. Students at the college had been living in Ireland for varying lengths of time. In fact, student <$4> had only been in Ireland for four weeks at the time of study. This shows that learners can acquire or 'pick up' such phrases at an early stage of immersion in the language and culture. As we have discuss throughout, the usage-based model seems to best capture this process. We are seeing evidence of learners acquiring items that are frequent in and salient to their everyday lives. For example in the idiomatic expressions chapter, we saw a number of examples of A2 learners using everyday expressions relating to feelings, senses and bodily functions. None of these feature in course book materials or language syllabi and yet they are crucial items for daily life. Learners, even at A2 level, hear them and experience them often enough to work out their syntactic form and their related unitary meaning. What we are seeing in the ACE data is the result of this process. That is, the learners are using the items that they have heard over and over again. They have worked their meaning out and now these items have become entrenched in their repertoires.
- The research has shown that the higher level cohort use more lexical strings than the A2 cohort which contributed to theory by showing what learners at each level can actually do rather than what they should be able to do. It also highlights that the A2 cohort do use the lexical strings. Tellingly, C1 learners have moved to abstraction of abstract meaning in terms of using more opaque items, as discussed above.
- Finally, the comparison with the BNC and EVP shows that students studying and living in Ireland can use these lexical strings at a high rate of frequency. The research

shows that it is not just students living in England that can use such strings. Through the analysis chapters, the EVP and the BNC have proved very useful points of reference because we were able to tell whether items are typically expected by learners and if so at what level. Also we could appraise whether the frequency of use in ACE was in line with the native speaker norm in the BNC. We also drew on the Limerick Corpus of Irish English in Chapter 8 when we wanted to examine usage of items in Irish English. There were some limitations to these comparisons, however, which we discuss below.

9.4 Limitations of this research

As mentioned in Chapter 1, there are a number of limitations to the present study. Firstly, this is a small study focusing on two groups of learners at one specific institution at a specific point in time. In other words, the study is a snapshot of these learners language use over a fixed period in time. The institution at which this research was conducted is an adult education centre located in a city in Ireland. The learners at the college ranged from 18 to 50 years of age, were of both genders and different nationalities. Though both groups analysed in this thesis were in this sense diverse, they only represent a certain type of learner: that is a mature learner only studying English language at the college. The majority of these learners were working full-time and only studying part-time so they are not typical of all learners. However, we argue that the sample offers a wide and rich cohort of adult ESOL learners. We do, however, accept that we cannot assume that we can generalise widely from our findings though they offer interesting insights nonetheless.

The study looked at just two levels of competence. These were the only classes available at the college and, as discussed in chapter 1, learners were streamed into levels based on a simple level test and oral interview. In reality, institutional funding only allowed for two classes so labels such as Elementary / A2 and Advanced / C1 are approximates. Within each of these two levels, there

were some students above and below these levels. We also note that learners were aware of the recordings being conducted and this could be cited as a limitation. However, given that the recordings continued over a whole term, it is argued that overtime, learners became very used to being recorded. Also, the written data provides a second source of data from the cohort which is not inhibited by recording processes. Another limitation of this study is that we do not have any data from learners' use of language outside of class. We have made many claims about learners acquiring language outside of the classroom because we know items were not taught within classroom syllabi, however, we cannot prove this absolutely.

Another limitation that we have mentioned frequently is the issue of opportunity of use. It may be that learners know many multi-word verbs, delexical patterns, collocations or idioms but they just have not had a chance to display or use them. Alternatively, a specific language task may have promoted a high degree of usage of certain items and so on. We can only mitigate this limitation by saying that because we have gathered the data over speaking and writing for a whole term that we are getting a typical sample from two learner cohorts in an ESOL setting.

Finally, we consider the limitations of the BNC and the EVP as baselines in this study. Without doubt, these comparisons brought many findings into relief but they also pointed to some anomalies and insufficiencies that researchers need to consider. Across chapters 5 – 8, the PMW frequencies of occurrences of items in ACE (across the four features under examination) were usually far higher than in BNC. This suggests that the broad representation of the BNC might not always be the most suitable comparison for learner language from the classroom. Alternatively, we could deduce from this that the language of the ESL language classroom represents one type of spoken and written language that is different to English as represented as a whole in the BNC. This point was often corroborated by the fact that items that appeared very frequently in ACE and had a very low occurrences in the BNC (PMWs) very often did appear, even at low levels, in the EVP. This points to the EVP being a useful representation of learner language in general though it too has its limitations as we shall discuss below.

We found, in particular when looking at informal uses of items and colloquialisms that appeared in ACE that they were not listed in the EVP. While many were included, items that an ESL learner frequently hears and uses in informal everyday situations are not in the EVP. Examples of these discussed in Chapter 8 include: *I'm dying for a drink* (A2); *I'm starving* (A2, C1) *I'm exhausted* (A2, C1); *I'm knackered* (C1); *I'm bursting to pee* (A2); *I could eat a horse* (C1). Of course, we don't expect these to appear in the EVP but it points to the EVP being a middle ground generic competency framework. This study brings to light so many rich uses of language, even at Elementary level, that are not noted in a generic framework such as the EVP. The EVP is built with foreign language learners in mind. Based on this study, we would argue that there is need for an ESL adaptation of the EVP.

We would also point to the limitation that the EVP does not capture the uses that learners put their lexical repertoire to. For example, in Chapter 8, as the examples above show, learners, at both levels, frequently used idiomatic expressions in hyperbole. While the EVP will tell us that a learner is expected to use the verb *eat* at A1, it does not tell us anything about its non-literal use and potential discourse functions associated with expressions like *I could eat a horse.*

9.5 Suggestions for Further Research

1) Looking to the future and as suggestions for further research, I discussed in the introduction chapter the fact that there is a lack of corpus research into spoken language and in particular into English as a second language learners and the linguistic features they use. There is somewhat of a gap in the literature, specifically with respect to the lower level learners, owing to this I would suggest a diachronic and thorough study of learners at A1 and A2 level. I would suggest that a corpus built solely on the language used by the lower level cohorts could aid our understanding and provide a more accurate account of the most frequent language features used by lower level learner, across a larger sample and across other learning sites.

2) In addition, it was extremely interesting for me to record and analyse the language and even the attempts made by the A2 students, particularly in respect of the manner in which they dealt with the communicative and interactive elements of the English language classroom. What is more, the noted language development and progression of this cohort over the twelve week course period was interesting to witness. However, it must be reiterated that at the college in question, and at the time of this research, there were only two language levels; A2 and C1. Due to this, some students in the A2 cohort would have been A1 level, and even at times real beginners. As a result, an analysis of a purely A2 cohort would provide more accurate results for this level. Furthermore, as these were two specific groups of learners, an analysis of different groups would yield a more accurate representation of learners in general.
3) In addition, the fact that the course ran for one semester of twelve weeks, with just two three hour class per week, only provides us with a 'snap shot' into the progression of these students. Personally, I feel that it would be interesting to conduct a longer diachronic observation of this cohort. I believe that such a study would provide us with fascinating insights into the second language acquisition process.
4) Moreover, I feel that a more thorough and comparative analysis with the CEFR levels, the English Vocabulary Profile, and the English Grammar Profile would provide accurate results into what these students really do know at the A2 and C1 level. A diachronic analysis would further contribute to the emerging research into language learners' abilities at different levels. As discussed above, this analysis could in turn inform a more ESL-friendly framework or profile.
5) As this present research focuses on the four lexical features: 1) Multi-word Verbs 2) Delexical Verbs 3) Collocations and 4) Idiomatic Expressions, the results provide us with an insight into specific language features. I believe that an analysis of the overall language features used by both cohorts compared with larger native speaker corpora, such as the

BNC, would generate fundamental insights into the differences and similarities of language in use between native and non-native speakers of English.

6) Additionally, comparing the ACE corpus with other varieties, genres and registers of English (formal, informal, ESP EAP etc.) with similar characteristics could expand some of the observations made in this study. The fact that the spoken ACE corpus was recorded during classroom time and the written corpus drawn from essay writing and examinations, the findings represent the language that these learners are using in the classroom. If it were at all possible, it could be interesting to further analyse the language they use in various social contexts; this would produce more data on sociolinguistic structures. A comparison with the ACE corpus could then give rise to the language behaviour of these students in different settings.

Being a language teacher for twelve years (at the time of data gathering) has made the present study even more interesting for me. I always thought that it would be advantageous for me as a teacher, for the learners themselves and for the institution in which I was working, to gain more insights into the second language acquisition process and to witness first-hand the progression of learners as they transverse proficiency levels over time. The intricacies and complexities of spoken and written language were always of great interest to me as a teacher but also as a German as a second language speaker. The possibility and opportunity to empirically analyse the syntactic, semantic and pragmatic use of the language spoken and written by non-native speakers of English intrigued me greatly. Corpus analysis with frequency counts, concordance lines and word lists presented me with the ability to identify the real language that my students were using in the context of the classroom; this in turn has taught me a great deal about the process of second language teaching and learning. Finally, I hope that this present research has contributed some authentic insights into the ever-growing and emerging fields of Corpus Linguistics, Second

Language Acquisition and the Teaching of English to speakers of other languages. To sum up, I hope that other researchers will be motivated by this research project to continue analysing language in use across various genres.

Most of all, it is hoped that this thesis adds to existing work on multi-word lexical items in learner language and that it brings to the fore the need to acknowledge learners' achievements beyond grammar and their ability to continually abstract meaningful units of language and add them to their lexical repertoire.

Bibliography

Aarts, J., (1984). *Corpus Linguistics: Recent Developments in the use of Computer Corpora in English Language Research,* Amsterdam: Rodopi.

Ahn, F., (1864). *A New, Practical and Easy Method of Learning the German Language. First and Second Course in one Volume*, London: Trübner and Nutt.

Aitchison, J., (2001). *Language Change: Progress or Decay*?, Cambridge: Cambridge University Press.

Akimoto, M., (1989). *A Study of Verbo-Nominal Structures in English,* Aoyama Gakuin University: Shinozaki Shorin.

Alderson, A. and North, B., (1991). *Language Testing in the 1990's. The Communicative Legacy*, Regional Language Centre Seminar on Language Testing and Language Programme Evaluation. Singapore: Regional Language Centre.

Alexander, R., (1992). *Fixed Expressions, Idioms and Phraseology in Recent English Learner Dictionaries,* London: Euralex.

Allan, Q., (1998). 'Delexical Verbs and the Degrees of Desemanticization.' *Word-Journal of International Linguistic Association, Vol.* 49, no.1, pp1-17.

Altenberg, B., (1998). 'On the Phraseology of Spoken English: The Evidence of Recurrent Word-Combinations' in Cowie, A., eds *Phraseology,* London: Routledge 101-122.

Altenberg, B., (2001). 'The Grammatical and Lexical Patterning of Make in Native and Non-Native Student Writing*,' Applied Linguistics*, vol. 22, no. 2, pp. 173-194.

Anderson, C., (2007). 'The CEFR and the need for more research.' *Modern Language Journal*, vol. 91, no. 4, pp. 659–663.

Anthony, E., (1975). 'Lexicon and Vocabulary.' *RELC Journal,* vol. 6, no.1, pp. 21-30.

Atkinson, T., (1997). *A Guide to Teaching Languages*. Cheltenham: Mary Glasgow.

Bachman, L., (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

Bahlsen, L., (1903/1905). *The Teaching of Modern Languages,* Columbia: Columbia University.

Bailey, K., (1998). *Learning about Language Assessment: Dilemmas, Decisions and Directions,* London, Heinle and Heinle Publishers.

Barkema, H., (1996). 'Idiomaticity and Terminology: A Multi-Dimensional Descriptive Model.' *Studia Linguistica,* vol. 50, no. 2, pp.125-160.

Barnard, H., (1971). *Advanced English Vocabulary*, Newbury: House Publications.

Barnard, H., (1971). *Teacher's Book for Advanced English Vocabulary*, Massachusetts: Newbury House.

Barnbrook, G., (2007). 'Sinclair on Collocation.' *International Journal of Corpus Linguistics*, vol.12, no.2, pp. 183-199.

Barns, A., and H., Marilyn (2003). *Effective Assessment in MFL*, London: Centre for Information on Language Teaching and Research.

Bauer, L., (1993). 'Word Families*.' International Journal of Lexicography, vol.* 6, no.4, pp. 253-279.

Bazell, C., Catford, J., Halliday, M., and R. Robins., (1966). *In memory of J.R. Firth*, London: Longman.

Beglar, D. (2000). *Estimating Vocabulary Size:* JALT Testing and Evaluation, available: http://jalt.org/test/beg_1.htm [accessed 2nd November 2010].

Bellot, J., (1580). *The Englishe Scholemaister. Containing many profitable perceptions for the natural borne French men, and other strangers that have their French tongue, to attain the true pronouncing of the English tongue*, London: Thomas Purfoote.

Bellot, J., (1586). *Familiar Dialogues, for the Instruction of them, that the desire to learn to speak English, and Perfectly to pronounce the same*: London: Scholar Press.

Benson, M., (1985). *Collocations and Idioms Dictionaries*, *Lexicography and Language Learning*, Oxford: Pergamon.

Benson, M., Benson, E., and Ilson, R., (1997). *The BBI combinatory dictionary of English: a guide to word combinations,* John Benjamins Publishing Co.

Berlitz, M., (1903). *The Berlitz Method for Teaching Modern Languages*, Berlin: English Part.

Biber, D. (1993). 'Representativeness in Corpus Design' *Literacy and Linguistic Computing*, vol. 8, no. 4, pp. 243-257.

Biber, D., Johansson, S., Leech, G., Conrad, S and E. Finegan. (1999). *Longman Grammar of Spoken and Written English,* London: Longman.

Biber, D., Conrad, S. (2009). *Register, Genre and Style.* Cambridge: Cambridge University Press.

Biskup D. (1992) 'L1 Influence on Learners' Renderings of English Collocations: A Polish/German Empirical Study' in: Arnaud, P, and Béjoint H. eds *Vocabulary and Applied Linguistics.* London: MacMillan, 128-146.

Blackie, M., Case, J., and J. Jawitz. (2010). 'Student Centredness: The Link Between Transforming Students and Transforming Ourselves.' *Teaching in Higher Education*, vol. 15, no. 6, pp. 637–646.

Baldwin, T., (2002). *Extracting the Un-extractable: A Case Study on Verb Particles. 6th Conference on Natural Language Learning,* Taiwan: Taipei.

Bloomfield, L., (1933). *Language,* New York: Holt.

Bloomfield, L., (1942). *Outline Guide for the Practical Study of Foreign Languages*, Baltimore, Linguistic Society of America.

Blum, S., (1978). 'Universals of Lexical Simplification.' *Language Learning Journal,* vol. 28, no. 22, pp. 399-415.

Blum, S., (1979). 'Lexical Simplification in Second Language Acquisition in Studies'. *Second Language Acquisition*, vol. 2, no. 2, pp. 43-63.

Bonelli, Tognini, E. (2001). *Corpus Linguistics at Work*, Amsterdam: John Benjamins.

Bonelli, Tognini, E. (2010). *Theoretical Overview of the Evolution of Corpus Linguistics,* The Routledge Handbook of Corpus Linguistics. Abington: Routledge.

Bovee, A., (1919). 'Teaching Vocabulary by the Direct Method.' *The Modern Language Journal*, vol. 4, no. 2, pp. 63-72.

Bright, J, and McGregor, G., (1970) *Teaching English as a Second Language*, New York: Longman.

Brown, H., (2004). *Language Assessment: Principles and Classroom Practices,* New York: Longman.

Bryson, B., (1990). *The Mother Tongue,* New York: Avon.

Buk, G., (2001). *Assessing Listening,* Cambridge: Cambridge University Press.

Busa, R., (1949) *Index Thomisticus*, Milan: Gallarate.

Bullokar, W., (1586). *Pamphlet of Grammar*, London: Henry Denham.

Callies, M., and Zaytseva, E., (2013). 'The Corpus of Academic Learner English (CALE) – A new resource for the study and assessment of advanced language proficiency' in Granger, S.,Gilquin, G. and Meunier, F. eds., *Twenty Years of Learner Corpus Research*: *Looking back, Moving ahead*. *Corpora and Language in Use*. Universitaires de Louvain: Louvain-la-Neuve: Presses, 49-59.

Campion, M., and Elley, W., (1971). *An Academic Vocabulary List.* Wellington, New Zealand Council for Educational Research.

Campbell, R., and Wales, R., (1970). 'The study of language acquisition' in Lyons, J., ed., *New Horizons in Linguistics*. Harmondsworth: Penguin Books, 26-32.

Canale, M., and Swain, M., (1980). 'Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing.' *Applied Linguistics*, vol. 1, no.1, pp. 1-47.

Capel, A., (2009). *A1-B2 Vocabulary: Insights arising from the English profile Wordlists Projects,* English Profile Seminar. C. ESOL. Cambridge: Cambridge ESOL.

Capel, A., (2010). 'A1 – B2 Vocabulary: Insights and issues arising from the English Profile Wordlists project.' *English Grammar Profile Journal,* vol. 1, no. 1, pp. 1–11.

Carroll, J., Davies, P., and Richman, B., (1971). *The American Heritage Word Frequency Book,* New York, American Heritage Publication Company.

Carlsen, C., (2012). 'Proficiency level – a fuzzy variable in computer learner corpora*.' Applied Linguistics*, vol. 3, no. 2, pp. 161–183.

Carlsen, C., (2014). *How valid is the CEFR as a Construct for Language Testing*. Paris: 5th ALTE Conference.

Carter, R., and McCarthy, M., (2001). 'Size isn't Everything: Spoken English Corpus and the Classroom.' *TESOL Quarterly,* vol. 35, no. 2, pp.337-340.

Carter, R., (1987). 'Is there a core vocabulary?' *Applied Linguistics, vol.* 8, no.1, pp. 178-198.

Carter, R., (1987). *Vocabulary*, London: Allen and Unwin.

Carter, R., (1987). 'Vocabulary and Second/Foreign Language Teaching*.' Language Teaching*, vol. 20, no.1, pp. 3-16.

Carter, R., (1987). *Vocabulary: Applied Linguistic Perspectives*. London: Allen and Unwin Ltd.

Carter, R., (1998). 'Reply to Guy Cook.' *ELT Journal, vol.* 52, no.1, pp.64-76

Carter, R., (1998). *Vocabulary,* New York: Routledge.

Carter, R., (2001). 'Vocabulary' in Carter, R., and Nunan, D., eds., *The Cambridge Guide to teaching English to speakers of Other Languages.* Cambridge: Cambridge University Press, 112-125.

Carter, R. and Nunan, D., (2001). *The Cambridge Guide to Teaching English to Speakers of other Languages.* Cambridge: Cambridge University Press.

Carter, R. and McCarthy, M., (1998). *CANCODE Corpus*. Cambridge: Cambridge and Newcastle.

Carter, R. and McCarthy, M., (1988). *Vocabulary and Language Teaching*. London: Longman.

Carter, R., and McCarthy, M. (1997). *Exploring spoken English,* Cambridge: Cambridge University Press.

Carter, R., (2004). *Language and Creativity: The Art of Common Talk*. London: Routledge.

Chomsky, N., (1959). 'A Review of B.F. Skinner's Verbal Behaviour*.' Language*, vol. 35, no.1, pp. 26-58.

Chomsky, N., (1956). 'Three Models for the Description of Language*.' IREC Transactions on Information Theory,* vol. 2, no. 1, pp. 113-124.

Chomsky, N., (1957). *Syntactic Structures*, The Hague: Mouton.

Chomsky, N., (1965). *Aspects of the Theory of Syntax,* Cambridge: MIT Press.

Choudhry, A., (2010). T*eaching Vocabulary in the EFL/ESL Classroom-Central Pedagogical Issues.* Department of English. Assan University: Silcher.

Chunjo, K., (2010). *Measuring Vocabulary Levels of English Textbooks and Tests using a BNC Lemmatised High Frequency Word List.* Retrieved November 2010 from http://www5d.biglobe.ne.jp/~chujo/eng/data/rodopi.pdf.

Clancy, B., (2010). *Hurry up baby son all the boys is finished their breakfast: A socio-pragmatic analysis of Irish settled and Traveller family discourse.* PhD Thesis. Mary Immaculate College, Limerick.

Clapen, E., (1970). *The Identification of non-native Speakers of English likely to under achieve in University Courser through inadequate command of Language English.* Unpublished PhD. Manchester: University of Manchester.

Clark, E., (1973). 'What's in a word? On the child's acquisition of semantics in his first language' in Moore, T., ed., (Ed.), *Cognitive development and the acquisition of language*. New York: Academic Press, 65-110.

Clark, H., (1970). *Word Associations and Linguistic Theory* in Lyons, J. *1970. New Horizons in Linguistics*. London: Pelican.

Clausen, W., (1994). *Virgil: Eclogues*. Oxford: Oxford University Press.

Coady, C., (1997). *Second Language Vocabulary Acquisition: A Rationale for Pedagogy.* Cambridge: Cambridge University Press.

Coady, J., (1995). *Research on ESL/EFL Vocabulary Acquisition: Putting it in Context. Second Language Reading and Vocabulary Learning*. New Jersey: Ablex Publishing Company.

Cobbett, W., (1984). *A Grammar of the English Language in a Series of Letters*. Oxford: Oxford University Press.

Coleman, J., (1996). *Studying languages: a survey of British and European students: the proficiency, background, attitudes and motivations of students of foreign languages in the United Kingdom and Europe.* London: CILT.

Coleman, J., (1966). *The Coleman Report: Equality of Educational Opportunity. Washington*, Johns Hopkins University.

Comenius, J., (1631). *Porta Linguarum Trilinguis Reserata et Aperta (The Gate of tongues unlocked and opened, or else a Seminarie or seed plot of all Tongues and Sciences).* London: George Millar Scholar Press.

Comenius, J., (1657). *Didactica Magna (The Great Didactic)*. Amsterdam: Keatinge.

Conklin, K., and Schmitt, N., (2008). 'Formulaic Sequences: Are they Processed more Quickly than Non-formulaic Language by Native and Non-Native Speakers?' *Applied Linguistics*, vol. 29, no. 1, pp. 72–89.

Corder, S. (1960). *English Language Teaching and Television*. London: Longmans, Green.

Council of Europe., (2001a). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment.* Cambridge: Cambridge University Press.

Coxhead, A., (2000). 'A New Academic Word List.' *TESOL Quarterly, vol.* 34, no.2, pp. 213-238.

Crystal, D., (2003). *English as a global language*. (2nd ed.). Cambridge: Cambridge University Press.

Crystal, D., (2005). *How language works*. London: Penguin Books.

Crystal, D., (1980a). 'Neglected Grammatical Factors in Conversational English' in Greenbaum, S. Leech, G., and Svartvik, J., eds., *Studies in Linguistics*. London: Longman, 25-64.

Cutler, D., (1979). 'The Access and Processing of Idiomatic Expressions*.' Journal of Verbal Learning and Verbal Behaviour,* vol. 18, no.1, pp. 523-534.

Davies, A., (2008). 'Applied Linguistics and the Butterfly Effect*.' TESOL Quarterly, vol.* 42, no.2, pp. 296-298.

De Beauyrande, R., (1984).' Forward to the Basics: Getting down to Grammar*.' College*

Composition and Communication, vol, 35, no. 3, pp. 358-367.

De Cock, S., (1998). *Corpora of learner speech and writing ELT.* International Conference on

Germanic and Baltic. University of Vilnius: Homo Liber.

De Jong, J., (1988). 'Rating Scales and Listening Comprehension. *Australian Review of Applied Linguistics,* vol. 11, no. 2, pp. 73-87.

Deese, J., (1965). *The Structure of Associations in Language and Thought*. Baltimore: The John Hopkins Press.

Duchan, J., (2004). *Framework in Language and Literacy: how they inform practice*. London: Guilford.

Duncan, D., (1731). *A New English Grammar*. London: Scholar Press.

Durrant, M., (2010). 'What can a Corpus tell us about Vocabulary Teaching Materials*?'* in McCarthy, M., and O'Keeffe, A., eds., *The Routledge Handbook of Corpus Linguistics*. Abingdon: Routledge, 112-128.

Eckersley, C., (1933). *A Concise English Grammar for Foreign Students.* London: Longmans Green.

Eckersley, C., (1938-42). *Essential English for Foreign Students.* London: Longmans, Green.

Ellis, N. C., (1995). 'The Psychology of Foreign Language Vocabulary Acquisition: Implications for CALL'. *International Journal of Computer Assisted Language Learning (CALL)*, vol. 8, no. 3, pp.103-128.

Ellis, N. C., (1997). *Vocabulary Acquisition: Word Structure, Collocation, Word Class and Meaning. Vocabulary Description, Acquisition and Pedagogy.* Cambridge: Cambridge University Press.

Ellis, N. C., (2003). 'Constructions, chunking, and connectionism: The emergence of second language structure' in Doughty, C. and Long, M., (eds), *Handbook of Second Language Acquisition*. Oxford: Blackwell, 33-68.

Ellis, N. C. and Larsen-Freeman, D., (2006). 'Language emergence: Implications for applied linguistics—Introduction to the special issue'. *Applied linguistics*, vol. 27, no. 4, 558-589.

Ellis, N .C., Simpson-Vlach, R., Römer, U., O'Donnell, M., & Wulff, S. (2015). 'Learner Corpora and Formulaic Language in Second Language Acquisition'in Granger, S., Gilquin, G., and F. Meunier, F., eds., *Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press, 357-358.

Ellis, R., (1997). *Second Language Acquisition*. Oxford: Oxford University Press.

Ellis, R., (1985). *Understanding Second Language Acquisition.* Oxford, England: Oxford University Press.

Epinosa, S., (2003). *Vocabulary Reviewing Trends in EFL/ESL Instruction and Testing. English as a Second Language.* Odiesa: University of Odiesa Spain.

Erman, B., and Warren, B., (2000). 'The Idiom Principle and the Open Choice Principle.' *Text and Talk*, vol. 20, no.1, pp. 29-62.

Espinosa, S., (2003). 'Reviewing Trends in EFL/ESL Instruction and Testing*.' Odisea*, vol. 4, no.1, pp.97-112.

Ewer, J., (1969). *A Course in Basic Scientific English*. London: Longman.

Farr, F., Murphy, B., and O'Keeffe., A., (2004) *The Limerick Corpus of Irish English.* Cambridge: Cambridge University Press.

Fairclough, N., (2001). *Language and Power* (2nd ed.). London: Longman.

Faucett, L., (1936). *The Oxford English Course. Four Volumes.* London: Oxford University Press.

Fernando, C., (1978). 'Towards a Definition of Idiom Its Nature and Function*.' Studies in Language,* vol. 2, no. 3, pp. 313-343.

Figueras, N., North, B., Takala, S., Verhelst, N., and Van Avermaet, P., (2005). *Relating examinations to the Common European Framework: A manual.* Language Testing, 22, 262–279.

Firth, A., (1996). 'The discursive accomplishment of normality. On lingua franca English and conversation analysis.' *Journal of Pragmatics,* vol. 26, no. 2, pp. 237–259.

Firth, J., (1951). *Modes of Meaning*. Indianapolis: Bobbs-Merrill.

Firth, J. (1957a). *Papers in Linguistics*. London: Oxford University Press.

Fisher, P., and Blachowicz, C., (2005). *Vocabulary Scale: Integrated Vocabulary Instruction. Meeting the needs of Diverse Learners.* North Central Regional Educational Laboratory.

Flood, W., (1952). *An Explaining and Pronouncing Dictionary of Scientific and Technical Words.* London: Longmans, Green.

Florio, J., (1578). *His firste Fruites: which yield familiar speech, merie Prouerbes, wittie Sentences, and golden sayings. Also a prefect Induction to the Italian, and English tongues*. London: Da Capo Press.

Florio, J., (1591). *Second Frutes, to be gathered of twelve trees, of diuers but delightsome tastes to the tongues of Italians and Englishmen*. London: De Capo Press.

Folse, K., (2007). *Implications of Second Language Vocabulary Research for Teachers. Teaching and Learning Vocabulary in another Language*. Dubai: TESOL Arabia.

Francis, W., and Kucera, H., (1982). *Frequency Analysis of English Usage.* Boston: Houghton Mifflin.

French, F., (1949). *The Teaching of English Abroad*. London: Oxford University Press.

Fries, C., (1945). *Teaching and Learning English as a Foreign Language*. Arbor: University of Michigan Press.

Fries, C., (1952). *The Structure of English, an Introduction to the Construction of English Sentences.* New York: Harcort, Brace.

Fries, C., (1945). *American Linguists and the Teaching of English as a Foreign Language.* Arbour: University of Michigan Press.

Frisby, A., (1957). *Teaching English. Notes and Comments on Teaching English Overseas*. London: Longmans, Green.

Fulcher, G., (2004). 'Deluded by artifices? The Common European Framework and Harmonization.', *Language Assessment Quarterly*, vol. 1, no. 4, pp. 253–266.

Gardner, D., and Davies, M., (2007). 'Pointing out Frequent Phrasal Verbs'. *TESOL Quarterly*, vol. 41, no. 2, pp. 339-359.

Gatenby, E., (1944). *English as a Foreign Language: Advice to non-English Teachers.* London: Longmans, Green.

Gibbs, R., (1986). 'Skating on Thin Ice: Literal Meaning and Understanding Idioms in Conversation.' *Discourse Processes*, vol. 9, no.1, pp.17-30.

Gouin, F., (1892). *The Art of Teaching and Studying Languages*. Paris: George Philip.

Gouin, F., (1892). *L'art denseigner et detuder les langues. Fischbecher G Paris.* Translated as: *The art of Teaching and Studying Languages*. London: George Philip and Son.

Gouldne, R., Nation, P., and Read, J., (1990). 'How long can a receptive vocabulary be?' *Applied Linguistics*, vol. 11, no. 49, pp. 341-363.

Granger, S., Gilquin, G. and Meunier, F. (2015) 'Introduction: learner corpus research – past, present and future' in Granger, S., Gilquin, G., and Meunier. F., eds., *The Cambridge Handbook of Learner Corpus Research.* Cambridge: Cambridge University Press, 1-5.

Grant, L., (2004). 'Criteria for re-defining idioms: Are we barking up the wrong tree?' *Applied Linguistics,* vol 25, no. 1, pp. 38-36.

Grant, L., (2005). 'Frequency of Core Idioms in the British National Corpus*.'* Jo*urnal of Corpus Linguistics,* vol*.* 10, no. 4, pp. 429-451.

Graver, B., (1963). *Advanced English Practice*. Oxford: Oxford University Press.

Gunkova, M., (2011). A *Corpus Based Linguistic and pedagogical investigation into Delexical Verbs.* Masaryk: Masaryk University.

Halliday, M., (1966). *Lexis as a Linguistic Level* in Bazell, C. London: Longman.

Halliday, M., (1985). *An Introduction to Functional Grammar.* London: E. Arnold.

Halliday, M., (1993). *Introduction to Functional Grammar*. Cambridge: Cambridge University Press.

Han, Z.-H. and Finneran R., (2013). Re-engaging the interface debate: Strong, weak, none, or all? *International Journal of Applied Linguistics* vol. 24, no. 3, pp. 370–389.

Hans, N., (1951). *New Trends in Education in the Eighteenth Century*. London: Routledge and Kegan Paul.

Harmer, J., (1991). *The Practice of English Language Teaching*. Essex: Longman.

Harmer, J., (2007). *How to teach English*. 2nd edition. London: Pearson Education Ltd.

Harmer, J., (2015). *The Practice of English Language Teaching*. 4th edition. London: Pearson Longman.

Harris, Z., (1951). *Methods in Structural Linguistics*. Chicago: Chicago University Press.

Hart, J. (1551). T*he Opening of the Unreasonable Writing of our English Tongue*. Stockholm: British Library.

Hartley, B., (1978). *Streamline.* Oxford: Oxford University Press.

Hartley, D., (1741). 'Section of the History of Medicine*.' Proceedings of the Royal Society of Medicine,* vol. 34, no. 1, pp.11-18.

Hartwell, P., (1985). 'Grammar, Grammars and the Teaching of Grammar*.' College English,* vol. 47, no.1, pp. 105-127.

Hawks, E., (1987). *Awareness of Language: An Introduction*. Cambridge: Cambridge University Press.

Hawkins, J.,and Filipović, L., (2012). *Criterial Features in L2 English: Specifying the Reference Levels of the Common European Framework*. Cambridge: Cambridge University Press.

Heilman, M. and Eskenazi, M., (1999). *Self-assessment in Vocabulary Tutoring,* available: www.cs.cmu.edu/~mheilman/papers/heilman_selfassessment_v3. [accessed 6th November 2010].

Herbert, A., (1965). *The Structure of Technical English*. London: Longmans Green.

Hickey, R., (2000). *English in Ireland*. Essen: University of Essen.

Hill, J., (2000). 'Revising Priorities: from the grammatical failure to collocational success. Teaching Collocation in Further Developments in the Lexical Approach'. *Language Teaching Publications*, vol 20, no.1, pp. 47-67.

Thanasoulas, D. (2013)**.** *History of Approaches to ELT, available:* http://www2.uni-wuppertal.de/FB4/anglistik/multhaup/Lehrveranstaltungen/introduction_to_elt/history_of_approaches_1.htm. [accessed 29th August 2013].

Hoey, M., (2005). *Lexical Priming: A New Theory of Words and Language.* London: Routledge.

Holyband, C., (1580). *De Pronuntiatione Linguae Gallicae.* London: T. Vautrollier.

Hornby, A., (1947). 'Linguistic Pedagogy.' *ELT Journal*, vol., no. 2, pp. 66-71.

Hornby, A., Gatenby, E and H. Wakefield., (1948). *A Learner's Dictionary of Current English*. London: Oxford University Press.

Hornby, A., (1953). 'Vocabulary Control-History and Principles*.' ELT Journal,* vol. 8, no. 1, pp. 15-21.

House, J., (1999). 'Misunderstanding in Intercultural Communication' in Gnutzmann, E. ed., *Interactions in English as a Lingua Franca and the Myth of Mutual Intelligibility*. Cambridge: Cambridge University Press, 73–89.

Howard-Hill, T., (1979). *British Bibliography and Textual Criticism: A Bibliography*. Oxford: Clarendon Press.

Howatt, A., (2004). *A History of English Language Teaching*. Oxford: Oxford University Press.

Hsu, P., (2011). *On the Effect of Idioms Comprehension of ESL College Students. English Language.* Taiwan: National Tawain Normal University.

Hübner A., and S, Laviosa., (1999). *Assessment and Accreditation for Languages: the Emerging Consensus*. London: London CILT.

Huckin, J., (1997). *Second Language Vocabulary Acquisition-A Rationale for Pedagogy.* Cambridge: Cambridge University Press.

Hume, D., (1748). *An Enquiry Concerning Human Understanding.* London: Jonathan Bennett.

Hulstijn J., Alderson J., and R. Schoonen., (2010). **'**Developmental stages in second-language acquisition and levels of second-language proficiency: are there links between them. Communicative proficiency and linguistic development: intersections between SLA and language testing research'. *European Journal of Oral Sciences*, vol. 1, no. 3, pp.11-20.

Hunston, S., (2002). *Corpora in Applied Linguistics.* Cambridge: Cambridge University Press.

Hymes, D., (1972). 'On Communicative Competence' in: Pride, J and Holmes, J., eds., *Sociolinguistics. Selected Readings*. Harmondsworth: Penguin, 269-293.

Jackendoff, R., (1983). *Semantics and Cognition*. Massachusetts: Massachusetts Institute of Technology.

Jackendoff, R., (1995). *Patterns in the Mind: Language and Human Nature.* New York*:* Basic Books.

Jacobs, H., Zinkgraf, S., Wormuth, D., Hartfiel, V., J. Hughey. (1981). *English Composition Program*. Newbury: House Publishers Inc.

Jenkins, J., (2000). *The Phonology of English as an International Language*. Oxford: Oxford University Press.

Jenkins, J., (2003). *World Englishes.* London: Routledge.

Jespersen, O., (1904). *How to Teach a Foreign Language*. London: Allen and Unwin.

Jenkins, J., (2006a). 'Current perspectives on teaching World Englishes and English as a Lingua Franca.' *TESOL Quarterly*, vol. 40, no.1, pp.10-40.

Jing, J., (1997). *Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy*. Taiwan: .International Conference Research on Computational Linguistics.

Johansson, S. (1998) 'Towards a Multilingual Corpus for Contrastive Analysis' in Borin, L., ed., *Paralleled Copora Paralleled Worlds.* Amsterdam: Rodopi, 12-22.

Johnson, S., (1755). *A Dictionary of the English Dictionary Language, in which the Words are deduced from their Originals and Illustrated in their Different Significations by Examples from the Best Writers.* London: W. Strahan.

Johnson, S., (1755). *A Dictionary of the English Language*. London: Princeton.

Jones, D., (1909). *The Pronunciation of English.* Cambridge: Cambridge University Press.

Jones, D., (1918). *Outline of English Phonetics.* Cambridge: Heffer and Teubner.

Jones, D., (1966). *Paul Passy*. Warwick: University of Warwick.

Jones, P. (1988). Vocabulary size as a Placement Indicator, available: Jones, P., (1988). http://www.lognostics.co.uk/index.htm, [accessed 13th July 2013].

Jonson, B., (1640). *The English Grammar. The Works.* London: Scholar Press.

Judd (1978). 'Vocabulary Teaching and TESOL: A Need for Revaluation of Existing Assumptions.' .*TESOL Quarterly*, vol. 12, no.1, pp.71-76.

Kaeding, W., (1898). *Häufkeitswörterbuch der deutschen Sprache: Festgestellt durch einen Arbeitsschlus der deutschen Stenographiesystem*. Berlin: Steglitz.

Kachru, B. (1992). *The Other Tongue: English across cultures*. IL: University of Illinois Press.

Kamimoto, T., (2010). *An Examination of Nation's 1990 Vocabulary Levels Test, available:* http://www1.harenet.ne.jp/~waring/vocab/colloquium/tad2001.htm. [2nd December 2010].

Kellerman, E., (1986). *An Eye for an Eye: Cross linguistic Constraints on the Development of the Lexicon. Cross linguistic Influence in Second Language Acquisition.* London: Pergman Press Ltd

Kelly, L., (1969). *Twenty-five centuries of language teaching.* Newbury House: Rowley.

Klein, W., (1986). *Second Language Acquisition.* Cambridge: Cambridge University Press.

Kling, L., (1971). *Woodworth and Schosberg's Experimental Psychology.* New York: Holt, Rhinehart and Winston.

Klinghardt, H., (1888). *Ein Jahr Erfahrungen mit der neuen Methode*. Marburg, N.G. Elwertsche Verlag.

Krashen, S., (1983). *The Natural Approach*: *Language Acquisition in the Classroom*. Oxford: Pergamon.

Krashen, S., (1985). *The Input Hypothesis: Issues and Implications.* London: Heinmann.

Krashen, S., (1982). *Principles and Practice in Second Language Acquisition.* New York: Pergamon Press

Krashen, S., (1981). *Second Language Acquisition and Learning.* Oxford: Pergamon Press.

Kroeh, C., (1887). 'Methods of Teaching Modern Languages: Issues and Implications.' *Transactions and Proceedings of the Modern Language Association of America, vol.* 3, no. 1, pp.169-185.

Kuchera, H., (1967). *A Computational Analysis of Present- Day American English Providence.* Rhode Island, Brown University Press.

Labov, W., (1972). *Language in the Inner City.* Philadelphia: University of Pennsylvania Press.

Labov, W., (1981). *The Study of Non Standard English*. IL: National Council of Teachers.

Larsen-Freeman, D., (1987)*.'* From Unity to Diversity: twenty-five years of language teaching methodology.' *Forum*, vol.25, no. 4, pp. 2-10.

Laufer, B., (1986)*.* 'The Lexical Plight in Second Language Reading: Words you don't know, Words you think you know*'. International Review of Applied Linguistics in Language Teaching, vol.* 29, no. 4, pp.11-28.

Laufer, B., (1995). 'Vocabulary Size and Use: Lexical Richness in L2 Written Production.' *Applied Linguistics, vol.* 16, no.3, pp. 307-322.

Laufer, B., (2000). 'Avoidance of idioms in a second language: The effect of L1–L2 degree of similarity.' *Studia Linguistica*, vol. 54, no.2, pp. 186–196.

Leacock, C., and Martin, C., (2001). *Research Reports Automatic assessment of Vocabulary usage without Negative Evidence.* Educational Testing Service Report 67, ETS Research Project: 19.

Leech, G., (1991). T*he State of the Art in Corpus Linguistics.* Aijmer/Altenberg: 8-29.

Leech, G., Rayson, P., and Wilson, A. (2001). *Word Frequencies in Written and Spoken English: Based on the BNC*. London: Longman.

Leech, G., (1992). *Corpora and Theories of Linguistic performance. Directions in Corpus Linguistics:* Proceedings of Nobel Symposium 82. J. Svartvik. Berlin and New York, Mouton de Gryter.
Leon, J., (2005). *Claimed and Unclaimed Sources of Corpus Linguistics* in the Henry Sweet Society Bulletin.

Levinston, E., (1979). 'Second Langua*ge* Acquisition: Issues and Problems*.' Interlanguage Studies Bulletin*, vol. 4, no.2, pp. 147-160.

Lewis, M., (1994). *The Lexical Approach: The State of ELT and a way Forward*. London, Language Teaching Publications.

Liao, Y.,and Fukuya, Y., (2004). 'Avoidance of Phrasal Verbs: The case of Chinese learners of English*.' Language Learning*, vol 54, no.1, pp. 193– 226.

Li Shengxi, L., Ma Huan, Y., (2010). *Second/Foreign Language Teaching Methodologies*. Malyashiya, University of Lehore.

Liontas, J., (1999). *Developing a pragmatic methodology of idiomaticity: The comprehension and interpretation of SL vivid phrasal idioms during reading* (Unpublished doctoral dissertation). The University of Arizona: Tucson.

Liontas, J., (2015). 'Developing Idiomatic Competence in the ESOL Classroom*:* A Pragmatic Account.' *TESOL Journal,* vol. 6, no. 4, pp. 621-658.

Lin, D., (1999). *Automatic Identification of non-compositional Phrases*. ACL Conference 1999. College Park, USA.

Little, D., (1988). *Authentic Texts in Foreign Language*. Authentic Publications: Dublin.

Locke, J., (1690). *An Essay Concerning Human Understanding*. London: Jonathan Bennett.

Locke, J., (1693). *Some Thoughts Concerning Education*. London: Awnsoham and Churchill.

Lord, R., (1974). 'Learning Vocabulary*.' International Review of Applied Linguistics,* vol. 12, no.3, pp. 239-247.

Lorge A., (1949). *The Semantic Count of the 570 Commonest English Words*. New York, Columbia University Press.

Lowth, R., (1762). *A Short Introduction to English Grammar, with Critical Notes.* London: Scholar Press 18.

Lynn, R., (1973). 'Preparing Word Lists: A Suggested Method.' *RELC Journal,* vol. 4, no. 1, pp. 25-32.

Mackin, R., (1964). *English Studies Series*. Oxford: Oxford University Press.

Macro, E., (2003). *Teaching and Learning a Second Language: A Review of Recent Research.* London: Continuum.

Madsen, H., (1983). *Techniques in Testing*. Oxford: Oxford University Press.

Maiguascha, R., (1993). 'Teaching and Learning Vocabulary in a Second Language: Past, Present and Future Directions.' *The Canadian Modern Language Review,* vol. 50, no. 1, pp. 83-110.

Maittaire, M., (1712). *The English Grammar, or an Essay on the Art of Grammar applied to and exemplified in the English Tongue.* London: W.B publications.

Makkai, A., (1972). *Idiom Structure in English*. Paris: The Hague.

Marcel, C., (1853). *Language as a Means of Mental Culture and International Communication; or manual of the Teacher and the Learner of Languages*. London: Chapman and Hall.

Martinez, E., (2010). *History of Vocabulary Teaching*. Granada: Granada University.

Mauranen, A., (2003). 'Academic English as Lingua Franca—A Corpus Approach.' *TESOL Quarterly*, vol. 37, no.1, pp. 513–27.

McArthur, T., (1998). *The English Languages*. Cambridge: Cambridge University Press

McCarthy, M., (2001). 'Discourse', in Carter, R. and Nunan, D., eds., *The Cambridge Guide to Teaching English to Speakers of Other Languages,* Cambridge: Cambridge University Press, 78-56.

McCarthy, M., and O'Keeffe, A., (2010). *The Routledge Handbook of Corpus Linguistics*. New York: Routledge.

McCarthy, M., (1984a). 'A New look at Vocabulary in EFL.' *Applied Linguistics,* vol. 5, no. 1, pp. 12-21.

McCarthy, M., (1984b). *Some Pragmatic Features of Written Text*. University of Birmingham: English Language Research.

McCarthy, M., (1987). *Interactive Lexis: Prominence and Paradigms. Discussing Discourse*. University of Birmingham: English Language Research.

McCarthy, M., (1990). *Vocabulary*. Oxford: Oxford University Press.

McCarthy, M., (1998). *Spoken language and Applied Linguistics.* Cambridge: Cambridge University Press.

McCarthy, M., (2001). *Issues in Applied Linguistics*. Cambridge: Cambridge University Press.

McCarthy, M and O'Dell, F., (2004). *English Phrasal Verbs in Use*. Cambridge: Cambridge University Press.

McCarthy, M., O'Keeffe, A., and S. Walsh., (2010). *The Vocabulary Matrix: Understanding, Learning and Teaching*: Hampshire: Heinle Cengage Learning.

McCarthy, M. and O'Keeffe, A., (2010). *The Routledge Handbook of Corpus Linguistics*. Abingdon: Routledge.

McCarthy, M. MacLean, A., and O'Malley, P., (1985). *Proficiency Plus: Grammar, Lexis*, *Discourse*. Oxford: Basil Blackwell.

McCarthy, M., (2002). 'A New Look at Vocabulary in EFL.' *ELT Journal,* vol. 5, no.1, pp. 12-22.

McCarthy, M., McCarten, J., and Sandiford, H. *Viewpoint Level 2*. Cambridge: Cambridge University Press**.**

McCarthy, M. and O'Keeffe, A., (2014) 'Spoken Grammar'. In Celce-Murcia, M., Brinton, D., and Snow, M., eds., *Teaching English as a Second or Foreign Language* (4th ed.). Boston, MA: National Geographic/Cengage, 271 - 287.

McEnery, T., Brezina, V., Gablasova, D., & Banerjee, J. (2019). Corpus Linguistics, Learner Corpora, and SLA: Employing Technology to Analyze Language Use. *Annual Review of Applied Linguistics,* vo. 3, pp.974-92.

McIntosh, A., (1966). *Patterns and Ranges in Patterns of Language: Papers in General, Descriptive and Applied Linguistics*. London: Longman.

McKay, S., (1980). 'Developing Vocabulary Materials with a Computer Corpus.' *RELC Journal,* vol. 11, no. 2, pp. 77-87.

McKay, S., (1980b). 'Teaching the Syntactic, Semantic and Pragmatic Dimensions of Verbs.' *TESOL Quarterly,* vol. 14, no. 1, pp. 17-26.

Meara, P., (1980). 'Vocabulary Acquisition: A neglected Aspect of Language Learning, Language Teaching and Linguistics.' *Abstracts,* vol. 15, no. 4, pp. 221-246.

Meara, P., (1982). 'Word Associations in a Foreign Language: A Report on the Birkbeck Vocabulary Project.' *Nottingham Linguistic Circular,* vol. 11, no. 2, pp. 29-37.

Meara, P., and Jones, G., (1988). *Vocabulary Size as a Placement Indicator. Applied Linguistics and Society.* London: CILT.

Meara, P., and Jones, G., (1990). *Eurocentres Vocabulary Size Test*. Zurich: Eurocentres.

Meara, P., (1990). *Some Notes on the Eurocentres Vocabulary Tests*. Foreign Language Comprehension and Production.

Meara, P., and Buxton, B. (1987). 'An Alternative to Multiple Choice Vocabulary Tests.' *Language Testing*, vol. 4, no.1, pp. 142-154

Meara, P., (1990). *Some Notes on the Eurocentres Vocabulary Tests*. London: Birkbeck College.

Meara, P., (1990). 'Vocabulary Acquisition. A Neglected Aspect of Language Learning. Language Learning and Linguistics'. *Abstracts*, vol. 13, no. 1, pp. 221-246.

Meara, P., (1997). 'Towards a New Approach to Modelling Vocabulary Acquisition' in Schmitt, N. and McCarthy, M., eds., *Vocabulary Description, Acquisition and Pedagogy*, Cambridge, Cambridge University Press, 86-102.

Meara, P., (2010). *EFL Vocabulary Tests*. Swansea: Lognostics.

Meara, P., (2000). 'Lex30: An Improved Method of Assessing Productive Vocabulary in an L2 System.' *System 28*, vol. 28, no. 1, pp. 19-30.

Mehring, J., (2005). *Developing Vocabulary in Second Language Acquisition: From Theories to the Classroom.* Hawaii: Pacific University.

Melchers, G. and Shaw, P., (2003). *World Englishes: an introduction*. London: Arnold

Melcuk, I., (1995). *Phrasemes in Language and Phraseology in Linguistics. Idioms: Structural and Psychological Perspectives.* Montreal: Mouton.

Miege, G., (1685). *New Method of Learning English.* London: Navigatium Bibliotheca.

Miege, G., (1685). *Nouvelle Methode pour apprende l'Anglois, avec une Nomenclature, Francoise et Angloise; un Recueil d'Expressions familieres; et des Dialogues, familiers et choisis*. London: Scholar Press.

Mifflin, H., (1969). *The American Heritage Dictionary*. Boston: Mifflin Houghton.

Milanovic, M. (2009). 'Cambridge ESOL and the CEFR. University of Cambridge ESOL Examinations' *Research Notes*, available http://www.cambridgeenglish.org/images/23156-research-notes-37.pdf (accessed 6th September 2017).

Mill, J., (1829). *Analysis of the Phenomena of the Human Mind*. London: Baldwin and Cradock.

Milton, J., (1644). *Of Education*. Reprinted in John Milton, 1971. *Selected Prose.* Harmondsworth: Penguin.

Milton, J., (2009). *Measuring Second Language Vocabulary Acquisition*. Bristol: Milton.

Montaigne, M., (1580). *On the Education of Children*. London: John Templeman.

Moon, R., (1996). *Data, Description, and Idioms in Corpus Lexicography*. Malaga: Euralex Conference 1996.

Moon, R., (1998). *Fixed Expressions and Idioms in English: A Corpus Based Approach*. Oxford: Oxford University Press.

Moon, R., (1998). *Fixed Expressions and Idioms in English-A Corpus Based Approach*. New York: Oxford University Press.

Moore, D., (2010). *Why Vocabulary Instruction Matters. Best Practices in Secondary Education*. State Education Research Centre: 4.

Morris, I., (1945). *The Teaching of English as a Second Language: Principles and Methods*. London: Macmillan.

Mukarto, F., Xaverius (2005). 'Assessing the depth of Second Language Vocabulary Knowledge*.' International Forum of Teaching and Studies, vol. 3, no. 3, pp. 40-47.*

Mulcaster, R., (1581). *Positions wherein those primitive circumstances be examined which are necessary for the training up of children*. London: T. Vautrollier.

Murray, L., (1795). *English Grammar, Adapted to the Different Classes of Learners.* York: Scholar Press.

Nakata, T., (2006). 'Implementing Optional Spaced Learning for English Vocabulary Learning: Towards Improvement of the Low-first Method derived from the Reactivation Theory.' *The Jalt CALL Journal,* vol. 2, no. 2, pp. 3-18.

Nation, P., (1983). *Teaching and Learning Vocabulary.* University of Wellington: English Language Institute.

Nation, P., (1983). 'Testing and Teaching Vocabulary*.' Guidelines,* vol*.* 5, no.1, pp. 113-124.

Nation, P., (1990). *Teaching and Learning Vocabulary*. New York: Newbury House Publications.

Nation, P., (2011). 'Research into Practice: Vocabulary.' *The Language Teaching Journal,* vol. 44, no. 4, pp. 529-539.

Nation, P., (1988). *Vocabulary and Reading*. *Vocabulary and Language Teaching*. London: Longman.

Nation, P., and Waring, R. (2010). Vocabulary Size, Text Coverage and Word Lists, available: http://www.fltr.ucl.ac.be/fltr/germ/etan/bibs/vocab/cup.html. [2nd December 2010].

Nation, P., (1997). 'Vocabulary Size, Text Coverage and Word Lists', in *Schmitt, N. and McCarthy, M., eds., Vocabulary: Description*, *Acquisition and Pedagogy*, Cambridge: Cambridge University Press, 6-19.

Nattinger, J., (1992). *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.

Neff van Aertselaer, J., and Bunce, C., (2011). 'The Use of Small Corpora for Tracing the Development of Academic Literacies*'* in Meunier, F, De Cock, S, Gilquin, G. and Paquot, E., eds., *A Taste for Corpora*, vol. 1, no. 45, pp. 63–84. Amsterdam: Benjamins.

Ogden, C., (1930). *Basic English: A General Introduction with Rules and Grammar*. London: Kegan Paul, Trench and Trubner.

O'Keeffe, A., (2010). *Historical perspective: What are corpora and how have they evolved?* The Routledge Handbook of Corpus Linguistics. Abington: Routledge Publishers.

O'Keeffe, A., Farr, F., (2003). 'Using Language Corpora in Initial Teacher Education: Pedagogic Issues and Practical Applications*.' TESOL Quarterly,* vol. 37, no. 3, pp. 389-418.

O'Keeffe, A. and McCarthy, M., (2003). 'What's in a Name?: Vocatives in casual Conversations and Radio Phone-in Calls,' in Leistyna, P. and Meyer, C., eds., *Corpus Analysis Language Structure and Language Use* . Cambridge: Cambridge University Press, 154-185.

O'Keeffe, A., McCarthy, M., and Carter, C., (2007). *From Corpus to Classroom: Language Use and Language Teaching.* Cambridge: Cambridge University Press.

O'Keeffe, A, and Mark, G., (2017). 'The English Grammar Profile of learner competence: Methodology and key findings*.' International Journal of Corpus Linguistics*, vol. 22, no. 4, pp. 457-489.

Oldmixon, J., (1712). *Reflections on Dr. Swift's Letter to the Earl of Oxford, about the English Tongue*. London: Scholar Press.

Oller, J., (1979). *Language Tests at School*. London: Longman.

O'Neill, R., (1979). *Kernel One*. London: Longman Publishing Group.

O'Sullivan, B., (2011). 'Language Testing*'* in Simpson, J., ed., *The Routledge Handbook of Applied Linguistics*. London: Routledge, 76-92.

Orfano, B., (2009). 'The Representation of Spoken language: A Corpus-Based Study of Spoken Sitcom Discourse.' PhD. Mary Immaculate College, University of Limerick. Limerick City.

Organisation, B., (1978). *Berlitz history in Brief. 100 Years of a Language Teaching Revolution.* New York: Berlitz Organisation.

Osbourne, J., (2014). 'Multiple Assessments of Oral Proficiency: Evidence from a Collaborative Platform*'* in Leclercq, P, Edmonds, A. and Hilton, H., eds., *Measuring L2 Proficiency: Perspectives from SLA*, Bristol: Multilingual Matters, 54-70.

Ortega, L., (2013). 'SLA for the 21st century: Disciplinary progress, transdisciplinary relevance, and the Bi/multilingual turn'. *Language Learning,* vol. 63, no. 1, 1-24.

Ovid, P., 20 BC. *The Amores*. Translated: (2011) A.S. Kline.

Oxford, R., (1990). *Language Learning Strategies: What every Teacher Should Know*. New York: Newbury House, Harper and Row.

Oxford, R., (2001). 'Language Learning Strategies' in Carter, R. and Nunan, D., eds., *The Cambridge Guide to Teaching English to Speakers of Other Languages*, Cambridge University Press, 166-173.

Oxford University Press, (1929). *Oversea Education: a Journal of Educational Experiment and Research in Tropical and Sub-Tropical Areas*. For the Secretary of State for the Colonies.

Palmer, H., West, M., and L. Faucett., (1936). *Interim Report on Vocabulary Selection for the Teaching of English as a Foreign Language*. New York 1934 and London 1935, Report of the Carnegie Conference.

Palmer, H., (1944/65). *Curso Internacional de Ingles*. London: Evans Bros.

Palmer, H., (1925/59). *English Through Actions*. Toyko: Longmans, Green.

Paribakht, T., (1996). 'Enhancing Vocabulary Acquisition Through Reading: A Hierarchy of Text-Related Exercise Types.' *The Canadian Modern Language Review,* vol. 52, no.2, pp.155-178.

Pennycook, A.D., (2016). *The Cultural Politics of English as an International Language*. London: Routledge.

Nation, P., (1997). 'Vocabulary Size, Test Coverage and Word Lists' in McCarthy, M. and Schmitt, N., eds., *Vocabulary, Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press, 52-74.

Pawley, A.,(1983). Two Puzzles for Linguistic Theory: Nativelike Selection and Nativelike Fluency.' *Language and Communication.* London: Longman: 191-227.

Pérez-Paredes, P., Mark, G. and O'Keeffe, A., (2018 In Press) 'The impact of usage-based approaches on second language learning and teaching'. *Research Notes*

Pestalozi, J., (1927-1996). *Complete Works: Critical Edition*. Zurich: Orell Füssei Verlag.

Peters, A., (1983). *The Units of Language acquisition*. Cambridge: Cambridge University Press.

Phillip, G., (2004). *Habeaus Corpus: Direct Access, Salience and Delexicalisation in Corpus Based metaphor Studies. Mind Language and Metphor EuroConference on the processing of Metaphor and Metonymy from Computers to Neuropsychology* Granada:

Postman, L., (1970). *Norms of Word Association*. New York: Academic Press.

Prabhu, N., (1987). *Second Language Pedagogy*. Oxford: Oxford University Press.

Prator, C., (1979). *An Outline of Language Teaching Approaches* in *Teaching English as a Second or Foreign Language*. New York: Newbury House.

Prendergast, T., (1864). *The Mastery of Languages, or, the Art of Speaking Foreign Tongues Idiomatically.* London: R. Bentley.

Priestly, J., 1761. *The Rudiments of English Grammar, adapted to the Use of Schools, with Observations on Style*. London: R. Griffiths.

Read, J., (1993). 'The Development of a new measure of L2 Vocabulary Knowledge'. *Language Testing*, vol. 10, no. 3, pp. 355-371.

Read, J., Schmitt, N., and M. McCarthy., (1997). *Vocabulary and Testing. Vocabulary, Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press.

Read, J., (1998). *Validating a Test to measure depth of vocabulary knowledge. Validation in Language Assessment.* A. Kunnan. Mahwah, NJ: Lawrence Erlbaum.

Redman, R., (1986). *Working with Words-A Guide to Teaching and Learning Vocabulary.* Cambridge: Cambridge University Press.

Renouf, A., (1983). *Corpus Development at Birmingham University English Language Research*: University of Birmingham.

Renouf, A., and Banerjee, J., (2007). 'Lexical Repulsion between Sense-Related Pairs.' *International Journal of Corpus Linguistics, vol.* 12, no.3, pp. 415-444.

Renouf, A., (2006). *The Changing face of Corpus Linguistics*. New York: Rodpoi.

Reppen, R., Fitzmaurice, S., Biber, D. (2002). *Using Corpora to Explore Linguistic Variation.* Amsterdam: John Benjaman Publishing Company.

Richards, J., and Rogers, T., (2001). *Approaches and Methods in Language Teaching*. Cambridge: New York, Cambridge University Press.

Richards, I., (1943). *Basic English and its Uses* London: Kegan Paul.

Richards, J., (1976). 'The Role of Vocabulary Teaching.' *TESOL Quarterly*, vol. 10, no.1, pp.77-89.

Richards, J., (1976). 'Teaching English for Science and Technology.' *English for Specific Purposes*, vol.1, no.1, pp. 29-35.

Richards, J., (1990). *The Language Teaching Matrix*. Cambridge: Cambridge University Press.

Richards, J., (2001). *Curriculum Development in Language Teaching*. Cambridge: Cambridge University Press.

Richards, J., (2006). *Communicative Language Teaching Today.* Cambridge: Cambridge University Press.

Richards, J., (2001). *Approaches and Methods in English Language Teaching: a Description and Analysis*. Cambridge: Cambridge University Press.

Richards, J., (2001). *Approaches and Methods in Language Teaching*. Cambridge: Cambridge University Press.

Rignall, J. (2016) *Realist Fiction and the Story Spectator*. London: Routledge.

Rivers, W., (1968). *Teaching Foreign Language Skills*. Chicago: University of Chicago.

Rivers, W., (1972). *The Foreign Language Teacher and Cognitive Psychology or Where do we go from here? Speaking in Many Tongues: Essays in Foreign Language Teaching.* Newbury House: Rowley Mass.

Rivers, W., (1981). 'Apples of Gold in Pictures of Silver: Where have all the words gone?' *Studia Linguistica,* vol. 35, no. 1-2, pp. 114-129.

Rivers, W., (1983). *Learning a Sixth Language: An Adult Learner's Diary in Communicating Naturally in a Second Language*. Cambridge: Cambridge University Press.

Roberts, J., (1975). *The Nature and Function of Porous Concepts*. Carolina: east Carolina University.

Rochet, B. (1984). *Applied Linguistics and Learning and Teaching of Foreign Languages*. London: Edward Arnold.

Roux, P., (2014).' English as an International Language: The Debate Continues.' *Polyclassic*, vol. 26, no. 4. pp. 45-52.

Saffran, R., (2001). 'The use of Predictive Dependencies in Language Learning'. *Journal of Memory and Language,* vol. 44, no.1, pp. 493-515.

Sari, L., (2004). *Assessing Speaking.* Cambridge: Cambridge University Press.

Sauveur, L., (1874a). *Introduction to the Teaching of Living Languages without Grammar or Dictionary.* Boston: Schoenhof and Moeller.

Savignon, S., (1972). 'Teaching for Communicative Competence: A Research Report.' *Audio Visual Language Journal*, vol. 10, no. 3, pp. 129-133.

Schneider, E. W. (2007). *Postcolonial English: Varieties around the world*. Cambridge University Press.

Schmitt, N., (1997). *Vocabulary Learning Strategies* in *Vocabulary Description, Acquisition and Pedagogy*. Schmitt, N. and McCarthy. Cambridge: Cambridge University Press.

Schmitt, N., and McCarthy, M., (1997). *Vocabulary, Description, Acquisition, and Pedagogy*. Cambridge: Cambridge University Press.

Schmitt, N., (2000). *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.

Schmitt, N., Schmitt, D. and Clapham, C., (2001).'Developing and Exploiting the Behaviour of two Versions of the Vocabulary Levels Test.' *Language Testing Journal*, vol. 18, no.1, pp. 24-42.

Schmitt, N. (2004). *Formulaic Sequences: Acquisition, Processing and Use.* Amsterdam: John Benjamins.

Schmitt, N., (2010). *Researching Vocabulary: A Vocabulary Research Manual*. Basingstoke: Palgrave Macmillan.

Schonell, F., Meddleton, I., (1956). *A Study of the Oral Vocabulary of Adults.* Brisbane: University of Queenland Press.

Scottt, J.and Nagey, W., (1997). 'Understanding the Definitions of Unfamiliar Verbs.' *Reading Research Quarterly*, vol. 32, no. 2, pp. 184-200.

Seidenstücker, J., (1811/13). *Elementarbuch Zur Erlernung der Französichen Sprache*. Dortmund: Leipzig.
Seidlhofer, B., (2001). 'Closing a Conceptual Gap: The Case for a Description of English as a Lingua Franca.' *International Journal of Applied Linguistics,* vol. 11, no. pp. 133–58.

Seidlhofer, B., (2004). *Annual Review of Applied Linguistics,* vol. 24, pp. 209–239. Cambridge: Cambridge University Press.

Sharifian, F., (2009). *English as an International Language. Perspectives and Pedagogical Issues.* Canada: Multilingual Matters.

Sherry, R. and Philip C., (2005). *Developing students' transferable skills in the language classroom: a report on the pilot programme of the Transferable Skills i*n *Third Level Modern Languages Curricula project.* Dublin: Transferrable Skills project 2005.

Simmons, A., (2010). 'Motivating Students through Power and Choice.' *English Journal*, vol. 100, no.1, pp. 65-69.

Sinclair, J., and Renouf, A. (1991). 'A Lexical Syllabus for Language Learning in Vocabulary and Language Teaching, in Carter, R., and McCarthy, M., eds., *Vocabulary and Language Teaching*. London: Longman, 28-42.

Sinclair, J., (1974). *English lexical Collocations a Study in Computational Linguistics*. Cahiers de Lexicologie.

Sinclair, J.,(1987). *Collins COBUILD English Language Dictionary.* COBUILD. London: Collins.

Sinclair, J., (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sinclair, J. (1997). Corpus Evidence in Language Description, in Wichmann, A, Fligelstone, T. and Knowles, G., eds., *Teaching and Language Corpora*, London: Longman, 27-39.

Skinner, B., (1953). *Science and Human Behaviour,* The Skinner Foundation.

Spolsky, B., (1995). *Measured Words: The Development of Objective Language Testing.* Oxford: Oxford University Press.

Stapleton, J., (2007). 'Comparing English Vocabulary in a Spoken Learner Corpus with a Native speaker Corpus: Pedagogical Implications arising from an Empirical Study in Japan.' *Language Teaching Research*, vol. 11, no. 3, pp. 1-93.

Stern, H., (1983). *Fundamental Concepts of Language Teaching*. Oxford: Oxford University Press.

Sweet, H., (1890). *A Primer of Phonetics*. Oxford: Clarendon Press.

Sweet, H., (1898). *A New English Grammar: Logical and Historical.* Oxford: Clarendon Press.

Sweet, H., (1964). *The Practical Study of Languages. A Guide for Teachers and Learners*. London: Oxford University Press.

Sweet, H., (1900). *The History of Language*. London: Dent.

Sweet, H., (1908). *The Sounds of English*. Oxford: Clarendon Press.

Swift, J., (1712). *A proposal for Correcting, Improving and Ascertaining the English Tongue*. . London.

Taylor Tricomi, E., (1986). 'Krashen's Second Language Acquisition Theory and the Teaching *of* Edited American English.' *Journal of Basic Writing, vol.* 5, no. 2, pp.12-26.

Terell, T., (1982). 'The Natural Approach to Language Teaching: An Update*.' The Modern Language Journal, vol.* 66, no. 2, pp. 127-254.

Thorndike, E., (1921). *The Teacher's Word Book*. New York: Columbia University.

Thorndike, E., (1932). *A Teacher's Word Book of 30,000 Words*. New York: Teacher's College, Columbia University.

Timmis, I., (2003). Corpora, Classroom and Context: The Place of Spoken Grammar in Language Teaching. PhD Thesis. University of Nottingham. Nottingham.

Timmis, I., (2015). *Corpus Linguistics for ELT: Research and Practice*. New York: Routledge.

Tom, C., (2002). *Review of Paul Nation Learning Vocabulary in another Language.* Canadian Journal of Linguistics.

Tomasello, M., (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tomlinson, B., (2003). *Developing Materials for Language Teaching*. London: Continuum.

Tschirner, E., (2004). 'Breadth of Vocabulary and Advanced English Study an Empirical Investigation.' *Electronic Journal of Foreign Language Teaching*, vol. 1, no. pp. 27-39

Türk, U., (2007). 'Higher Level Learners and Vocabulary Acquisition'. *41st IATEFL Conference*. University of Aberdeen.

Twaddell, F., (1973). 'Vocabulary Expansion in the TESOL Classroom*.' TESOL Quarterly,* vol. 7, no.1, pp. 61-78.

Van Ek, J., (1973). *The Threshold Level in a unit/credit system*. Strasbourg: Council of Europe.

Vietor, W., (1902). *Die Methodik des neusprachlichen Unterrichts*. Leipzig: Teubner.
Cobb, N. (1990) *Vocabulary Levels Test*, available: www.er.uquam.ca/nobel/r21270/level. [1st October 2011].

Wallis, J., (1653). *Grammatica Lingvae Anglicanae*. Oxford: Scholar Press 142 1969.

Warren, C., (2010). *What can a Corpus Tell us about Multi-Word Units?* The Routledge Handbook of Corpus Linguistics. McCarthy, M. and O'Keeffe, A., Abingdon: Routledge.

Webbe, J., (1622). *An Appeal to Truth, in the Controuersie* between Art, and Use*; about the best, and most expedient Course in Languages*. London: George Latham.

Webster, N., (1783). *A Grammatical Institute of the English Language, Part 1.* Hartford: Hudson and Goodwin.

Weir, C., (1990). *Communicative Language Testing*. Hertfordshire: Prentice Hall International Ltd.

Weir, C., (2005). *Language Testing and Validatio*n: *An Evidence-Based Approach.* Basingstoke: Palgrave MacMillan.

West, M., (1914). *Education and Psychology*. London: Longmans, Green.

West, M., (1930). 'Speaking Vocabulary in a Foreign Language.' *Modern Language Journal*, vol. 14, no.1, pp. 509-521.

West, M., (1938). *New Method Composition*. London: Longmans, Green.

West, M., (1953). *A General Service List of English Words, with Semantic Frequencies and a Supplementary Word-List for the Writing of Popular Science and Technology*. London: Longmans' Green.

West, M., and Endiet, J., (1935). *New Method English Dictionary*. London: Longman.

West, M., (1935). *The New Method English Dictionary*. London: Longmans, Green.

Widdowson, H., (1978). *Teaching Language as Communication*. Oxford: Oxford University Press.

Widdowson, H., (2003). *Defining Issues in English Language Teaching*. Oxford: Oxford University Press.

Widgery, W., (1888). *The Teaching of Languages in Schools*. London: D. Nutt.

Wilkins, D., (1972). *The Linguistic and* Situational Content *of the Common Core in a Unit/Credit System.* Strasbourg: Council of Europe.

Wilkins, D., (1972). *Linguistics in Language Teaching*. London: Edward Arnold Publishers Ltd.

Wilkins, D., (1972). *The Linguistic and Situational Content of the Common Core in a Unit/Credit System.* Strasbourg: Council of Europe.

Willis, D., (1990). *The Lexical Syllabus*: *A New Approach to Language Teaching*. London: HarperCollins.

Wolter, J., (2002). *Teaching Literacy Using a Multiple-Linguistic Word-Study Spelling Approach: A Systematic Review.* Utah: Utah University Press.

Wray, A., (1999). 'Formulaic Language in Learners and Native Speakers.' *Language Teaching,* vol. 32, no.4, pp. 213-231.

Xue, G. and Nation, P., (1984). 'A University Word List.' *Language Learning and Communication*, vol. 3, no. 2, pp. 215-229.

Yonge, C., (1856). *The Orations of Marcus Tullius Cicero*.

Youngqi, G., (2003). 'Vocabulary Learning in a Second Language: Person, Task, Context and Strategies.' *TESL-EJ Journal*, vol. 7, no.2, pp.96-110.

Zhang, Q., (2007). *EFL Teacher's Attitudes toward Information and Communication Technologies and Attributing Factors*. Pecking: Pecking University.

Zimmerman, C., (1997). 'Historical Trends in Second Language Pedagogy' in Coady, J. *Second Language Vocabulary Acquisition-A Rationale for Pedagogy*. Cambridge: Cambridge University Press, 162-208.

Zoltan, D., (2003). 'Attitudes Orientating and Motivations in Language Learning.' *A Journal of Research on Language Studies,* vol. 53, no. 1, pp. 3-32.

Zyzik, E., (2009). 'Teaching and Learning Idioms: The Big Picture.' *Clear News, vol.* 13, no. 2, pp. 1-7.

Appendices

Appendix 1 Specific learner outcomes for FETAC A

Component Specification

English as a Second Language

Level 3

3N0870

1. Introduction

The Further Education and Training Awards Council is the single national awarding body in further education and training in Ireland. It is responsible for determining the standards for named awards at levels 1 to 6 on the National Framework of Qualifications. All named awards are devised in line with the National Qualifications Authority of Ireland's determinations and guidelines.

2. The National Framework of Qualifications

The National Framework of Qualifications comprises 10 levels ranging from initial learning (level 1) to the most advanced levels of learning (level 10).

At each level there are one or more award types. An award type is a grouping of awards that share similar features. The National Qualifications Authority of Ireland has determined Award Type Descriptors for each award type. See www.nqai.ie. The Award Type Descriptor identifies the key strands and sub-strands of knowledge, skill and competence for that award type.

3. Guide to Level

Learning outcomes at this level relate to a low volume of practical capability and of knowledge of theory. The outcomes relate to the performance of relatively simple work and may be fairly quickly acquired. Outcomes at this level may also confer a minimum employability for low skilled occupations and include functional literacy and numeracy

Strand	Sub-strand	Nature of learning
Knowledge	Breadth	Knowledge moderately broad in range
	Kind	Mainly concrete in reference and with some comprehension of relationship between knowledge elements
Know How & Skill	Range	Demonstrate a limited range of practical and cognitive skills and tools
	Selectivity	Select from a limited range of varied procedures and apply known solutions to a limited range of predictable problems
Competence	Context	Act within a limited range of contexts
	Role	Act under direction with limited autonomy; function within familiar, homogeneous groups
	Learning to Learn	Learn to learn within a managed environment
	Insight	Assume limited responsibility for consistency of self- understanding and behaviour

Extract from 'Determinations for the Outline National Framework of Qualifications': NQAI (www.nqai.ie)

1. Introduction

The Further Education and Training Awards Council is the single national awarding body in further education and training in Ireland. It is responsible for determining the standards for named awards at levels 1 to 6 on the National Framework of Qualifications. All named awards are devised in line with the National Qualifications Authority of Ireland's determinations and guidelines.

2. The National Framework of Qualifications

The National Framework of Qualifications comprises 10 levels ranging from initial learning (level 1) to the most advanced levels of learning (level 10).

At each level there are one or more award types. An award type is a grouping of awards that share similar features. The National Qualifications Authority of Ireland has determined Award Type Descriptors for each award type. See www.nqai.ie. The Award Type Descriptor identifies the key strands and sub-strands of knowledge, skill and competence for that award type.

3. Guide to Level

Learning outcomes at this level relate to a low volume of practical capability and of knowledge of theory. The outcomes relate to the performance of relatively simple work and may be fairly quickly acquired. Outcomes at this level may also confer a minimum employability for low skilled occupations and include functional literacy and numeracy

Strand	Sub-strand	Nature of learning
Knowledge	Breadth	Knowledge moderately broad in range
	Kind	Mainly concrete in reference and with some comprehension of relationship between knowledge elements
Know How & Skill	Range	Demonstrate a limited range of practical and cognitive skills and tools
	Selectivity	Select from a limited range of varied procedures and apply known solutions to a limited range of predictable problems
Competence	Context	Act within a limited range of contexts
	Role	Act under direction with limited autonomy; function within familiar, homogeneous groups
	Learning to Learn	Learn to learn within a managed environment
	Insight	Assume limited responsibility for consistency of self- understanding and behaviour

Extract from 'Determinations for the Outline National Framework of Qualifications': NQAI (www.nqai.ie)

1. Introduction

The Further Education and Training Awards Council is the single national awarding body in further education and training in Ireland. It is responsible for determining the standards for named awards at levels 1 to 6 on the National Framework of Qualifications. All named awards are devised in line with the National Qualifications Authority of Ireland's determinations and guidelines.

2. The National Framework of Qualifications

The National Framework of Qualifications comprises 10 levels ranging from initial learning (level 1) to the most advanced levels of learning (level 10).

At each level there are one or more award types. An award type is a grouping of awards that share similar features. The National Qualifications Authority of Ireland has determined Award Type Descriptors for each award type. See www.nqai.ie. The Award Type Descriptor identifies the key strands and sub-strands of knowledge, skill and competence for that award type.

3. Guide to Level

Learning outcomes at this level relate to a low volume of practical capability and of knowledge of theory. The outcomes relate to the performance of relatively simple work and may be fairly quickly acquired. Outcomes at this level may also confer a minimum employability for low skilled occupations and include functional literacy and numeracy

Strand	**Sub-strand**	**Nature of learning**
Knowledge	Breadth	Knowledge moderately broad in range
	Kind	Mainly concrete in reference and with some comprehension of relationship between knowledge elements
Know How & Skill	Range	Demonstrate a limited range of practical and cognitive skills and tools
	Selectivity	Select from a limited range of varied procedures and apply known solutions to a limited range of predictable problems
Competence	Context	Act within a limited range of contexts
	Role	Act under direction with limited autonomy; function within familiar, homogeneous groups
	Learning to Learn	Learn to learn within a managed environment
	Insight	Assume limited responsibility for consistency of self- understanding and behaviour

Extract from 'Determinations for the Outline National Framework of Qualifications': NQAI (www.nqai.ie)

1. Introduction

The Further Education and Training Awards Council is the single national awarding body in further education and training in Ireland. It is responsible for determining the standards for named awards at levels 1 to 6 on the National Framework of Qualifications. All named awards are devised in line with the National Qualifications Authority of Ireland's determinations and guidelines.

2. The National Framework of Qualifications

The National Framework of Qualifications comprises 10 levels ranging from initial learning (level 1) to the most advanced levels of learning (level 10).

At each level there are one or more award types. An award type is a grouping of awards that share similar features. The National Qualifications Authority of Ireland has determined Award Type Descriptors for each award type. See www.nqai.ie. The Award Type Descriptor identifies the key strands and sub-strands of knowledge, skill and competence for that award type.

3. Guide to Level

Learning outcomes at this level relate to a low volume of practical capability and of knowledge of theory. The outcomes relate to the performance of relatively simple work and may be fairly quickly acquired. Outcomes at this level may also confer a minimum employability for low skilled occupations and include functional literacy and numeracy

Strand	Sub-strand	Nature of learning
Knowledge	Breadth	Knowledge moderately broad in range
	Kind	Mainly concrete in reference and with some comprehension of relationship between knowledge elements
Know How & Skill	Range	Demonstrate a limited range of practical and cognitive skills and tools
	Selectivity	Select from a limited range of varied procedures and apply known solutions to a limited range of predictable problems
Competence	Context	Act within a limited range of contexts
	Role	Act under direction with limited autonomy; function within familiar, homogeneous groups
	Learning to Learn	Learn to learn within a managed environment
	Insight	Assume limited responsibility for consistency of self- understanding and behaviour

Extract from 'Determinations for the Outline National Framework of Qualifications': NQAI (www.nqai.ie)

1. Introduction

The Further Education and Training Awards Council is the single national awarding body in further education and training in Ireland. It is responsible for determining the standards for named awards at levels 1 to 6 on the National Framework of Qualifications. All named awards are devised in line with the National Qualifications Authority of Ireland's determinations and guidelines.

2. The National Framework of Qualifications

The National Framework of Qualifications comprises 10 levels ranging from initial learning (level 1) to the most advanced levels of learning (level 10).

At each level there are one or more award types. An award type is a grouping of awards that share similar features. The National Qualifications Authority of Ireland has determined Award Type Descriptors for each award type. See www.nqai.ie. The Award Type Descriptor identifies the key strands and sub-strands of knowledge, skill and competence for that award type.

3. Guide to Level

Learning outcomes at this level relate to a low volume of practical capability and of knowledge of theory. The outcomes relate to the performance of relatively simple work and may be fairly quickly acquired. Outcomes at this level may also confer a minimum employability for low skilled occupations and include functional literacy and numeracy

Strand	**Sub-strand**	**Nature of learning**
Knowledge	Breadth	Knowledge moderately broad in range
	Kind	Mainly concrete in reference and with some comprehension of relationship between knowledge elements
Know How & Skill	Range	Demonstrate a limited range of practical and cognitive skills and tools
	Selectivity	Select from a limited range of varied procedures and apply known solutions to a limited range of predictable problems
Competence	Context	Act within a limited range of contexts
	Role	Act under direction with limited autonomy; function within familiar, homogeneous groups
	Learning to Learn	Learn to learn within a managed environment
	Insight	Assume limited responsibility for consistency of self- understanding and behaviour

Extract from 'Determinations for the Outline National Framework of Qualifications': NQAI (www.nqai.ie)

**The Further Education and Training Awards Council (FETAC)
was set up as a statutory body on 11 June 2001
by the Minister for Education and Science.
Under the Qualifications (Education & Training) Act, 1999,
FETAC now has responsibility for making awards
previously made by NCVA.**

FETAC
Further Education and Training Awards Council
Comhairle na nDámhachtainí Breisoideachais agus Oiliúna

Module Descriptor

Language

Level 4

September 2001

www.fetac.ie

Level 4 Module Descriptor

Summary of Contents

Introduction	Describes how the module functions as part of the national vocational certificate framework.
Module Title	Indicates the module content. This title appears on the learner's certificate. It can be used to download the module from the website www.fetac.ie.
Module Code	An individual code is assigned to each module; a letter at the beginning denotes a vocational or general studies area under which the module is grouped and the first digit denotes its level within the national vocational certificate framework.
Level	Indicates where the module is placed in the national vocational certificate framework, from Level 3 to Level 6.
Credit Value	Denotes the amount of credit that a learner accumulates on achievement of the module.
Purpose	Describes in summary what the learner will achieve on successfully completing the module and in what learning and vocational contexts the module has been developed. Where relevant, it lists what certification will be awarded by other certification agencies.
Preferred Entry Level	Recommends the level of previous achievement or experience of the learner.
Special Requirements	Usually 'none' but in some cases detail is provided here of specific learner or course provider requirements. There may also be reference to the minimum safety or skill requirements that learners must achieve prior to assessment.
General Aims	Describe in 3-5 statements the broad skills and knowledge learners will have achieved on successful completion of the module.
Units	Structure the learning outcomes; there may be no units.
Specific Learning Outcomes	Describe in specific terms the knowledge and skills that learners will have achieved on successful completion of the module.
Portfolio of Assessment	Provides details on how the learning outcomes are to be assessed.
Grading	Provides details of the grading system used.
Individual Candidate Marking Sheets	List the assessment criteria for each assessment technique and the marking system.
Module Results Summary Sheet	Records the marks for each candidate in each assessment technique and in total. It is an important record for centres of their candidate's achievements.
Appendices	Can include approval forms for national governing bodies.
Glossary of Assessment Techniques	Explains the types of assessment techniques used to assess standards.
Assessment Principles	Describes the assessment principles that underpin FETAC approach to assessment.

Level 4 Module Descriptor

Summary of Contents

Introduction	Describes how the module functions as part of the national vocational certificate framework.
Module Title	Indicates the module content. This title appears on the learner's certificate. It can be used to download the module from the website www.fetac.ie.
Module Code	An individual code is assigned to each module; a letter at the beginning denotes a vocational or general studies area under which the module is grouped and the first digit denotes its level within the national vocational certificate framework.
Level	Indicates where the module is placed in the national vocational certificate framework, from Level 3 to Level 6.
Credit Value	Denotes the amount of credit that a learner accumulates on achievement of the module.
Purpose	Describes in summary what the learner will achieve on successfully completing the module and in what learning and vocational contexts the module has been developed. Where relevant, it lists what certification will be awarded by other certification agencies.
Preferred Entry Level	Recommends the level of previous achievement or experience of the learner.
Special Requirements	Usually 'none' but in some cases detail is provided here of specific learner or course provider requirements. There may also be reference to the minimum safety or skill requirements that learners must achieve prior to assessment.
General Aims	Describe in 3-5 statements the broad skills and knowledge learners will have achieved on successful completion of the module.
Units	Structure the learning outcomes; there may be no units.
Specific Learning Outcomes	Describe in specific terms the knowledge and skills that learners will have achieved on successful completion of the module.
Portfolio of Assessment	Provides details on how the learning outcomes are to be assessed.
Grading	Provides details of the grading system used.
Individual Candidate Marking Sheets	List the assessment criteria for each assessment technique and the marking system.
Module Results Summary Sheet	Records the marks for each candidate in each assessment technique and in total. It is an important record for centres of their candidate's achievements.
Appendices	Can include approval forms for national governing bodies.
Glossary of Assessment Techniques	Explains the types of assessment techniques used to assess standards.
Assessment Principles	Describes the assessment principles that underpin FETAC approach to assessment.

Introduction

A module is a statement of the standards to be achieved to gain an FETAC award. Candidates are assessed to establish whether they have achieved the required standards. Credit is awarded for each module successfully completed.

The standards in a module are expressed principally in terms of specific learning outcomes, i.e. what the learner will be able to do on successful completion of the module. The other elements of the module - the purpose, general aims, assessment details and assessment criteria - combine with the learning outcomes to state the standards in a holistic way.

While FETAC is responsible for setting the standards for certification in partnership with course providers and industry, it is the course providers who are responsible for the design of the learning programmes. The duration, content and delivery of learning programmes should be appropriate to the learners' needs and interests, and should enable the learners to reach the standard as described in the modules. Modules may be delivered alone or integrated with other modules.

The development of learners' **core skills** is a key objective of vocational education and training. The opportunity to develop these skills may arise through a single module or a range of modules. The core skills include:

- taking initiative
- taking responsibility for one's own learning and progress
- problem solving
- applying theoretical knowledge in practical contexts
- being numerate and literate
- having information and communication technology skills
- sourcing and organising information effectively
- listening effectively
- communicating orally and in writing
- working effectively in group situations
- understanding health and safety issues
- reflecting on and evaluating quality of own learning and achievement.

Course providers are encouraged to design programmes which enable learners to develop core skills.

1	**Module Title**	**Language**	
2	**Module Code**	**French**	**E10106**
		German	**E10107**
		Spanish	**E10108**
		Italian	**E10109**
		Irish	**E10110**
		Russian	**E10111**
		Other languages	**contact FETAC**
3	**Level**	**4**	
4	**Credit Value**	**1 credit**	

5 Purpose

This module is a statement of the standards to be achieved to gain an FETAC credit in Language at Level 4. It is designed to be taken across a wide range of FETAC certificates.

This module is designed to give beginners or near-beginners a basic level of language competence.

The module may be included in a wide range of vocational programmes. It is intended that language skills be developed through broad vocationally related activities. The module may be delivered in conjunction with Cultural Studies G10036.

The emphasis is on building the learners' confidence through integrating skills of listening, speaking, reading and writing.

6 Preferred Entry Level

Level 3 Certificate, Junior Certificate or equivalent qualifications and/or relevant life and work experiences.

7 Special Requirements

None.

8 General Aims

Learners who successfully complete this module will:

8.1 develop the ability to use the language effectively for practical communication

8.2	acquire the basic language skills, knowledge and attitudes necessary for advancement to further study or into employment
8.3	gain an awareness of the culture and civilisation of countries where the target language is spoken.

9 Units

The specific learning outcomes are grouped into 4 units.

Unit 1	**Interpersonal Communication**
Unit 2	**Travel and Accommodation**
Unit 3	**Food and Drink**
Unit 4	**Shopping and Making Purchases**

10 Specific Learning Outcomes

Unit 1 **Interpersonal Communication**

Learners should be able to:

10.1.1 enquire about and express the following:
- notions of time, including point of time, length of time, age, calendar, reference to present
- notions of quality, including colour, comparison, similarity
- notions of quantity, including numerals, amounts, sizes, weights and measures
- notions of space, including location, relative position, distance, motion, direction

10.1.2 give and seek information

10.1.3 communicate in a social setting by greeting, attracting attention, taking leave

10.1.4 express and find out attitudes such as satisfaction, dissatisfaction, likes, dislikes, preferences, regret, apology

10.1.5 get things done by requesting, accepting, asking for assistance

10.1.6 repair a breakdown in communication by signalling non-understanding, asking for repetition

10.1.7 state and spell their names, address, telephone numbers, email addresses and those of others

10.1.8	state age and make comparative comments about age, such as older than, younger than
10.1.9	state nationality, date of birth, marital status, occupation
10.1.10	give information about family, interests and elicit similar information from others
10.1.11	describe a loss or theft
10.1.12	describe missing items
10.1.13	give and request information about daily routine.

Unit 2 **Travel and Accommodation**

Learners should be able to:

10.2.1	ask and pay for a ticket using public transport
10.2.2	request and understand information regarding time of departure and arrival
10.2.3	use a ticket vending machine
10.2.4	book and pay for accommodation
10.2.5	make and answer enquiries about accommodation and leisure facilities
10.2.6	ask for and understand directions.

Unit 3 **Food and Drink**

Learners should be able to:

10.3.1	order and pay for food and drink in a variety of settings
10.3.2	understand a menu, which course etc
10.3.3	recognise basic ingredients such as types of meat, vegetables and dairy products
10.3.4	know the typical food and drink of the target country
10.3.5	enquire about and tell meal times in the target country.

Unit 4	**Shopping and Making Purchases**	
	Learners should be able to:	
10.4.1	know the different types of shops and the various departments within a department store	
10.4.2	ask about the location and opening times of shops and department stores	
10.4.3	ask and pay for items of clothing/gifts/souvenirs/food/drink etc.	
10.4.4	ask about the size, quantity, colour, price	
10.4.5	say something is suitable/unsuitable/too big/small/expensive	
10.4.6	understand the currency and rate of exchange	
10.4.7	enquire about postal rates	
10.4.8	ask and pay for stamps/telephone card in a post office/shop	
10.4.9	follow operating instructions when using a public telephone, vending machine, cyberpost	
10.4.10	read signs, notices, advertisements, articles, brochures, telephone book/directory.	

11 Portfolio of Assessment

Please refer to the glossary of assessment techniques and the note on assessment principles at the end of this module descriptor.
All assessment is carried out in accordance with FETAC regulations.
Assessment is devised by the internal assessor, with external moderation by FETAC.

Summary	**Collection of Work**	**40%**
	Examination (Aural)	**30%**
	Skills Demonstration	**30%**

11.1 Collection of Work

The internal assessor will devise guidelines for candidates on gathering a collection of work that demonstrates evidence of a range of specific learning outcomes.

The collection will include a minimum of four items. Three reading items and one writing item must be included.

Reading

Candidates will be required to extract essential, factual information from a variety of authentic and semi-authentic texts. Items could include:

- letter
- simple operational instructions/signs/notices
- short passage from a magazine/newspaper article
- menu
- advertisement for café/restaurant
- brochure

Writing

Candidates will be required to write at least one of the following:

- about themselves
- a short passage on a specific topic
- short emails/faxes/postcards.
- fill in a simple form

All instructions and questions to candidates must be in English/ mother tongue.

Evidence of reading: candidates must answer questions in English/mother tongue.

Evidence of writing: candidates must write in the target language, and text may be handwritten or word processed.

Dictionaries may be used.

11.2 Examination

The internal assessor will devise an aural examination that tests candidates' listening and interpretation skills.

The examination will be based on a range of specific learning outcomes and will be 30 minutes in duration.

The format of the examination will be as follows:

Candidates are required to answer 5 questions (6 marks each).

Candidates will listen to each passage **three times**. Candidates should be given sufficient time to read through the questions before the tape/CD begins.

All instructions and questions must be in English/mother tongue.

Candidates must answer in English/mother tongue.

Dictionaries may not be used.

11.3 Skills Demonstration

In one or more skills demonstrations, candidates will be assessed in oral skills in the target language. The skills may be assessed during the course of the module or in test conditions at a specified time.

Candidates will be required to:

- talk about and respond to questions on themselves, family and interests
- participate in a dialogue set in a social/private domain eg shopping, asking/receiving directions, eating out etc.

The interview and the dialogue must each be at least 5 minutes in duration.

Evidence presented will include

- tape (audio or video)
- documents/list of questions/role play scenarios.

All instructions to candidates must be in English/mother tongue.

The dialogue, all verbal questions and responses must be in the target language.

The candidate may be given 10 minutes to prepare for the dialogue.

Dictionaries may not be used.

12 Grading

Pass	50 - 64%
Merit	65 - 79%
Distinction	80 - 100%

Individual Candidate Marking Sheet 1		Language Level 4 Collection of Work 40% Examination (Aural) 30%

Language (see page 1 for details): _____ **Code:** _____

Candidate Name: _____ **PPSN.:** _____

Centre: _____ **Centre No.:** _____

Assessment Criteria		Maximum Mark	Candidate Mark
Collection of Work **Reading** • relevant information extracted • information understood • context of text understood • a variety of different texts dealt with effectively **Writing** • texts grammatically correct • vocabulary correct • texts convey meaning • texts contain essential information		**30** **10**	
	Subtotal	**40**	
Examination 5 questions (6 marks each)			
Question No.:	4	**6**	
	5	**6**	
	6	**6**	
	4	**6**	
	5	**6**	
	Subtotal	**30**	
	TOTAL MARKS *This mark should be transferred to the Module Results Summary Sheet*	**70**	

Internal Assessor's Signature: _____ ***Date:*** _____

External Authenticator's Signature: _____ ***Date:*** _____

Individual Candidate Marking Sheet 2		Language Level 4 Skills Demonstration 30%

Language (see page 1 for details): _____ **Code:** _____

Candidate Name: _____ **PPSN.:** _____

Centre: _____ **Centre No.:** _____

Assessment Criteria	Maximum Mark	Candidate Mark
Interview		
Communication		
• questions understood and responded to appropriately	**5**	
Fluency		
• speech rhythm appropriate, diction and pronunciation clear	**5**	
Language		
• grammatical structures used effectively, vocabulary used correctly	**5**	
Dialogue		
Communication		
• dialogue completed, messages understood and responded to appropriately	**5**	
Fluency		
• speech rhythm appropriate, diction and pronunciation clear	**5**	
Language		
• grammatical structures used effectively, vocabulary used correctly	**5**	
TOTAL MARKS *This mark should be transferred to the Module Results Summary Sheet*	**30**	

Internal Assessor's Signature: _____ ***Date:*** _____

External Authenticator's Signature: _____ ***Date:*** _____

Glossary of Assessment Techniques

Assignment

An exercise carried out in response to a brief with specific guidelines and usually of short duration.

Each assignment is based on a brief provided by the internal assessor. The brief includes specific guidelines for candidates. The assignment is carried out over a period of time specified by the internal assessor.

Assignments may be specified as an oral presentation, case study, observations, or have a detailed title such as audition piece, health fitness plan or vocational area profile.

Collection of Work

A collection and/or selection of pieces of work produced by candidates over a period of time that demonstrates the mastery of skills.

Using guidelines provided by the internal assessor, candidates compile a collection of their own work. The collection of work demonstrates evidence of a range of specific learning outcomes or skills. The evidence may be produced in a range of conditions, such as in the learning environment, in a role play exercise, or in real-life/work situations.

This body of work may be self-generated rather than carried out in response to a specific assignment eg art work, engineering work etc.

Examination

A means of assessing a candidate's ability to recall and apply skills, knowledge and understanding within a set period of time (time constrained) and under clearly specified conditions.

Examinations may be:

- practical, assessing the mastery of specified practical skills demonstrated in a set period of time under restricted conditions
- oral, testing ability to speak effectively in the vernacular or other languages
- interview-style, assessing learning through verbal questioning, on one-to-one/group basis
- aural, testing listening and interpretation skills
- theory-based, assessing the candidate's ability to recall and apply theory, requiring responses to a range of question types, such as objective, short answer, structured, essay. These questions may be answered in different media such as in writing, orally etc.

Learner Record

A self-reported record by an individual, in which he/she describes specific learning experiences, activities, responses, skills acquired.

Candidates compile a personal logbook/journal/diary/daily diary/ record/laboratory notebook/sketch book.
The logbook/journal/diary/daily diary/record/laboratory notebook/sketch book should cover specified aspects of the learner's experience.

Project

A substantial individual or group response to a brief with guidelines, usually carried out over a period of time.

Projects may involve:

research – requiring individual/group investigation of a topic
process – eg design, performance, production of an artefact/event

Projects will be based on a brief provided by the internal assessor or negotiated by the candidate with the internal assessor. The brief will include broad guidelines for the candidate. The work will be carried out over a specified period of time.

Projects may be undertaken as a group or collaborative project, however the individual contribution of each candidate must be clearly identified.

The project will enable the candidate to demonstrate: (*some of these – about 2-4*)

- understanding and application of concepts in (specify area)
- use/selection of relevant research/survey techniques, sources of information, referencing, bibliography
- ability to analyse, evaluate, draw conclusions, make recommendations
- understanding of process/planning implementation and review skills/ planning and time management skills
- ability to implement/produce/make/construct/perform
- mastery of tools and techniques
- design/creativity/problem-solving/evaluation skills
- presentation/display skills
- team working/co-operation/participation skills.

Skills Demonstration

Assessment of mastery of specified practical, organisational and/or interpersonal skills.

These skills are assessed at any time throughout the learning process by the internal assessor/another qualified person in the centre for whom the candidate undertakes relevant tasks.

The skills may be demonstrated in a range of conditions, such as in the learning environment, in a role-play exercise, or in a real-life/work situations.

The candidate may submit a written report/supporting documentation as part of the assessment.

Examples of skills: laboratory skills, computer skills, coaching skills, interpersonal skills.

FETAC Assessment Principles

1 Assessment is regarded as an integral part of the learning process.

2 All FETAC assessment is criterion referenced. Each assessment technique has **assessment criteria** which detail the range of marks to be awarded for specific standards of knowledge, skills and competence demonstrated by candidates.

3 The mode of assessment is generally local i.e. the assessment techniques are devised and implemented by internal assessors in centres.

4 Assessment techniques in FETAC modules are valid in that they test a range of appropriate learning outcomes.

5 The reliability of assessment techniques is facilitated by providing support for assessors.

6 Arising from an extensive consultation process, each FETAC module describes what is considered to be an optimum approach to assessment. When the necessary procedures are in place, it will be possible for assessors to use other forms of assessment, provided they are demonstrated to be valid and reliable.

7 To enable all learners to demonstrate that they have reached the required standard, candidate evidence may be submitted in written, oral, visual, multimedia or other format as appropriate to the learning outcomes.

8 Assessment of a number of modules may be integrated, provided the separate criteria for each module are met.

9 Group or team work may form part of the assessment of a module, provided each candidate's achievement is separately assessed.

Appendix 3 Metadata for all participants

Speaker	Cohort	Spoken word count	Written word count	Age	Gender	First Language	Nationality	Occupation	Hobbies	Years studying English
<$1>	A2	1,065	548	22	Male	Polish	Polish	Retail Assistant	Reading, listening to music and socialising.	8 months
<$2>	C1	10,022	958	25	Female	Latvian	Latvian	Librarian	Swimming, running and going to the gym.	3 years
<$3>	C1	14,528	1020	23	Male	Polish	Polish	Barman	Translating and writing	3 years
<$4>	A2	4,012	697	30	Female	Italian	Italian	Unwaged	Listening to music, travelling.	8 weeks

<$5>	A2	3,426	867	45	Male	Portuguese	Brazilian	Unwaged	Travelling and learning languages.	1 year
<$7>	C1	16,245	1158	26	Male	Portuguese	Brazilian	Teacher	Extreme sports and hiking.	4 years
<$8>	C1	12,458	959	50	Female	French	French	Retail Assistant	Playing and watching football and socialising.	6 years
<$9>	C1	18,879	1059	25	Female	German	German	Student	Cooking and eating out.	5 years
<$10>	A2	3,425	717	19	Male	Filipino	Filipino	Unwaged	Travelling and learning languages.	10 weeks
<$11>	C1	13,246	1118	32	Female	Spanish	Spanish	Au Pair	Travelling, listening to	6 years

									music and going to live concerts.	
<$12>	C1	12,645	858	45	Male	German	German	Secretary	Reading and watching horror movies.	3 years
<$13>	A2	5,014	718	24	Female	Italian	Italian	Student	Shopping and socialising.	1 year
<$14>	A2	3,658	958	52	Male	Spanish	Chilean	Teacher	Travelling and blogging.	6 months
<$15>	C1	14,424	964	34	Female	Spanish	Mexican	Engineer	Swimming and horse riding.	2 years
<$16>	A2	5,467	808	45	Female	Portuguese	Brazilian	Engineer	Travelling.	1 year

<$17>	C1	17,269	858	30	Male	Russian	Russian	Lawyer	Watching TV, going to the cinema and eating out.	6 years
<$18>	C1	15,625	960	25	Female	Polish	Polish	Journalist	Writing, reading and researching.	4 years
<$19>	C1	11,253	913	22	Male	Spanish	Spanish	Teacher	Travelling and learning languages.	5 years
<$20>	A2	6,865	660	25	Female	French	French	Au Pair	Socialising and technology.	5 months
<$21>	C1	17,262	814	52	Female	Swahili	African	Dentist	Keep fit and hairdressing.	6 years

<$22>	C1	12,206	758	50	Male	French	Mongo	Teacher	Bible group and writing.	10 years
<$23>	A2	6,258	670	20	Female	Mandarin	Chinese	Au Pair	Travelling and make up.	1 years
<$24>	C1	16,259	960	30	Male	Polish	Polish	Student	Socialising, eating out and football.	3 years

Appendix 4: Letter of Consent

Mary Immaculate College

Information and Consent Form

Information Section:

The title of this PhD research project is 'take him to the cleaners and make him do your homework': a corpus-based analysis of lexical structures used by English language learners

1. Participants will be assessed on performance of writing and speaking in relation to the framework.
2. Samples of oral and written work will be taken during the extra class: every Tuesday at 6pm for 12 weeks.
3. They will be required to complete Nation's vocabulary levels test at both the beginning and end of the project.
4. Completing the test should take about 30 minutes and the research will take place at Limerick College of Further Education.

5. I hope that students and teachers will benefit from this project. Students should improve their English vocabulary, oral and written performance. With the compilation of the corpus teachers will be able to see, on what level of the framework students lie, and this will help with future streaming of learners.
6. All students have the right to withdraw from this research at any time.
7. Researcher: Justin McNamara Department of Language and Linguistics, Mary Immaculate College Limerick. Phone: 0851290574. Email: Justin.McNamara@mic.ul.ie.
8. Supervisor: Dr. Anne O'Keeffe Mary Immaculate College Limerick. Email: Anne.okeeffe@mic.ul.ie

Your Contribution

As a participant you will be asked to:

1) Attend the extra class of a Tuesday at 6pm at Limerick College of Further Education room M38.
2) Complete a background questionnaire (profile).

3) Complete Nation's online test twice (one at the beginning of research and the second at the end).
4) Take part in oral presentations, conversations and role-plays which will be recorded.
5) Complete a self -assessment grid based on the can dos of the framework.
6) Complete a questionnaire at the end of the study.

Consent Section:

I, the undersigned, declare that I am willing to take part in this research project which is part of course assessment for the PhD in Linguistics. The title of the study **is '**take him to the cleaners and make him do your homework': a corpus-based analysis of lexical structures used by English language learners/

- I declare that I have been fully briefed on the nature of this study and my role in it and have been given the opportunity to ask questions before agreeing to participate.
- The nature of my participation has been explained to me and I have full knowledge of how the information collected will be used.
- I am also aware that my participation in this study will be recorded and I agree to this. However, should I feel uncomfortable at any time I can request that the recording equipment be switched off. I am entitled to copies of recordings made and am fully informed as to what will happen to these recordings once the study is completed.

- I fully understand that there is no obligation on me to participate in this study and that I am free to withdraw my participation at any time without having to explain or give a reason.
- I am also entitled to full confidentiality in terms of my participation and personal details. Anonymity is guaranteed.

_____ _____

Signature of participant Date

Appendix 5: Descriptors for CEFR levels and IELTS bands 8 and 9

Common European Framework of Reference Descriptors

Proficient User	**C2**	Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.
	C1	Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.
Independent User	**B2**	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics, which are familiar, or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.

Basic User	**A2**	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
	A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.

IELTS Band 8 and 9 Descriptors

BRITISH COUNCIL idp IELTS AUSTRALIA UNIVERSITY of CAMBRIDGE ESOL Examinations

IELTS Speaking band descriptors (public version)

Band	Fluency and Coherence	Lexical Resource	Lexical Resource	Pronunciation
9	▪ speaks fluently with only rare repetition or self correction; any hesitation is content-related rather than to find words or grammar ▪ speaks coherently with fully appropriate cohesive features ▪ develops topics fully and appropriately	▪ uses vocabulary with full flexibility and precision in all topics ▪ uses idiomatic language naturally and accurately	▪ uses a full range of structures naturally and appropriately ▪ produces consistently accurate structures apart from 'slips' characteristic of native speaker speech	▪ uses a full range of pronunciation features with precision and subtlety ▪ sustains flexible use of features throughout ▪ is effortless to understand
8	▪ speaks fluently with only occasional repetition or self-correction; hesitation is usually content-related and only rarely to search for language ▪ develops topics coherently and appropriately	▪ uses a wide vocabulary resource readily and flexibly to convey precise meaning ▪ uses less common and idiomatic vocabulary skilfully, with occasional inaccuracies ▪ uses paraphrase effectively as required	▪ uses a wide range of structures flexibly ▪ produces a majority of error-free sentences with only very occasional inappropriacies or basic/non-systematic errors	▪ uses a wide range of pronunciation features ▪ sustains flexible use of features, with only occasional lapses ▪ is easy to understand throughout; L1 accent has minimal effect on intelligibility
7	▪ speaks at length without noticeable effort or loss of coherence ▪ may demonstrate language-related hesitation at times, or some repetition and/or self-correction ▪ uses a range of connectives and discourse markers with some flexibility	▪ uses vocabulary resource flexibly to discuss a variety of topics ▪ uses some less common and idiomatic vocabulary and shows some awareness of style and collocation, with some inappropriate choices ▪ uses paraphrase effectively	▪ uses a range of complex structures with some flexibility ▪ frequently produces error-free sentences, though some grammatical mistakes persist	▪ shows all the positive features of Band 6 and some, but not all, of the positive features of Band 8
6	▪ is willing to speak at length, though may lose coherence at times due to occasional repetition, self-correction or hesitation ▪ uses a range of connectives and discourse markers but not always appropriately	▪ has a wide enough vocabulary to discuss topics at length and make meaning clear in spite of inappropriacies ▪ generally paraphrases successfully	▪ uses a mix of simple and complex structures, but with limited flexibility ▪ may make frequent mistakes with complex structures, though these rarely cause comprehension problems	▪ uses a range of pronunciation features with mixed control ▪ shows some effective use of features but this is not sustained ▪ can generally be understood throughout, though mispronunciation of individual words or sounds reduces clarity at times

5	▪ usually maintains flow of speech but uses repetition, self-correction and/or slow speech to keep going ▪ may over-use certain connectives and discourse markers ▪ produces simple speech fluently, but more complex communication causes fluency problems	▪ manages to talk about familiar and unfamiliar topics but uses vocabulary with limited flexibility ▪ attempts to use paraphrase but with mixed success	▪ produces basic sentence forms with reasonable accuracy ▪ uses a limited range of more complex structures, but these usually contain errors and may cause some comprehension problems	▪ shows all the positive features of Band 4 and some, but not all, of the positive features of Band 6
4	▪ cannot respond without noticeable pauses and may speak slowly, with frequent repetition and self-correction ▪ links basic sentences but with repetitious use of simple connectives and some breakdowns in coherence	▪ is able to talk about familiar topics but can only convey basic meaning on unfamiliar topics and makes frequent errors in word choice ▪ rarely attempts paraphrase	▪ produces basic sentence forms and some correct simple sentences but subordinate structures are rare ▪ errors are frequent and may lead to misunderstanding	▪ uses a limited range of pronunciation features ▪ attempts to control features but lapses are frequent ▪ mispronunciations are frequent and cause some difficulty for the listener
3	▪ speaks with long pauses ▪ has limited ability to link simple sentences ▪ gives only simple responses and is frequently unable to convey basic message	▪ uses simple vocabulary to convey personal information ▪ has insufficient vocabulary for less familiar topics	▪ attempts basic sentence forms but with limited success, or relies on apparently memorised utterances ▪ makes numerous errors except in memorised expressions	▪ shows some of the features of Band 2 and some, but not all, of the positive features of Band 4
2	▪ pauses lengthily before most words ▪ little communication possible	▪ only produces isolated words or memorised utterances	▪ cannot produce basic sentence forms	▪ speech is often unintelligible
1	▪ no communication possible ▪ no rateable language			
0	▪ does not attend			

Appendix 6: Transcription conventions

Feature	Symbol and Comment	Example
Speaker	<$1>, <$2>, <$3> etc.	Each speaker is numbered in order of entering conversation. Where the speaker cannot be identified, it can be guessed: <$1?> or entered as an unknown speaker: <$M> (male) or <$F> (female)
Punctuation	**. ? '** NOTE: (i) If an utterance ends in a code bracket, the punctuation mark comes after the final code brackets. (i) Only use full stops, question marks and apostrophes. **A full stop is used to show a complete utterance rather than a complete grammatical sentence.** (ii) No commas are used.	<u>Example 1</u> <$1> They're acorns. <$2> Acorns no they're not they're pine things. <$3> Pine cones. <$2> Pine cones. <$1> Acorns. <u>Example 2</u> <$1> Yeah you've seen it obviously I mean. <$2> Oh I have yes.

		<$1> Did you discuss whether or not it is permanent <$G3>?
Spacing	(i) Single line spacing. (ii) Use a single space between the speaker symbol <$> and the beginning of the utterance. (iii) After each speaker turn, leave one blank line.	<$1> Well I didn't know it. <$2> Oh no? <$1> No.
Tabs	Do not use tabs	Example <$2> You know ah I know am he's at the age I suppose it's an impressionable age aam and maybe that's why you know I feel later down the road he would probably want to have it removed.
Capitalisation	Capitalise as normal. At the beginning of each utterance except in the case of interruptions (see section on *Interrupted Utterances*).	<$2> Yes he was on holidays in England and he came back with a nice tattoo on his shoulder and am I was actually listening to your programme yesterday and it prompted me to phone you am with the ear piercing. <$1> Yeah.
Incomplete Words	= (i) Mark incomplete words with an equals sign	Example <$2> No it's aam I w= I don't know how you'd really describe its design.

	(ii) There should be no space between '=' and the incomplete word, e.g. wh=	<$1> And wha= you really did like it?
Truncation\ Repetitions	Only mark truncated utterances if they occur more than once <$=> marks the beginning of a truncated utterance. <\$= > marks the end of a truncated utterance. NOTE: (i) Note there must be a space between the '=' and the final code bracket e.g. <$= > (ii) If the repetition is intentional, this should not be marked as a truncation, see example 3.	<u>Example 1</u> <$1> Come here <$=> your bottom your bottom <\$=> your bottom branch. <u>Example 2 (not marked as a truncation)</u> <$1> Put that on on in the bottom. <u>Example 3 (intentional repetition)</u> <$1> I said I would give him a certain sum of money. <$2> How much? How much? How much?
Interrupted Utterances	+ is used to mark the end of the interrupted utterance and also to mark the beginning of a resumed utterance.	<u>Example 1</u> <$2> Here Mam will you throw me a couple of+ <$1> Get out of my way.

	NOTE: (i) Resumed utterances begin with lower case letters. (i) There should be no space between '+' and the final word of the interrupted utterance. The same applies to the resumed utterance (i) Not all interrupted sentences are resumed see examples 3. In these cases, do not use the + symbol	<$2> +boxes. Example 2 <$1> Am I was just wondering if you can have laser treatment and how successful+ <$2> It is. <$1> +it's to have it removed. Example 3 <$1> Oh look the start of the ones that Mum hates because they're. <$2> It's awful. <$1> Can I put these up? These have to go on the tree.
Overlaps	**<$O>** marks the beginning of an overlap. **<\$O>** marks the end of an overlap.	Example 1 <$1> I just <$O1> never liked them <\$O1>.

	NOTE: (i) These symbols first appear in the main utterance marking where the overlap begins and ends. (ii) The actual overlapping utterance is given in the next speaker turn. (i) The number in the overlap symbol must correspond with the speaker who has been overlapped.	<$2> <$O1> It's not just <\$O1> a case of saying ''what a pity". <u>Example 2</u> <$1> Reflection <$O1> in that tone <\$O1>. <$2> <$O1> Yes do you know <\$O1> it's religious it's the birth of Jesus.
Uncertain or Unintelligible Utterances	<$G?> Unintelligible utterance where the number of syllables cannot be guessed. <$G1>, <$G2> … <$G5>, <$G5+> The number of unintelligible syllables can be guessed up to a maximum of five after which <$G5+> will suffice. <$H> marks a guessed utterance. <\$H> marks a the end of a guessed utterance.	<u>Example 1</u> <$2> Now I'd better <$G?> that has a big enough string to go around it. <u>Example 2</u> <$1> Yes it's religious it's the birth of Jesus and we should all be aware of that <$H> oh God <\$H>. <u>Example 3</u> <$1> <$H> I'm not <\$H> sure about <$H> these <\$H>. <u>Example 4</u> <$1> <$G3> where they went <$G5+> Mike.

Reported Speech		Example 1 <$1> I definitely heard him say ‘‘I’ll be there at five o’clock”. Example 2 <$1> Well there was one am ‘‘love” on one hand and ‘‘hate” on the other hand then of course the usual ones you know ‘‘I love Mum” and whatever the girlfriend’s name was at the time.
Titles	Do not use quotation marks and capitalise all initial letters	Example <$1> Have you seen East Of Eden?
Numbers	Number, including dates, should be written in full. EXCEPTION: Use the symbol 0 (zero) when a speaker pronounces zero as ‘oh’, see example 2.	Example 1 <$1> I’d say there was at least five people there. Example 2 <$1> The code for Limerick is 0 six one.
Acronyms	Capitalise acronyms and allow a single space between each.	Example <$1> But the one thing that worries me is a risk of infection like H I V.

Abbreviations and short forms	These are not followed by full stops. NOTE the following conventions	Examples etc Dr Mrs ie Prof okay cos tis all right
Extra-linguistic features (laughing, coughing, significant background noise)	<$E> dog barking <\$E> Any relevant extra information can be added as shown in the examples.	Example 1 <$1> <$E1 > shouting <\$E1> That's a good idea. Example 2 <$1> Yeah is it true you though you <$E1> laugh <\$E1> might be able to scrape it off? Example 3 <$2> <$E> singing <\$E> oh oh oh the sweetest thing.

Non-standard contractions	**<$X>** this symbol goes before the non-standard contraction. **<\$X>** this symbol goes after the standard version of the non-standard contraction. **|** precedes the standard version. NOTE : This convention does not apply to standard contractions, e.g., I'll, there's, can't, haven't	Example 1 <$1> He <$X> shouldn't've | shouldn't have <\$X> done that. Example 2 <$1> <$X> D'you | do you <\$X> know what I mean? Example 3 <$1> Is that your bag? <$2> <$X> 'Tis | it is <\$X>.
Fillers, backchannels, hesitations unconventional phrases etc.	NOTE the following conventions	Jeekist \Jeez\ Jeenie am ah oh eh ur mmhm (agreeing) hm? (question) mm yeah

		yep ahah hah? shure nah shush ye (you plural) hey okay all right aw yera

Appendix 7: Writing task sample

Writing task samples

Task: Letter home

Description of the task: Based on Headway Upper and Pre-intermediate, unit 1 (John and Liz Soars, 2010) students were tasked with filling in the grammatical gaps in letters before writing their own letter home to their family or friends.

Level that task was given to: A2 and C1

Details of the example: Student <$15>, A2

Dear Jane
Thank you to receive me in your house. I'm also very happy and impatient to stay with you. I'm sure that I am going to stay with a lovely family as I could see in the picture you send me.
My airplane arrives at Gatwick airport in the morning at 9h35 the 25th of August. I'll probably get on the train twenty or thirty minutes later but don't worry I have a mobile phone I'm going to phone you as soon as I'm on the train and we'll meet us at the station at time.
I don't have any problems with the food but I don't eat after 9pm, I don't smoke. I prefer to share my room with a student especially if it's a girl and I could study with her. I'm going to stay a couple of months after the end of my course I would like to take my diploma before leaving you and eventually spend more time with you.
Best wishes

Task: Essay on communication

Description of the task: Based on Advanced Masterclass (Cambridge University Press; 2010) students took part in a spoken debate on the topic of communication they also reads a text from the course book. Then, for homework, they had to write 100 words about what communication meant to them.

Level that task was given to: C1
Details of the example: Student <$2>, C1

Communication
There are so many different ways to communicate. You can use the phone, write on the internet, write a letter or talk face to face. But which one is the best way to communicate? I think the best way is to talk face to face because if you talk on the phone to someone you can only hear the other person. And if you write in the internet you can only read. On the internet it can be that they missunderstand each other because they can't see any body language and can't hear the voice of each other. It is also easyer to tell not the truth if you don't talk face to face. But I have to say all those inventions like internet or mobile phones are really good if you aren't live side by side. If you need to talk to someone who isn't at home you can ring him on his mobile phone it's great. For me, for example it's at the moment really handy to text with friends in the internet because I am in

Ireland and my friends are in Germany. So without internet I will be happy when I am back and can talk

Task: Write an essay comparing Ireland with your home country

Description of the task: Students at both levels had to write a minimum 100 word essay comparing and contrasting Ireland with their home country.

Level that the task was given to: A2 and C1.

Details of the example: Student <$12>, C1.

<$12> Clementia

Ireland compared to my country

I would say Ireland and Germany are really different. But that is my own experience. First of all one completely different thing: Ireland is an island and the people are driving on the left side of the road. The Germans are driving on the right side and they don't live on an island. In Ireland you can find much more roundabouts than in Germany. The German traffic contains more traffic lights than roundabouts. But not only the traffic is different also the countryside. The country side looks in Ireland greener because it is raining heavier and more. Instead of the rain we have in Germany much snow in the winter months. Also the winter is colder and longer. Because of the snow and the high mountains in Germany you can go skiing. Snow in Ireland is seldom and the mountains are smaller. So skiing is impossible. Another thing which looks unlike is the fields. Between the Irish fields are often small walls or hedges. In Germany we don't have those walls or hedges in our countryside. It's the same with their houses The Irish people like it to build a wall around their houses and gardens with a gate in front of the house. In Germany it is unusual to have suche walls with gates. But also not only the country side and the traffic are different also the people. Ireland is known for friendly and helpful people and for hospitality. I agree with this statement. One typical and friendly thing is to ask "how are you" after the salutation. It doesn't matter if they know the person or not. They ask everyone. In Germany we just ask friends this question. It would be curious if you say hello to anyone on the street in Germany and ask how are you. In opposition to Germany it is normal to say this in Ireland. I thing it is simply part of the salutation. Germany and Ireland are completely different in the organization. Germans like it to organise or plan their day or even better their week. They are also known for punctuality. The Irish people are more unorganized and not in time. But I think because of that they are more spontaneous and they live for the moment. So I would say the Irish people like their life more relaxed than the Germans. One example for a typical organisation of a German is to plane the dinner. A German person would think about the dinner one day before or in the morning time. An Irish person would take a look into the fridge one hour before he wants start cooking and decide then. One of the Irish relaxed things which I noticed immediately was the time of school beginning. The German school doesn't start so late than the Irish one. I would ay there is different food in those countries. The Irish people eat much more toast so they have a bigger selection of toast. In Germany we eat more brown bread than toasts. The typical beans in red sauce we don't eat and also fish and chips are unusual in Germany. In Ireland you can get fish and chips nearly everywhere. Something which I really like on Ireland are the pubs because all people at all ages can go into a pub to have fun. In Germany there are not that much places where everyone in all ages can go. Finally I have to say

that is my impression of Ireland compared to Germany. I thing both countries have their good and not so good things.

Task: Job application and CV

Description of the task: Both groups were tasked with completing their own job application, curriculum vitae and participated in a job interview with the teacher as part of their final assessment.

Level that the task was given to: A2 and C1

Task: Myself

Description of the task: Students of the A2 cohort were tasked with writing a minimum of 75 words about themselves. This was the first written topic for this level.

Level that the task was given to: A2

Details of the task: Student <$19>, A2

Myself

My name is Ilona Karavajeva. I am 25 years old and in this year 11th November I will be 26. I came from Latvia Riga. I have one sister and her name is Viktorija. She is younger for one year. We look completely differently with my sister-I have brown eyes and dark hair, but she has blond hair and blue eyes. Even she is a younger sister, she is 10 centimetres higher than me. My sister and I are very friendly and kind; we help each other at any time. At this moment Viktorija lives in Belfast, United Kingdom. She is still studying in University if Latvia. She will get award that she graduate University with Bachelor Organization and management diploma after exam s on May. My mother and father are living in Riga in the nice private house that is located very close to Riga centre. Our house is located in perfect place, where is 1 to 2 km away from bus stop, and by bus you can get to city centre and Riga international airport only with 20 minutes. We lived in this place for more than 20 years. Latvia, officially the Republic of Latvia, is a country in the Baltic region of Northern Europe. It is bordered to the North by Estonia, to the south by Lithuania, to the east by the Russia Federation, and to the southeast by Belarus. Across the Baltic Sea to the west lies Sweden. The territory of Latvia covers 64,589 km2 and it has a temperate seasonal climate. The Latvians are Baltic people culturally similar to the Estonians and Lithuanians. Riga is the nice and beautiful city. Riga is the capital and largest city of Latvia with more than 700,000 inhibitions. It is largest city of the Baltic States. Riga's territory covers 307.17 km2 and situated on the month of the Daugava. The river Daugava flowing through Russian Federation and Belarus country. I came to Ireland in January 2006 and lived all this time in Limerick. I couldn't find job for first month. That time was really distressful. It was also so hard to leave Latvia, parents, sister and friends, but in Ireland everything was strange and new. I was glad that I met some good people. Especially Irish people are very friendly and always smiling. After long searching I got job in O'Brien's Sandwich Bar, where I started work in 2006 February. I'm still working at the same place, it is long time. In Ireland now I have good friends with who I spend my free time. Also I have very good colleagues at work and nice people to share house with. I am renting double room in very beautiful house. There are two Latvian dancing bands in Ireland- "Karakums" in Dublin and "Nemiers" in Limerick. I am member in

Limerick dancing band. It is not big dancing group, just 6-8 people. I enjoy time what I am spending at training. I met new friends also. I like Ireland. It is nice country, but weather could be better. ☺

Task: Fairy tale

Description of the task: Both cohorts studied Irish fairy tales and legends such as *Tir na nÓg*, the Banshee and Drunken Tady (a Limerick folklore tale). As homework, both groups had to write about a fairy-tale famous in their home country.

Level that the task was given to: A2 and C1

Task: Fill in form

Description of the task: Based on activities from *Anseo* (National Adult Literacy Agency, 2010), students at both levels learned how to fill in a variety of forms in the target language. Students were tasked with competing job applications, medical card applications, library membership applications and accommodation details.

Level that the task was given to: A2 and C1

Task: Past Tenses

Description of the task: After working on the past simple tense and past progressive aspect in class, the A2 cohort were tasked with writing a text in the past for homework. This text could be about any topic as long as it was written in the past.

Level that the task was given to: A2

Details of the example: Student <$15>, A2

Last summer Bob and Anna went camping in France. When they arrived at the entrance of the village, the sun was shining through their window's car and it was warm inside. First they wanted to visit the most famous castle in this village. After that, they planned to dinner in the restaurant who was in the main square of this castle. Then when they were having dinner, the weather changed suddenly. When they arrived in the campsite it was still raining it was terrible it was windy as well. They try to put up the tent but it was too windy. Finally they slept in their car.

Task: Dear Sue

Description of the task: Having focused on modal verbs in class, students wrote letters of advice to each other based on an agony aunt activity taken from *Reward Intermediate* (Kay, 2006). Students had to read a problem and work together in groups to create a written response with advice. The problems range from private and personal to romantic and professional.

Level that the task was given to: A2 and C1.

Task: Picture story

Description of the task: A further example drawn from *Reward Intermediate* (Kay, 2006) is picture story. Each student was given a picture and from that picture they had to create a narrative.

Level that the task was given to: C1

Details of the example: Student <$13>, C1.

I am a typical woman-sensible romantic that is why my favourite movie is-"P.S. I Love You".
This film was *produced in* the USA in 2007. The director of the film is Richard LaGravenese. The stars of movie: Hilary Swank as Holly and Gerard Butler as Gerry.
This film is very close to me because the subject was related to Ireland and Irish traditions. It shows the Irish culture and Irish people. In the background you can hear Irish music and Irish songs. It also shows the beauty of the Irish nature and allows you to admire the hansomeness and *untamed nature* of the Irishmen. Since seeing the movie, every woman *dreams about* knowing a handsome, charming with a specific sense of humour Irishman, who will love her even after his death.
The film-"P.S. I love You" is a *touching story about* true and infinite love of man to woman and about the way how every woman wants to be loved. It talks about the fact that even in the *face of death*, disease, the man who loves a woman is not *thinking about* himself but about her. He *thinks about love of his life*. He also *thinks about* the changes that will be in her life after his death, when he will go away and how can he help her to survive this time of despair and grief. He wonders what to do to make her *start a new life* without him easier. He *worried about* her, not for himself. This film also shows the nature of women and their relationships with men. It describes what women feel and how they think, what *gives them joy and happiness,* and what *makes them to feel anger* or *shed a tears.*
Here is a brief synopsis of the film:
Holly is beautiful, smart and colourful woman. She is married to a passionate, funny and impetuous Irishman named Gerry. Gerry is *the love of her life*. So, when Gerry's *life is taken* by an illness, it takes the life out of Holly. Holly loses husband far too early and tragically. She very hard grieves for husband. Nobody knows Holly better than Gerry. He is the only one who can help her and understand but he is no longer there. It is a good thing he planned ahead. Before he died, Gerry wrote Holly a series of letters, each one causing *a fresh stream of tears to fall.* All of those letters will guide her, not only through her grief, but also in rediscovering herself. The first message arrives on Holly's 30th birthday in the form of cake. On the cake is a tape recording from Gerry. She hears Gerry's voice to her *utter shock,* who tell her to *go out* and celebrate herself. In the weeks and months that follow, more letters from Gerry are delivered in surprising ways. All of the letters are sending her on a new adventure and are *signing off* in the same way: "P.S. I Love You". During all this time, Holly's mother and best friends are worrying that Gerry's letters are keeping Holly *tied to the past.*
This film is not a light comedy or a date movie. It is very, very sad film. Even the happy parts are sad. If you go to see this, you have to *take many tissues* because you will *shed a tears.* After watching this story every woman dreams that although once in her life like Holly and meet the charming Irishman named Gerry. Unfortunately, men don't like this film. They say that is too sentimental, tender and dramatic, without any action and the main theme of the film is the death of a loved one. They say that it is a sad story about love.
Maybe that is why-"P.S. I love You" is a typical woman' film.

Task: Fold line text

Description of the task: In order to test learner's spelling and grammatical accuracy, twice during the term, students were given a predictive text. A sheet of paper with sentences folded was given to both groups. Each learner can only see one sentence due to the folds in the paper and must create a sentence that they feel would fit with the one which they can see. These sentences were added to the corpus and form part of the twenty thousand written corpus**.**

Level that the task was given to: A2

Task: Grammar gap fill

Description of the task: Throughout the twelve week course, students of both levels completed various grammar gap fills. These gap fills included tenses, aspects and lexis and were also included in the word count of the written corpus.