

**An examination of successful language use at B1, B2 and C1
level in UCLanESB speaking tests in accordance with the
Common European Framework of References for Languages**

by

Shelley Byrne

A thesis submitted in partial fulfilment for the requirements for the degree of
Doctor of Philosophy at the University of Central Lancashire

June 2016



STUDENT DECLARATION FORM

Concurrent registration for two or more academic awards

I declare that while registered as a candidate for the research degree, I have not been a registered candidate or enrolled student for another award of the University or other academic or professional institution

Material submitted for another award

I declare that no material contained in the thesis has been used in any other submission for an academic award and is solely my own work

Collaboration

Where a candidate's research programme is part of a collaborative project, the thesis must indicate in addition clearly the candidate's individual contribution and the extent of the collaboration. Please state below:

Signature of Candidate Shelley Byrne

Type of Award PHD

School SCHOOL OF LANGUAGES AND GLOBAL STUDIES

ABSTRACT

The concept of success in second language learning has received growing attention over the last few years. With the earlier dominance of the native speaker as a model and measure for success, learner language had previously been seen as inferior or deviant from the native speaker target and norm. However, with the arrival of the *Common European Framework of Reference for Languages* [CEFR] (CoE, 2001), a shift in emphasis saw learner proficiency across all language use viewed in terms of what learners could do, rather than what they could not. Despite this more positive outlook, the CEFR, in its aim of being non-language specific to maintain applicability across language learning contexts, fails to effectively illustrate how its many descriptors and scales apply to learner language at different levels. Though it extensively documents *what* learners are able to do with their language, it fails to specify exactly *how* they can actually do it.

The aim of this study was therefore to examine what makes B1, B2 and C1 learners successful in their speech in accordance with the CEFR. In being successful, learners' spoken performance should correspond with the criteria presented in the CEFR's descriptors and scales for their current proficiency level. Employing corpus linguistics methodology, speech was chosen as a focus for this study as it i) represents the skill by which learners are most commonly judged, and ii) represents an area that has not received as much attention in corpus linguistics and learner corpus research. Via the compilation of a 70,578 token learner corpus of speaking exam data, analysis examined learners' use of their core vocabularies and the occurrence and realisation of can-do statements for speech. The study's research questions asked: What percentage of the words used by successful B1, B2 and C1 learners came from the first 1000 and second 1000 most frequent words in English? What were the 20 most frequent words at B1, B2 and C1 and their notable collocations, colligations and functions? What were the 20 most frequent keywords at B1, B2 and C1 and their notable collocations, colligations and functions? What were the most frequent 3- and 4-word chunks at B1, B2 and C1 and their notable collocations and functions? and What CEFR indicators are present in terms of spoken interaction, spoken production and strategies at B1, B2 and C1 and how are they realised?

Overall, the study ascertained that B1, B2 and C1 learners differed very little in their use of the 2000 most frequent words in English. Though B1 learners evidenced significantly fewer word types and tokens than their B2 and C1 counterparts, only 1 in 33 words at all three levels came from beyond the 2000 word vocabulary limit. The findings

indicated that quantitative measures alone were not able to distinguish learners at different levels nor the changes for vocabulary range suggested in the CEFR. With 97% token coverage achieved, however, this core vocabulary of 2000 words did allow learners to be successful in their speech. In addition, frequent and keyword data alongside lexical chunk analysis established that lexis such as *we, er, erm, think, so, like, a lot of, agree with you*, exhibited several similarities and differences across the levels. They supported the argument that for learners to be successful in their speech, they need to be supplied with multifunctional lexis which not only aims to broaden their vocabularies, but which intends to deepen it too. By making learners more aware of what can be achieved with the vocabulary they already possess, transitions across proficiency levels can seem more within reach. Finally, analysis of CEFR can-do statements revealed several key points, in particular for successful pragmatics and discourse at B1, B2 and C1. It identified that the objective of interaction evolved across levels: B1 learners expressed their thoughts and opinions, B2 learners asked for the opinions of others and C1 learners, though demonstrating the same skills as B1 and B2 learners, sought mostly to elaborate more fully on the reasons for their thoughts. In terms of the combination of ideas, simplistic chains were evidenced via the use of conjunctions but whilst B1 learners concerned themselves mostly with simple addition of ideas, B2 and C1 learners looked towards utterances expressing cause-and-effect.

In sum, the study demonstrated that some of the misconceptions about learner language do have a considerable impact on learner success. Progression across proficiency levels presented itself more in the flexibility and multifunctionality of lexis, rather than its complexity or level of difficulty. It concluded that by making expectations more realistic and not by presuming that learners would do more in their speech than any native speaker, more can be learnt about what learners *are* able to achieve. Also, by supplementing teaching with materials based on real examples of successful learner speech, learners can be presented with more relevant, more realistic and more attainable models of language use.

ACKNOWLEDGEMENTS

I would first like to express my sincere gratitude and personal thanks to Christian Jones and Daniel Waller. Without your continued support, guidance and encouragement over my many years at UCLan, I doubt that I would have got this far and have grown so much as a teacher and researcher; you've certainly led by example. I also want to say a special thank you to all those who have advised or contributed to my research: Svenja Adolphs, June Challis, Patrycja Golebiewska, Anna Gregson, Nick Gregson, Nicola Halenko, Tania Horak, Mike Scott, Sylvia Tito, Nicola Walker, CEG campus, the independent language school and, of course, the examiners and students who agreed to take part in the study. I also send an extended, heartfelt thank you to the WISER team Nick Gregson, Ruth Bavin, Hazel Gourley, Kelly Mcateer and Karen Smith for being a continued and invaluable source of support, humour (and biscuits!) over the past few years.

To Paul and my family, I know that none of this would even have been possible without your unwavering encouragement, support and belief which has got me to this point today. You have and continue to inspire me every day and know that I could never thank you enough for all that you have done not only during this PhD, but all throughout my life. To my friends, thank you for just being you.

PS. Paul, regarding the bet we made a few years ago, for once, it looks like you were right!

CONTENTS

Abstract.....	iii
Acknowledgements.....	v
Table of figures.....	xi
List of Tables.....	xii
List of abbreviations.....	xiv
1. Introduction.....	1
1.0 Chapter introduction.....	1
1.1 Preliminary rationale for exploring success.....	1
1.2 The NS model of success and the gaps that remain.....	2
1.3 Success at different levels.....	5
1.4 Statement of research aims.....	7
1.5 A focus on speaking.....	8
1.6 The research questions.....	9
1.7 Chapter conclusion and overview of thesis.....	9
2. Literature Review (part one): Communicative competence and gaps in the CEFR.....	11
2.0 Introduction.....	11
2.1 Theories of communicative competence and their impact on learner success.....	11
2.1.1 Hymes' theory of communicative competence.....	11
2.1.2 Canale & Swain's theory of communicative competence.....	15
2.1.3 Canale's theory of communicative competence.....	19
2.1.4 Bachman and Palmer's model of language use and performance.....	22
2.2 Communication skills and the effect of pragmatic and strategic competences on success.....	28
2.3 Linking communication, communicative competence and learner success to the CEFR.....	33
2.3.1 A brief introduction to the CEFR and its relevance to the study of success.....	34
2.3.2 Language use, general competences and communicative competences.....	36
2.3.3 Linking the CEFR to success.....	40
2.4 Chapter conclusion.....	44
3. Literature Review (part two): Findings from corpora and implications for learner success in speech.....	46
3.0 Introduction.....	46
3.1 An introduction to corpus linguistics and its significance for speech.....	46
3.1.1 What is corpus linguistics and what do corpora contain?.....	46
3.1.2 The benefits and drawbacks of corpus linguistics.....	48
3.1.3 The significance of corpus linguistics for this study of spoken success.....	50
3.2 Corpus findings for spoken grammar and their implications for success.....	51
3.3 Corpus findings for lexical competence and their implications for spoken success.....	55
3.3.1 Vocabulary size.....	56

3.3.2 Formulaic language: Collocation and lexical chunks	58
3.4 A brief rationale of the research questions.....	63
3.5 Chapter conclusion.....	65
4. Pilot study	66
4.0 Introduction.....	66
4.1 Pilot corpus type, level and research population.....	67
4.2 Procedure for constructing the C1 pilot corpus.....	69
4.3 Pilot study demographics	71
4.4 Pilot study research questions and brief rationale.....	72
4.5 Methods used for pilot corpus analysis.....	74
4.5.1 Analysis used for research question 1	74
4.5.2 Analysis used for research questions 2a-c	75
4.5.3 Analysis used for research question 3.....	76
4.6 Pilot study findings and discussion.....	77
4.6.1 RQ1: What percentage of the words used by successful C1 learners come from the first 1000 and second 1000 most frequent words in English?.....	77
4.6.2 RQ2a: What were the 20 most frequent words at C1 and their notable collocations, colligations and functions?.....	82
4.6.3 RQ2b: What were the 20 most frequent keywords at and their notable collocations, colligations and functions?.....	89
4.6.4 RQ2c: What were the most frequent 3- and 4-word chunks at C1 and their notable collocations and functions?.....	94
4.6.5 RQ3: What Common European Framework of Reference for Languages (CEFR) indicators are present in terms of spoken interaction, spoken production and strategies at C1?	99
4.7 Chapter conclusion.....	102
4.8 Changes to methodology following the pilot study	103
5. Methodology	106
5.0 Introduction.....	106
5.1 UCLanESB corpus demographics	106
5.1.1 B1, B2, C1 sub-corpora and combined UCLanESB corpus information.....	107
5.1.2 Candidate L1	108
5.1.3 Candidate nationality	108
5.1.4 Exam topics per level.....	109
5.1.5 Exam duration per level	109
5.1.6 Tokens per exam section.....	109
5.2 Corpus construction rationale	109
5.2.1 Research tool for collecting learner language.....	110
5.2.2 Corpus type	114
5.2.3 Corpus compiler.....	115

5.2.4 Corpus representativeness.....	116
5.2.5 Corpus balance.....	120
5.2.6 Corpus size.....	122
5.2.7 Corpus sampling.....	125
5.2.8 Transcription.....	127
5.2.9 Ethics.....	130
5.3 Data analysis.....	133
5.3.1 RQ1: What percentage of the words used by successful B1, B2 and C1 learners come from the first 1000 and second 1000 most frequent words in English?.....	133
5.3.2 RQ2a: What were the 20 most frequent words at B1, B2 and C1 and their notable collocations, colligations and functions?.....	135
5.3.3 RQ2b) What were the 20 most frequent keywords at B1, B2 and C1 and their notable collocations, colligations and functions?.....	141
5.3.4 RQ2c) What were the most frequent 3- and 4-word chunks at B1, B2 and C1 and their notable collocations and functions?.....	144
5.3.5 RQ3) What Common European Framework of Reference for Languages (CEFR) indicators are present in terms of spoken interaction, spoken production and strategies at B1, B2 and C1 and how are they realised?.....	145
5.4 Chapter conclusion.....	151
6. Results and discussion.....	152
6.0 Introduction.....	152
6.1 RQ1: What percentage of the words used by successful B1, B2 and C1 learners come from the first 1000 and second 1000 most frequent words in English?.....	152
6.1.1 RQ1 Results.....	152
6.1.2 RQ1 Discussion.....	156
6.1.3 RQ1 Summary.....	159
6.2 RQ2a: What were the 20 most frequent words at B1, B2 and C1 and their notable collocations, colligations and functions?.....	160
6.2.1. <i>We</i>	162
6.2.2 Discussion of the frequent word <i>we</i>	169
6.2.3 <i>Er/Erm</i>	171
6.2.4 Discussion of the frequent words <i>er</i> and <i>erm</i>	173
6.2.5 RQ2a Summary.....	175
6.3 RQ2b: What were the 20 most frequent keywords at B1, B2 and C1 and their notable collocations, colligations and functions?.....	176
6.3.1 <i>Think</i>	180
6.3.2 Discussion of the keyword <i>think</i>	184
6.3.3 <i>So</i>	186
6.3.4 Discussion of the keyword <i>so</i>	192
6.3.5 <i>So, very, really</i>	193
6.3.6 Discussion of the keywords <i>so, very, really</i>	195

6.3.7 <i>Like</i>	196
6.3.8 Discussion of the keyword <i>like</i>	200
6.3.9 RQ2b summary	201
6.4 RQ2c: What were the most frequent 3- and 4-word chunks at B1, B2 and C1 and their notable collocations and functions?	202
6.4.1 <i>A lot of</i>	204
6.4.2 <i>Agree with</i>	207
6.4.3 Topic-related chunks.....	209
6.4.4 B1, B2 and C1 lexical chunk discussion.....	210
6.4.5 RQ2c summary	211
6.5 RQ3: What Common European Framework of Reference for Languages (CEFR) indicators are present in terms of spoken interaction, spoken production and strategies at B1, B2 and C1 and how are they realised?.....	212
6.5.1 B1 can-do occurrence.....	213
6.5.2 B2 can-do occurrence.....	215
6.5.3 C1 can-do occurrence.....	218
6.5.4 A closer comparison of abilities and language at B1, B2 and C1	220
6.5.5 Production	223
6.5.6 Production statement summary	227
6.5.7 Interaction	228
6.5.8 Interaction statement summary	233
6.5.9 Production strategies	234
6.5.10 Production strategy summary.....	236
6.5.11 Interaction strategies	237
6.5.12 Interaction strategy summary.....	240
6.5.13 RQ3 summary	241
7. Conclusion:	244
7.0 Chapter introduction	244
7.1 The aims and main findings of this study	244
7.2 Limitations	251
7.3 Implications for teaching and learning.....	252
7.4 Implications for research.....	256
7.5 Concluding remarks	258
8. References.....	259
9. Appendices.....	276
Appendix 1: Exam structure	277
Appendix 2: UCLanESB marking criteria	278
Appendix 3: UCLanESB marking scheme	281
Appendix 4: Number of UCLanESB exams taken at Preston campus	282

Appendix 5: Nationality figures for students in the School of language, literature and international studies	283
Appendix 6: Advance notice email	284
Appendix 7: Presentation for obtaining consent from pre-sessional learners	285
Appendix 8: UCLanESB consent form.....	286
Appendix 9: Vocabulary profile: B1.....	287
Appendix 10: <i>Text Inspector</i> analysis of B1, B2 and C1 speech across exam parts.....	290
Appendix 11: LINDSEI 20 most frequent words	291
Appendix 12: Final can-do statements.....	292
Appendix 13: Can-do statement verification exercise	295
Appendix 14: Interaction can-do statements: Seeking opinions	296

TABLE OF FIGURES

Figure 1: Hymes' (1972) theory of communicative competence	13
Figure 2: Canale & Swain's (1980) theory of communicative competence	16
Figure 3: Canale's (1983) theory of communicative competence	21
Figure 4: Bachman and Palmer's (1996: 63) language use and performance with Bachman's (1990: 87) components of language ability.....	24
Figure 5: A break-down of production skills (Bygate, 1987)	31
Figure 6: A break-down of interaction skills (Bygate, 1987)	31
Figure 7: Bygate's (1987: 50) summary of oral skills	33
Figure 8: CEFR Common Reference Levels (CoE, 2001: 23).....	35
Figure 9: The CEFR's (Cambridge ESOL, 2011: 10) model of general and communicative competences.....	38
Figure 10: An example 'users may wish to consider' box (CoE, 2001: 102)	39
Figure 11: CEFR Qualitative aspects (CoE, 2001: 29) with vague words in bold	42
Figure 12: CEFR example of can-do scale. Spoken interaction (CoE, 2001: 79).....	43
Figure 13: CEFR scales for vocabulary range (CoE, 2001: 112)	58
Figure 14: Uses of <i>you</i> in questions. Part B, pilot study.	88
Figure 15: Pilot study can-do statement occurrence in parts A, B and C at C1	100
Figure 16: Procedure followed for checking transcription accuracy and consistency	130
Figure 17: Overview of informed consent process	131
Figure 18: Collocates of 'so' with high MI scores in C1 data	140
Figure 19: Summary of RQ2a's analytical procedure	141
Figure 20: Overview of analytical process for RQ3	147
Figure 21: Can-do statement similarity.....	149
Figure 22: Usage of <i>we</i> for confirmation and discourse management at B2 and C1	165
Figure 23: Modal and auxiliary verbs in R1 position in Part C at B1.....	166
Figure 24: Modal and auxiliary verbs in R1 position in Part C at B2.....	167
Figure 25: Modal and auxiliary verbs in R1 position in Part C at C1.....	167
Figure 26: CEFR descriptors for cohesion and coherence (CoE, 2001: 125).....	175
Figure 27: BNC Spoken sample of <i>I think maybe</i>	184
Figure 28: C1 concordance lines for <i>so</i> and <i>things</i>	191
Figure 29: C1 concordance lines for <i>so</i> and <i>really</i>	191
Figure 30: C1 concordance lines for <i>so</i> and <i>actually</i>	192
Figure 31: Uses of <i>like</i> (% proportions).....	196
Figure 32: <i>Like</i> trends across UCLanESB data.....	197
Figure 33: Concordances for <i>something like</i> (B2)	199
Figure 34: Concordances for <i>like if</i> (C1)	199
Figure 35: Concordances for <i>have like</i> (C1)	200
Figure 36: <i>I agree with you but</i> B2	208
Figure 37: <i>I agree with you because</i> C1	208
Figure 38: B1 production statement occurrence	213
Figure 39: B1 interaction statement occurrence.....	213
Figure 40: B1 production strategy statement occurrence.....	214
Figure 41: B1 interaction strategy statement occurrence	214
Figure 42: B2 production statement occurrence	215
Figure 43: B2 interaction statement occurrence.....	216
Figure 44: B2 production strategies statement occurrence	216
Figure 45: B2 interaction strategies statement occurrence	217
Figure 46: C1 production statement occurrence	218
Figure 47: C1 interaction statement occurrence.....	218
Figure 48: C1 production strategy statement occurrence.....	219
Figure 49: C1 interaction strategy statement occurrence	219
Figure 50: Second-most frequent production statements (CoE, 2001: 58-59).....	224
Figure 51: Cohesion device list.....	224
Figure 52: Opinion-related interaction statements (CoE, 2001: 74, 77, 79).....	229
Figure 53: Proportions of opinion patterns at B1, B2 and C1.....	231

Figure 54: Can-do statements for clarification (CoE, 2001: 87).....	237
Figure 55: Questions from B1, B2 and C1 data which may sound more natural.....	240
Figure 56: B1 Marking criteria (UCLanEB, n.d.-b: 24)	278
Figure 57: B2 Marking criteria (UCLanEB, n.d.-b: 27)	279
Figure 58: C1 Marking criteria (UCLanEB, n.d.-b: 30)	280
Figure 59: Marking scheme for B1, B2 and C1 (UCLanEB, n.d.-b: 31)	281
Figure 60: Advanced notice email	284
Figure 61: Presentation for pre-sessional students.....	285
Figure 62: UCLanESB consent form	286
Figure 63: Learner speech at B1, B2 and C1 according to CEFR levels	290
Figure 64: Can-do statement verification exercise.....	295

LIST OF TABLES

Table 1: Pilot study corpus demographics	72
Table 2: Pilot study C1 Vocabulary Profile - Entire exam	78
Table 3: Pilot study K1 and K2 words in Parts A-C.....	81
Table 4: 20 most frequent words in pilot study - Entire exam.....	82
Table 5: Pilot study frequency lists for Parts A, B and C	83
Table 6: Usage of <i>er</i> and <i>erm</i> after questions.....	85
Table 7: Top 20 keywords in pilot study - Entire exam.....	89
Table 8: Top 20 pilot study keywords in Parts A, B and C	90
Table 9: Uses of <i>so</i> in the C1 pilot corpus	92
Table 10: 20 most frequent 3-word chunks in pilot study - Entire exam.....	94
Table 11: 20 most frequent 4-word chunks in pilot study - Entire exam.....	95
Table 12: Amendments to the main study	104
Table 13: B1 sub-corpus	107
Table 14: B2 sub-corpus	107
Table 15: C1 sub-corpus	107
Table 16: UCLanESB corpus composition.....	107
Table 17: UCLanESB learner L1.....	108
Table 18: UCLanESB learner nationality	108
Table 19: Topics discussed at B1, B2 and C1.....	109
Table 20: Exam length at B1, B2 and C1	109
Table 21: Tokens in Parts A, B and C at B1-C1	109
Table 22: Applying Sinclair's (2004) criteria to UCLanESB corpus.....	117
Table 23: Applying Tono's (2003) learner corpus criteria to UCLanESB corpus.....	117
Table 24: Final external criteria for UCLanESB corpus.....	120
Table 25: Percentage of K1 and K2 words at B1, B2 and C1.....	152
Table 26: Types, tokens, means and SDs for K-1 and K-2.....	154
Table 27: <i>Text Inspector</i> analysis of B1, B2 and C1 learner speech.....	156
Table 28: 20 most frequent words at B1, B2 and C1	161
Table 29: Normalised frequencies for <i>we</i>	162
Table 30: Collocates of <i>we</i> at B1, B2 and C1	163
Table 31: Descriptive statistics for <i>er</i> and <i>erm</i>	171
Table 32: Collocates of <i>er</i> and <i>erm</i>	172
Table 33: B1 keywords - BNC World	177
Table 34: B1 keywords – LINDSEI.....	177
Table 35: B2 keywords - BNC World	178
Table 36: B2 keywords – LINDSEI.....	178
Table 37: C1 keywords - BNC World	179
Table 38: C1 keywords – LINDSEI.....	179
Table 39: <i>Think</i> at B1, B2 and C1.....	180
Table 40: Colligational patterns of <i>think</i>	181
Table 41: Collocations of <i>I think</i>	183
Table 42: Frequency of <i>so</i> across Parts A, B and C at B1, B2 and C1	186

Table 43: Normalised frequencies and dispersion of <i>so</i>	187
Table 44: Function of <i>so</i> at B1, B2 and C1.....	188
Table 45: <i>So</i> as a summarising discourse marker	189
Table 46: Common collocates of <i>so</i> at B1, B2 and C1 levels.....	190
Table 47: Collocates of <i>so</i> specific to particular CEFR levels.....	190
Table 48: Adverbs of degree and their collocates	194
Table 49: Uses of <i>like</i> (raw frequencies)	196
Table 50: B1, B2 and C1 three-word chunks	203
Table 51: B1, B2 and C1 four-word chunks	203
Table 52: Accuracy of <i>much</i> and <i>many</i> at B1, B2 and C1	205
Table 53: Common collocates of <i>a lot of</i> across UCLanESB, Spoken BNC and LINDSEI data	206
Table 54: Most frequent production statements.....	221
Table 55: Most frequent interaction statements	221
Table 56: Most frequent production strategy statements	222
Table 57: Most frequent interaction strategy statements	222
Table 58: Cohesion devices across B1, B2 and C1.....	225
Table 59: Opinion patterns at B1, B2 and C1	230
Table 60: Two most common question types at B1, B2 and C1	233
Table 61: Repair across B1, B2 and C1	235
Table 62: Reasons clarification was sought at B1, B2 and C1	238
Table 63: Language used for clarification at B1, B2 and C1	240
Table 64: UCLanESB speaking exam structure (UCLanEB, n.d.-b: 23, 26, 29).....	277
Table 65: Number of exams taken at Preston campus	282
Table 66: Nationality of students from the School of Languages, Literature and International Studies.....	283
Table 67: B1 vocabulary profile	287
Table 68: B2 vocabulary profile	288
Table 69: C1 vocabulary profile	289
Table 70: LINSEI 20 most frequent words	291
Table 71: B1 can-do statements	292
Table 72: B2 can-do statements	293
Table 73: C1 can-do statements	294
Table 74: B1 seeking opinion questions	296
Table 75: B2 seeking opinion questions	297
Table 76: C1 seeking opinion questions	298

LIST OF ABBREVIATIONS

BNC – British National Corpus

CANCODE – Cambridge and Nottingham Corpus of Discourse in English

CLAWS - Constituent Likelihood Automatic Word-tagging System

EGP – English Grammar Profile

EVP – English Vocabulary Profile

L2 – Second language

NS – Native speaker

OPI – Oral proficiency interview

RQ – Research question

SLA – Second language acquisition

TL – Target language

TTR – Type-token ratio

MWU – Multi-word unit

1. INTRODUCTION

1.0 Chapter introduction

In many contexts, the word ‘success’ seems straightforward and does not necessitate any in-depth investigation. Referring to “the accomplishment of an aim or purpose” (Oxford English Dictionary, 2015: paragraph 1), many would concede that this basic definition portrays a rather simplistic, dichotomous relationship: people are either successful in the attainment of their goals or they are not. For learners of a second language [L2], however, such a definition of success proves problematic and shortsighted. It suggests that an objective is clearly identifiable and distinguishable, and that being successful in learning an L2 involves a finite process with a clear start- and end-point when, in fact, varying needs and motivations could result in one learner’s end-goal being another learner’s half-way point in the process. Most importantly, though, the above definition risks overlooking the evolving features of language that can make acquiring all of its aspects an endless, unattainable and, perhaps, unnecessary task for many second language learners. This chapter therefore seeks to introduce the views of some learners and writers regarding what success in a second language constitutes and the barriers that have been created. It will demonstrate how a tendency to judge learners according to native-speaker norms has the potential to associate success with an all-or-nothing, “right or wrong” phenomenon (Lewis, 2012: 45), rather than with the gradual, changeable entity that has been recognised previously in second language acquisition [SLA] (see Corder, 1967; Selinker, 1972). In meeting its overall aim of presenting the rationale as to why a thorough examination of success is needed, this chapter will offer a statement of the study’s specific aims, the justification surrounding its focus on spoken language, the research questions, and an overview of the thesis’ chapters.

1.1 Preliminary rationale for exploring success

As an English language tutor and assessor, personal experience has confirmed the view that many learners aspire towards a native or native-like proficiency in English. A noticeable trend, however, is the tendency for some learners to overlook the fact that they can still be successful in their use of English without achieving such a level. In fact, I have experienced many encounters, similarly to Cook (2002), in which an English

language learner has apologised for their level of English despite their ability to communicate effectively and successfully. The “comparative fallacy” (Bley-Vroman, 1983: 1), involving the comparison and assessment of non-native speakers against native-speaker norms, would seem, therefore, not only to be a feature found in SLA research, but one which appears in learners’ self-assessments. In spite of native-speaker [NS] variation in linguistic and sociocultural abilities causing considerable debate regarding the identification of a sole NS norm (see Andreou and Galantomos, 2009; Cook, 1999; Kramsch, 2003; Lee, 2005; Lyons, 1996; Rampton, 1990), learners still strive towards the NS ideal (Timmis, 2002; 2005). Whilst doing so, it is the author’s view that learners often do not realise that they, as non-native learners, can be “successful users of English” [SUEs] (Prodromou, 2008: xiv), learners who are capable of drawing on their linguistic resources to operate effectively in the sociocultural contexts encountered. With studies of SUEs (see Piller, 2002; Prodromou, 2008) often focussing on more advanced learners or those who can ‘pass’ as NSs, general populations of learners are denied ‘success’ as it is seemingly associated only with the exclusive, elite group of learners who are able to reach the highest levels in language learning.

The preliminary rationale for exploring learner success in English thus stems from the sense that the NS target set by learners, and indeed sometimes by practitioners (see Canagarajah, 2007; Kramsch, 2003; Timmis, 2002), forms the basis of a learner’s assessment of their own achievement: success equates to attaining native-like proficiency whereas other proficiencies equate to failed or incomplete attempts (Birdsong, 2004). The next section aims to highlight some of the problems presented by the NS model as a target and measure for success. It is not the author’s opinion that the NS is a poor model, but by critiquing it, the constraints and effects it has on learners can be identified and linked to the need for greater emphasis on *learner* success.

1.2 The NS model of success and the gaps that remain

Section 1.0 introduced the notion that success is fundamentally reliant upon the accomplishment of a goal. As already explained, native or native-like proficiency is assumed to be the learner’s ideal target in second language learning (Canagarajah, 2007; House, 2003; Piller, 2002; Timmis, 2002; Timmis, 2005). As such, NSs not only provide a model for success, but they also provide a yardstick for interlanguage development (see Selinker, 1972) towards which learners can aim throughout the acquisition process (House, 2003; Jenkins, 2006; Piller, 2002). It the author’s perspective, however, that

success should not solely be associated with reaching a level comparable to that of NSs. Firstly, for adult learners such as those who will form the basis of this study, research into individual differences in SLA has concluded that ultimate attainment, “the outcome or end point of acquisition”, differs greatly (Birdsong, 2004: 82). In fact, when studies have focussed on high-achieving adult L2 learners (see Coppetiers, 1987; Selinker, 1972), they have often been used to emphasise the *infrequency* with which complete mastery or native-like proficiency is reached (see Han, 2004; Wray, 2002). With adult L2 learners invariably differing in their language achievement (Schachter, 1996), it is clear that simply aiming to be native-like, does not mean that the target will be met; most learners will never be comparable to NSs in their accuracy, fluency and creativity (Graddol, 2006; Medgyes, 1992) nor will they reach the point at which they become “indistinguishable from native speakers” (Selinker, Swain & Dumas, 1975: 139).

Secondly, adopting NS proficiency as a model for success in English can disregard the context surrounding L2 learners’ language usage. Since L2 speakers of English greatly outnumber NSs (Canagarajah, 2007; Modiano, 1999) and since the majority of communication in English occurs without the presence of a native speaker the NS model can be irrelevant and superfluous for learners using English as a language for communication (Kramsch, 2003). As a lingua franca, “a contact language between persons who share neither a common native tongue nor a common (national) culture” (Firth, 1996: 240), judging oneself as successful according to NS norms may not only be unnecessary, but alienating both *linguistically*, i.e. in terms of a ‘standard’ English, and *socioculturally*, i.e. in terms of learner identities and L2 attitudes (see Ellis, 1994; House, 2003; Kachru, 1992a; Kachru, 1992b; Phillipson, 1996; Rampton, 1990; Seidlhofer, 2005; Widdowson, 1994). Ultimately, a model for success may be suitable for “*most*” situations but it will not apply to “*all*” situations (Kachru, 1992a: 52). The model chosen should thus be applicable to the students’ needs and contexts.

Finally, the appropriacy of the NS model of success becomes increasingly disputable given that it is probably impossible for adult learners to ever attain the standard often given in definitions of NS speakers. Not only can this realisation make success unreachable for learners targeting NS levels, but it can result in them being forever treated as language learners rather than the language users they are (Firth & Wagner, 1997). Put simply, a NS is a speaker of the language that was learnt first (Bloomfield, 1933) or they are a speaker of a language used since early childhood (McArthur, 1992). They are also often assumed to be monolingual, monocompetent language users (Milambiling, 2000).

The impact of these statements is that the inclusion of a criterion dependent upon place of birth or first language seems instantly excluding of adult L2 learners. Before they embark on the process of acquiring a TL, they are discounted and prevented from ever achieving NS proficiency: “you can no more change the historical fact of which language you spoke first than you can change the mother who brought you up” (Cook, 2008: 171). More detailed definitions of NSs offered by Chomsky (1965) and Stern (1983) place further distance between NSs and adult L2 learners as they are said to i) possess an innate ability to learn a language, and ii) to develop an intrinsic capacity to exploit structures freely and creatively without undue attention in order to satisfy the linguistic demands placed upon them. The NS can therefore be seen as an unmatched “ultimate standard” (Chomsky, 1965: 21), a gifted individual who enjoys an “uncontroversial privilege of birth” (Kramersch, 2003: 254). The NS model, rather than being the standard for success, instead becomes a “utopian, unrealistic” goal (Alptekin, 2002: 57), a type of “linguistic unicorn” (Edge, 1988: 154) eluding most adult L2 learners. As a model, such strict definitions create the impression that learners will forever be seen as failures or outsiders regardless of the proficiency level achieved or their current stage of development (Graddol, 2006; Widdowson, 1994).

It must be acknowledged, however, that the NS definitions offered above do not acknowledge the inconsistencies which make the notion of an English native speaker “confusing, misleading and a cause of fuzziness” (Andreou & Galantomos, 2009: 202). NSs do not always demonstrate a ‘standard’ use of language nor a consistent level across all of its modes; they may be mono- or multilingual; they may not originate from traditional ‘inner circles’; and they may also display much variation in their registers and styles meaning that a sole, ideal NS, as posited by Chomsky (1965) does not exist (Andreou & Galantomos, 2009; Cook, 1999; Kachru, 1992a; Kramersch, 2003; Lee, 2005; Lyons, 1996; Rampton, 1990). For instance, a particularly bold but valid point by Andreou and Galantomos (2009: 202) illustrates this argument when they question “who can be considered to be the ideal native speaker of, let’s say, Italian, a college professor, an accountant or a bus driver?” It would seem that just as non-native speaker language is an imitation of or deviation from native language, it would be equally permissible to query whether all NSs of a language represent an equal, uniform proficiency such as that provided in teaching materials: most typically, the model of the educated NS (Seidlhofer, 2001). Indeed, if “variable characteristics” (Cook, 1999: 186) such as region, occupation, generation and social class (see Andreou & Galantomos, 2009; Kramersch, 2003; Saniai,

2011) do not qualify or disqualify individuals from being native speakers, questions could be asked as to whether individual differences should prevent *learners* of a language from achieving a comparable standard. It also casts doubt as to whether a single native model should be used as a target and measure for learner success.

Relating the previous paragraphs to the preliminary rationale for examining success, it is clear that the NS goal of many learners can seem unrealistic, sometimes irrelevant and unattainable for many adult learners of English. Relating them to personal teaching experience, such models also have the potential to undermine students' perceptions of their own L2 abilities and have led to the impression that learner language was in some way a "degenerate form" of the target language (Bley-Vroman, 1983: 4). As has been shown, however, the fact remains that many learners still consider the NS a good or ideal target to aim for, a sentiment which on balance is also supported by the author. As a result, although this study aims principally to *examine* successful L2 language rather than generate a *model* for it, it will do so in accordance with the belief that such language does not have to originate from NSs alone. It is the learners themselves who can also provide models for each other and whose language can supplement NS models and teaching materials with more realistic, more achievable goals. Taking this perspective as a basis, the following section presents additional motivation for this study. It highlights that there is still a need to lend attention to the different stages reached by learners as they progress in second language learning rather than to the 'end-product' if success is to be determined adequately. Once again using the comparative fallacy as a platform for discussion, it demonstrates how tools such as the *Common European Framework of References for Languages* [CEFR] (Council of Europe [CoE], 2001), whilst instrumental in documenting what learners can do versus what they cannot, still require further enhancement if they are to be truly useful in defining success at different levels.

1.3 Success at different levels

As already discussed, the comparative fallacy results in comparisons of learners against NS definitions and abilities. Personal experience of reading SLA literature suggests that these comparisons, whether intentional or unintentional, explicit or inferred, tend to cast a NS shadow over L2 learners and their efforts (Cook, 2008; Firth & Wagner, 1997; Kachru, 1992a; Prodromou, 2008). They can create a "deficit relationship"

(Prodromou, 1997: 439) and rather than underlining a learner's success in using an L2, they can heighten their sense of failure towards L2 development and the NS model itself (Cook, 2002; 2008). As is remarked by Naiman et al. (1978: 2):

“...failure is accompanied by dissatisfaction, awareness of one's own inadequacy, and sometimes annoyance, disappointment, frustration, and even anger at the colossal waste of time.”

The continual pressure to replicate the NS not only creates “stereotypes that die hard” (Nayar, 1994 cited by Alptekin, 2002: 60), but it potentially discourages learners from persevering with the acquisition of the TL. It also overlooks the potential that could be unlocked by viewing learners not as “failed native speakers” but instead as “successful multicompetent speakers” of more than one language (Cook, 1999: 204).

Of vital importance to this study, however, is the barrier created by the NS model. By focussing on the overall goal of L2 development, little assistance is offered to learners whose goals for success may simply be to develop their interlanguage to become more proficient in English than they currently are. For instance, the NS model cannot be fully relevant to their needs since in representing the ‘finished article’ it cannot demonstrate the nuances in interlanguage between the beginner, pre-intermediate, intermediate, upper-intermediate and advanced learner levels. Whereas this realisation has prompted some writers to propose alternative models of greater relevance to L2 learners’ multicompetencies and language learning strategies (see Alptekin, 2002; Cook, 2008; Coperías Aguilar, 2008; Edge, 1988; Medgyes, 1992; Modiano, 1999; and Preston, 1981 for alternative models; see also Cook, 1992; Cook, 1999; Coppieters, 1987; Galambos & Golin-Meadow, 1990 for multicompetence and O’Malley & Chamot, 1990 for language learning strategies), relatively few studies have sought to elaborate on learner proficiency levels and what makes the learners within them *successful* in their own right. By focussing on learners, this study therefore intends to address the imbalance caused by the dominance of the NS model by establishing not what learners of English are *unable* to do at different stages, but rather what they *are able* to do in terms of their spoken production. In accordance with the levels and descriptors offered in the CEFR (CoE, 2001) it will detail how success manifests itself in the speech of B1, B2 and C1 learners and how their interlanguage progresses in ways that might not be recognised in some syllabi and textbooks.

1.4 Statement of research aims

This study aims to examine what makes B1, B2 and C1 learners successful in their spoken interactions according to the CEFR. Through the construction of a learner corpus of spoken data, it will establish which lexical features are typical of learners at each of the three proficiency levels, and how they differ across levels. As has been recently stated, such enquiry is not simply concerned with describing levels, it should also concentrate on describing progression across them (English Profile, 2016). In a similar vein, it aims to reveal the nature of the collocational and colligational patterns employed by learners in their speech and how such lexis can satisfy relevant can-do statements from the CEFR. Though highlighting methodology can be somewhat unorthodox at this stage of a thesis, it is important to acknowledge that a range of methodological options do exist for researchers investigating learner language success. However, very few offer the same advantages afforded by learner corpus research which, using real samples of learner language as a basis for frequency and pattern analyses, objectively extends understanding of language use in ways that lie beyond the parameters of researcher intuition (McEnery & Xiao, 2010). By not focussing on the errors occurring in real learner language and by not highlighting spoken can-do statements which have not yet been evidenced or mastered, this study thus intends to demonstrate what successful B1, B2 and C1 learners are able to do in their own right and not what they ought to be able to do according to NS targets.

Ultimately, the research intends to complement studies such as English Vocabulary Profile (Cambridge University Press [CUP], 2015a) and English Grammar Profile (CUP, 2015b) by constructing a corpus which will add more detail to the objectives detailed in the CEFR, and which will provide “more realistic examples of language use” (McEnery, Xiao & Tono, 2006: 98) for learners requiring more attainable targets than those provided by the native speaker. Just as corpora can collect NS examples of language, by obtaining real examples of learner language, it is hoped that this study will support the current change in attitudes where language learners are no longer judged according to “unrealistic criteria” (Weir, 2005a: 107) which have no real bearing on their spoken linguistic capabilities or spoken success. It will respond to calls demanding that learner interlanguage be recognised not as an inferior system of a language user’s first or second language, but as its own system (Tenfjord et al., 2006 cited by Gilquin, De Cock

& Granger, 2010): a system capable of showing what students can do with their language use as opposed to a system used to highlight what they cannot.

1.5 A focus on speaking

Speaking has been chosen as a focus for various reasons. Considered “the most important” skill (Foot, 1999: 36), speaking is now given a certain prestige in linguistics and is considered the skill to which language teaching is directed (Underhill, 1987). As the “primary form” of language, it also lays the foundation for other modes (e.g. written language) and provides the source from which language evolution and change stems (Hughes, 2011: 14). The second reason is that despite this theoretical ‘reverence’, it is occasionally “undervalued” as a skill (Bygate, 1987: vii) seen, sometimes, as problematic due to the difficulty in isolating it entirely from other skills (Brown, 2004) and researched rather less frequently (Hughes, 2011; Weir, 2005a). Furthermore, as speech often occurs in real-time communicative, interactive settings, it is a multifaceted and demanding mental process (Ellis, Simpson-Vlach & Maynard, 2008; Taylor, 2011). A speaker cannot attend solely to the process of constructing meaning. They have to make decisions spontaneously about what they will say, how they will say it, the effect it may have on the interlocutor and the demands made of them in an evolving communicative encounter (Bygate, 1997; Hughes, 2011; McCarthy, 1998; Taylor, 2011). It also involves extremely rapid decision-making, the negotiation of unexpected problems, the following of communication norms, the maintenance or construction of relations and the handling of turn-taking and time pressures. As is remarked by Levelt (1989: 199), “[t]here is probably no other cognitive process...whose decision rate is so high”. Finally, and possibly most importantly for the topic of success, speaking represents the skill which is “most frequently judged” (Bygate, 1987: vii), a perhaps unsurprising fact given that the majority of language use is spoken rather than written (Lewis, 2012). All of these challenges combined can create obstacles for individuals speaking in their first language. For learners wishing to be successful and for those wishing to be judged as successful in an L2, the skill of speaking would seem to be, therefore, the most obvious starting point in a study of learner success.

1.6 The research questions

In order to achieve the targets specified above, the following research questions [RQs] will be answered:

RQ1) What percentage of the words used by successful B1, B2 and C1 learners come from the first 1000 and second 1000 most frequent words in English?

RQ2a) What were the 20 most frequent words at B1, B2 and C1 and their notable collocations, colligations and functions?

RQ2b) What were the 20 most frequent keywords at B1, B2 and C1 and their notable collocations, colligations and functions?

RQ2c) What were the most frequent 3- and 4-word chunks at B1, B2 and C1 and their notable collocations and functions?

RQ3) What Common European Framework of Reference for Languages (CEFR) indicators are present in terms of spoken interaction, spoken production and strategies at B1, B2 and C1 and how are they realised?

1.7 Chapter conclusion and overview of thesis

This opening chapter aimed to introduce the NS's once overbearing influence on learner success and the revisions that prompted a shift away from language models emphasising the 'end-product' to those stressing the incremental stages in the second language learning process. It also presented the aim and research questions of the study and explained the specific focus on learner speech. Chapters 2 and 3 continue by exploring the literature which has contributed and shaped the notion of success in second language learning. The former revisits theories of communicative competence so as to provide a platform for discussion of the CEFR's position on learner competence and proficiency in a target language. The latter presents the field of corpus linguistics and the findings it has generated which are of direct relevance for learner speech and language achievement so as to demonstrate the unique advantages the method affords researchers. Chapter 4 then documents the pilot study that was conducted using a corpus of C1 learner speech. Detailing the initial procedures followed and the results obtained, it will also

explicate the changes that were made to the main study, procedures for which are discussed in full in Chapter 5. Chapter 6 then presents the results and discussion which answer the questions detailed in the previous section. It draws together the findings per research question and clarifies their relevance to previous literature and to the notion of learner success in speech. The conclusion in Chapter 7 will draw together the main results found in this study. It will highlight the study's implications for language learning, teaching and for future research into learner success in speech, and will document the limitations of the investigation. The bibliography and appendices are then provided in Chapters 8 and 9.

2. LITERATURE REVIEW (PART ONE): COMMUNICATIVE COMPETENCE AND GAPS IN THE CEFR

2.0 Introduction

The opening chapter introduced the concerns and gaps generated by previous attempts to identify suitable models of second language learning. It juxtaposed the NS model's popularity and characterisation of the 'finished article', with the developing language learner's ability to model achievable interlanguage progression at different stages. Though such models allow evaluations and descriptions of language progression to be generated, they neglect a fundamental question that must be addressed if learner success is to be fully understood: "What does it mean to know a language?" (Spolsky et al., 1968: 79). The matters of knowing and using a language are, of course, very different, but learner success cannot be fully examined if one does not know what knowledge is being evidenced in speech. The purpose of this chapter is therefore to highlight the communicative competences required if learners are to use language adequately and appropriately within their specific communicative contexts and to show how these are measured or categorised by the CEFR.

To do so, it first reviews Hymes (1972), Canale and Swain (1980), Canale (1983) and Bachman and Palmer's (1996) theories of communicative competence and language use so that their composition and relevance to success can be fully comprehended. Following an examination of communication and the production and interaction skills integral to its fulfilment, the chapter then culminates with a consideration of the CEFR's position of communicative competence and its implications for investigating learner speech.

2.1 Theories of communicative competence and their impact on learner success

2.1.1 Hymes' theory of communicative competence

The first theory of communicative competence, created by Hymes (1972), arose from dissatisfaction with Chomsky's (1965) competence-performance distinction. Adopting a nativist stance, Chomsky viewed language not as a tool for communication, but instead as a tool for cognitive growth. He posited that to possess a language was to

have available “a set...of sentences, each finite in length and constructed out of a finite set of elements” (Chomsky, 1965: 13). These elements, acquired via a language acquisition device during first language acquisition [FLA], are then governed by a person’s Universal Grammar which determines the production of language (Chomsky, 1982). Clearly, language production here was associated only with the well-formedness of utterances in the first language: a quality reliant upon a monolingual speaker-hearer’s innate sense of accuracy as opposed to their experience using the language (Chomsky, 1965).

Hymes therefore reacted to the stringent polarisation of competence and performance in this theory. Put simply, Chomsky’s notion of competence referred not to a person’s ability to *use* language, as the term might infer, but to their tacit *knowledge* of language (Spolsky, 1989). As an inaccessible, “static” entity (Taylor, 1988: 153), or “deep-seated mental state” (Widdowson, 1989: 129), competence could be utilised to create meaning, but could not be fully reflected in performance, “the actual use of language”, due to the psycholinguistic influences influencing it (Chomsky, 1965: 4). For Hymes, the two aspects cannot be separated so easily. In a sense, whereas Chomsky’s theory facilitated the application of linguistic rules for generating sentences, it neglected to address how a language’s grammar may be adapted to suit communicative contexts and situations (Campbell & Wales, 1970; Canale & Swain, 1980; Hymes, 1972). It essentially failed to acknowledge the abilities and linguistic norms “without which the rules of grammar would be useless” (Hymes, 1972: 278), since being able to use language appropriately in a range of situations is equally as important as the accuracy of the grammar underlying it (Lyons, 1970). With Chomsky’s theory of competence being seen as little more than a theory of grammar (Canale & Swain, 1980), Hymes sought to recognise the features of language that made it not only *well-formed*, but *acceptable* as well.

In this renewed attitude to competence, Hymes challenged the notion that sociolinguistic aspects were viewed as performance-related imperfections (Hymes, 1972; Llorca, 2000; Taylor, 1988). Features of performance relating to social, situational or functional demands were previously associated with grammatically irrelevant performance flaws, resulting in performance being an all encompassing term for features not described by competence. Hymes (1972) makes the key distinction that sociocultural aspects of language use *also* symbolise a type of competence that develops during FLA. Governed by their own set of rules and systems, social experiences not only contribute to

performance, but also to a language user’s internal knowledge of their first language. Hymes highlights that an individual’s knowledge and the ability to use knowledge of the language in performance are linked and are of equal worth to descriptions of language:

“We have then to account for the fact that a normal child acquires knowledge of sentences not only as grammatical but also as appropriate. He or she acquires competence as to when to speak, when not, and as to what to talk about with whom, when, where, in what matter. In short, a child becomes able to accomplish a repertoire of speech acts, to take part in speech events, and to evaluate their accomplishment by others...”

(Hymes, 1972: 277)

The key argument in this theory, therefore, is that linguistic and sociocultural competences are not separate entities with the latter ‘grafted’ onto the former, nor is sociocultural competence irrelevant to a language user’s overall language competence. Similarly, competence no longer remains a static state of knowledge due to its involvement in the process of creating and comprehending meaning in a range of communicative contexts (Ellis, 1994). Competence instead embodies the “capabilities of a person”, dependent both on tacit knowledge *and* the “ability for use” (Hymes, 1972: 282) (see Fig. 1). It incorporates four distinct qualities: that language is *formally possible*, in terms of grammar, culture or communication; *feasible*, regarding the implementation of language according to psycholinguistic and cultural factors; *appropriate*, involving awareness of contextual features and tacit knowledge of sentences and situations; and *done* in the sense that the forms *do* actually occur in a language (see Hymes, 1972: 281-286).

Figure 1: Hymes' (1972) theory of communicative competence

Model of knowledge	Model of performance	Actual use
Communicative competence		Performance
Knowledge	Ability for use	

Hymes’ (1972) theory does much to extend the concept of competence in communication and its applicability to second language learning and success. For instance, as a theory fundamental to social interaction (Paulston, 1992), it stipulates that grammatical accuracy alone is insufficient in the learning and use of a language. Although focussing on FLA, similarly to Chomsky, it advocates that language users need

knowledge of the L2's grammar and its use in production. For adult L2 learners, this means that the learning of grammar rules alone would not suffice; they have to also be capable of adapting language to suit its mode, audience, genre and context. Furthermore, by recognising that language users possess "differential knowledge of a language", emphasis is detracted from the ideal-speaker hearer proposed by Chomsky (Hymes, 1972: 270). It supports the view that individuals, whether native or not, do not behave identically; neither is their language use always entirely accurate. The appearance of errors in learner language may, therefore, not be irrelevant to discussions of competence. For language learners with incomplete and evolving competence of the target language (Ellis, 1990), errors in performance may, to some extent, expose their current interlanguage state (Selinker, 1972). Errors in learner performance should not be immediately discounted as a feature of performance as they offer vital indicators as to learners' current L2 knowledge.

Hymes' theory ultimately asserts that neither linguistic nor sociocultural competence can thrive without the other. A speaker with high grammatical accuracy but inappropriate language will be equally as unsuccessful in communication as a speaker with high appropriacy but low grammatical accuracy. However, as valuable as Hymes' theory has been, it has still been subject to criticism. Firstly, by changing the notion of competence from a state of knowledge to an ability, concern exists as to the validity of the understanding and criticisms underpinning the revision of Chomsky's theory (Taylor, 1988; Widdowson, 1989). More significantly, the lack of precision involved in the modification of competence results in its 'growth' into a term encompassing more than one element of language thus weakening its comprehension (Taylor, 1988; Widdowson, 1989). For instance, the lack of detail in illustrating the composition and interplay of its components complicates its application to real-life teaching and language use (Canale, 1983; Canale & Swain, 1980; Celce-Murcia & Dornyei, 1995; Widdowson, 1989). Specifically, with regards to the implications of its four elements for L2 learners and teaching, some ambiguities remain: it is unknown whether L1 competence shapes or affects the learning and use of an L2, a second language learner may not have knowledge about what is feasible or 'done' in an L2 and likewise, little expansion is offered as to potential changes occurring as L2 language experience grows (see Brumfit, 2001 for more detailed discussion). Finally, though Chomsky deliberately omits language use from his definition of competence, Hymes' theory potentially risks emphasising language in use at the expense of grammatical accuracy due to the omission of grammar in key terms

(Widdowson, 1989). Clarification is therefore needed of their exact relationship so that its application to language teaching can be determined. There is hence a clear need for a theory, explicit in its construction, of communicative competence based on the second language learner rather than the native monolingual. To understand what makes learners more or less successful, a theory clarifying how competence is utilised in performance is also key.

2.1.2 Canale & Swain's theory of communicative competence

Section 2.1.1 has so far documented the progression of competence as a theory of grammar to a theory of ability for use. The shift from concentrating attention on a monolingual NS's innate knowledge of grammar to the combination of linguistic and sociocultural competences has shown that relating terms to L2 learners' competence has been problematic. Whereas Chomsky's definitions are difficult to apply to non-native adult learners, Hymes' explanations do not clearly describe the exact nature of competence and the interactions involved in making language knowledge available for use. Despite communicative competence assuming "buzzword" status in the language learning field (Canale, 1983: 2), communicative competence theory still required development. Canale and Swain's (1980) theory of communicative competence thus aimed to clarify terms, propose a new model building on Hymes' work, and, in stark contrast to Chomsky and Hymes, place the language learner at the heart of discussion.

The omission of explicit reference to grammatical competence in Hymes' theory led Canale and Swain (1980: 5) to identify a significant trend in the second language field: that of communicative competence being linked "exclusively" to ability for use resulting in the assumption that it was only a theory of performance. To address misconceptions that the two types of competences are distinct and should be taught separately, Canale and Swain instead stipulate that grammar should be reinstated in definitions and models of communicative competence:

"Just as Hymes (1972) was able to say that there are rules of grammar that would be useless without rules of language use, so we feel that there are rules of language use that would be useless without rules of grammar."

(Canale & Swain, 1980: 5)

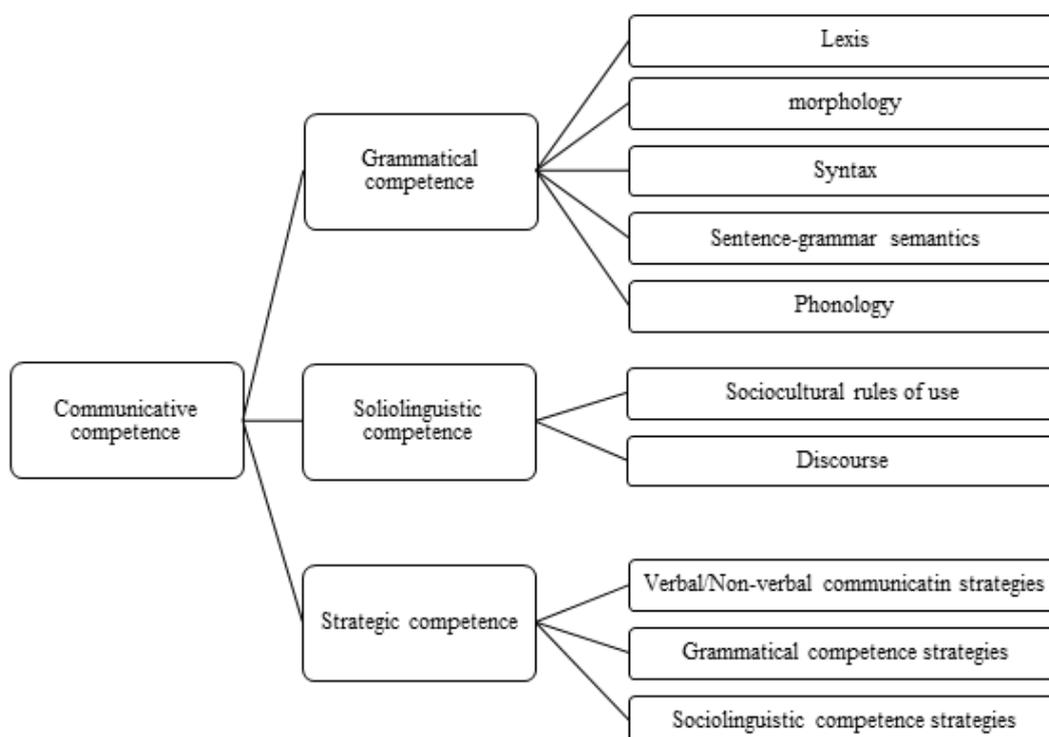
They stress that grammatical, sociolinguistic and strategic competence are of equal importance to communicative competence which is thus reliant upon the “relation and interaction” of all its elements (Canale & Swain, 1980: 6). Communicative performance, on the other hand, is defined as:

“the realization of these competences and their interaction in the actual production and comprehension of utterances (under general psychological constraints that are unique to performance)...[it is the] actual demonstration of this knowledge in *real* second language situations for *authentic* communication purposes.”

(Canale & Swain, 1980: 6)

As is displayed in Figure 2, Canale and Swain’s theory comprises three distinct elements: *grammatical* competence, *sociocultural* competence and *strategic* competence, a new addition based on communication skill research focussing on the oral skills learners need to “get along in...or cope with” most of the situations they are likely to encounter (Canale & Swain, 1980: 9).

Figure 2: Canale & Swain's (1980) theory of communicative competence



Grammatical competence concerns knowledge of lexis, morphology, syntax, semantics and phonology which merge with L2 pedagogy to help learners “determine and express accurately the literal meaning of utterances” (Canale & Swain, 1980: 30). Sociolinguistic

competence, on the other hand, relates more to the non-literal traits of language use involved in making sense of the grammar of an utterance and the language user's intentions. It deals specifically with i) the characteristics involved in topic, participant role and interaction norms, and ii) the employment of grammar, attitude and register that are required for communication to be seen as *appropriate*, a case in point being a university student using imperative forms to demand that a tutor offer an assignment deadline extension and then using 'ta' rather than 'thank you' if it is granted. Similarly, sociolinguistic competence relies on rules of discourse, examinations based on the cohesion and coherence of utterances on a collective rather than an individual basis; put simply, language as a whole rather than isolated utterances. Whereas judgements of accuracy can report on utterances *separately*, discourse incorporates the way utterances *combine* to complete communicative transactions appropriately. Finally, strategic competence plays a role in communication breakdowns attributed to a lack of competence or the influence of performance variables (Canale & Swain, 1980). Relating either to obstacles facing grammatical or sociolinguistic competences, they help learners to counteract problems they contend with during communication. Although these three features compose Canale and Swain's (1980: 31) theory of communication competence, the writers are keen to emphasise one over-arching condition affecting each component: the contingent existing on "probability rules of occurrence". Perhaps similar to Hymes' condition necessitating knowledge of whether a linguistic term is 'done', communicative competence in the target language will only be achievable if the learner acquires knowledge as to whether linguistic features *actually occur* in the second language.

Canale and Swain's (1980) theory has significant implications for second language learning success. Their theory firstly shifts focus away from FLA and the monolingual NS to concentrate on L2 learners, in particular within the fields of L2 teaching and testing. In doing so, it helps to achieve their aim of establishing a "more valid and reliable measurement of second language communication skills" (Canale & Swain, 1980: 1) which was denied by Chomsky and Hymes' theories. Secondly, despite Hymes' assertions that grammar was encapsulated in his theory, Canale and Swain readdress the perceived imbalance between grammatical and sociolinguistic competence. Though learners are, in fact, judged by their language *use* rather than their use of *grammar* (Widdowson, 1978), this amended theory of communicative competence affirms that overall judgements of success are equally reliant on both appropriacy *and* accuracy in communication transactions. Another distinguishing feature of the model is that, in

contrast with the two prior theories, gaps in grammatical and sociocultural competences can be minimised by strategic competence. Learners will encounter a range of communicative obstacles but strategic competence may still allow learners to communicate efficiently and successfully. Since Canale and Swain (1980: 30) remark that communication strategies can be “called into action”, the teaching of such strategies would presumably be of benefit to learners participating in communication. Not only would learners be able to manoeuvre around particular difficulties, but the teaching of strategies would also “prepar[e] second language learners to *exploit*” language (Canale & Swain, 1980: 29 [emphasis added]) meaning that they could make the most of their available resources. Finally, however, it must be noted that as a ‘demonstration of knowledge’, Canale and Swain are keen to state that performance does not exhibit a speaker’s competence; it can provide abstractions perhaps allowing for features to be deduced, but communicative competence can still not be measured directly. In terms of learner development and success, the production of an accurate and appropriate utterance, for instance, can only provide evidence that it has been produced adequately on this “particular occasion”, rather than it providing evidence of complete mastery (Lewis, 2012: 10).

The revisiting of communicative competence theory has emphasised the role of the second language learner and the way that gaps in competence can be ‘by-passed’ with the use of strategies. However, it still receives close scrutiny. Just as Hymes’ theory was criticised for a lack of explanation illustrating the interplay between grammatical and sociocultural competence, Canale and Swain’s theory has also been deemed underdeveloped in this manner. Though a more evident attempt is made to show how the three competences are not entirely separate from each other, it has been judged “more as a list than a model” (Baker, 2015: 137); likewise, the placement of grammatical competence first in the ‘list’ may lend it a heightened sense of significance in comparison to sociolinguistic and strategic competences. For success, though Canale and Swain (1980: 27) assert that grammatical competence cannot be viewed as “more or less crucial to successful communication”, such organisation could suggest otherwise. Correspondingly, if this were the case, the components of grammatical competence would require further clarification. As Celce-Murcia (2008) identifies, there remains a lack of specification as to the role of lexis, in particular, the role of formulaic language (to be discussed in Section 3.3.2). For instance, an L2 beginner in English may be quickly introduced to communication strategies such as ‘*could you repeat that?*’ and *could you*

say it slower please?” without them knowing the syntax underpinning them. Such strategic competence, therefore, could facilitate success in the short-term, but it might only be in the long-term that such questions form part of grammatical competence. Furthermore, it raises doubt as to whether strategic competence constitutes language knowledge or, inadvertently, whether it constitutes ability for use (see Shohamy, 1996).

The final aspect for success concerns the role of the *learner* in Canale and Swain’s theory of communicative competence. For language teaching goals, it is unknown whether the theory serves temporarily, until a point has been mastered, or permanently to describe past uses of language, or whether they ‘prescribe’ future usage (see Brumfit, 2001). With this in mind, Brumfit (2001: 52) argues that language users and learners become “passive victim[s] of the inherited rule systems”. Learners have little bearing therefore on their own development of communicative competence and teaching, which insinuates that a more dynamic model of language knowledge would be called for. Finally, by declaring that grammatical, sociolinguistic and strategic competences are reliant or influenced by ‘probability for use’, Canale and Swain (1980: 27) advocate the need for “meaningful communicative interaction with highly competence speakers of the language” and authentic teaching texts. This indicates that despite language learners’ needs being foregrounded, the targets to which they should aim once again associate themselves with the NS (Baker, 2015). As Chapter 1 identified, however, a NS model would be unable to demonstrate the differences between B1, B2 and C1 learners. Hence the question arises as to whether ‘highly competent speakers’ could in fact be made a relative concept. In doing so, the basis of this study is that learners may indeed represent the only source of truly authentic language when models for high competency at B1, B2 and C1 are sought.

2.1.3 Canale’s theory of communicative competence

Sections 2.1.1 and 2.1.2 have identified the slow shift in communicative competence theories from discussions centring on the NS to discussions of language learner competence. Though learners have been moved to the forefront of communicative theory, the models have so far presented a lack of relevance to the potential of using learner language as models of success. Theories have also been criticised i) for an absence of explicit detail regarding how the proposed components of communicative competence interact, ii) for insinuations that learners remain passive in the communicative

competence process, and iii) for confusion as to the role of formulaic lexis in grammatical competence and the nature of strategic competence as knowledge or ability. Of particular relevance to this study is the absence of detail regarding how competence is actually used in performance. This section presents Canale's (1983) theory of competence. In essence, as will be seen shortly, it bridges the move from theories of competence towards models of language ability.

Responding to criticisms of Canale and Swain's theory, once again Canale attempts to clarify and expand on some of the ambiguities that arose. In reaction to assertions that the theory included a concept similar to Hymes' ability for use (see Shohamy, 1996) despite their overt reluctance to do so (see Canale & Swain, 1980: 7), Canale (1983: 5) acknowledges that their theory referred, alternatively, to a communicative competence incorporating "underlying systems of knowledge and skill required for communication." The knowledge fundamental to communicative competence, or likewise the *declarative* knowledge of 'knowing about' language is therefore accompanied by skill, the *procedural* knowledge concerning the extent to which knowledge 'can be performed' or put to use in 'actual communication': the new term created for performance in order to avoid confusion with Chomsky's 1965 definition (Canale, 1983; McNamara, 1995). Transitioning from knowledge-oriented approaches of language teaching, Canale asserts that adopting a more skill-oriented approach is a much-needed shift if students are to learn how to employ such knowledge adequately:

"such [knowledge-oriented] approaches do not seem to be sufficient for preparing learners to use the second language well in authentic situations: they fail to provide learners with the opportunities and experience in handling authentic communication situations in the second language, and thus fail to help learners to master the necessary skills in using knowledge."

(Canale, 1983: 15)

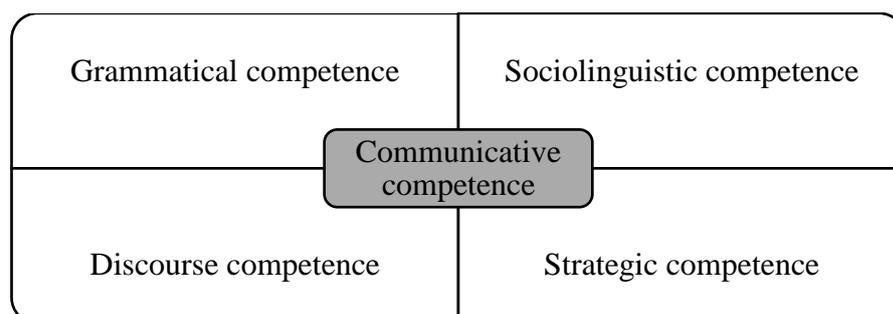
For learners of a second language aspiring to be seen as successful, competence is therefore not only reliant upon *having* knowledge but also being exposed to it in real situations and having the means to *utilise it*. As Lewis (2012: 33) remarks, knowledge is necessary but "What matters is not what you know, but what you can do. 'Knowing' a foreign language may be interesting; the ability to use it is life-enhancing." With learners clearly varying in different levels of knowledge and in the skills to execute it, this means that communicative competence could be a quality that changes as target language

experience develops. Furthermore, a learner with a high degree of knowledge may, in fact, be considered unsuccessful if they are not able to put that knowledge into action whereas a learner with less knowledge, may still be able to exploit it to some extent.

Canale's (1983: 7-11) theory of communicative competence (illustrated in Figure 3), thus contains four key areas of knowledge and skill, briefly summarised here:

- *Grammatical competence*: “concerned with mastery of the language code” and once again with the comprehension and creation of literal meaning grammatical competence relates to vocabulary, word formation, sentence formation, pronunciation, spelling and linguistic semantics.
- *Sociolinguistic competence*: involves the “extent to which utterances are produced and understood *appropriately* in different sociolinguistic contexts depending on contextual factors.” With appropriacy thus relying on the form and meaning (of functions, attitudes and ideas), sociolinguistic competence displays a contrast with Canale and Swain's theory in that it no longer encompasses discourse rules.
- *Discourse competence*: “concerns mastery of how to combine grammatical forms and meanings to achieve a unified spoken or written text in different genres.” As a separate entity, discourse competence refers to the structural connections enabling interpretation of a text, or *cohesion*, and the interaction of meanings among texts conveying the literal, functional and attitudinal, or *coherence*.
- *Strategic competence*: relating to the “mastery of verbal and non-verbal communication strategies” allows for i) the compensation of communication breakdowns which occur due to a lack of competence or due to “limiting conditions”, and ii) the boosting of communicative “effectiveness” relating to the context and function of language i.e. rhetorical effect, a new addition to the previous definition of strategic competence in Canale and Swain.

Figure 3: Canale's (1983) theory of communicative competence



As an outline of what competence should, as a minimum, contain, less emphasis once again (as criticised earlier by Widdowson, 1989) is placed on how these four areas interrelate both with the psychological and environmental factors of communication, and with each other (Canale, 1983). Essentially, it still presents a ‘list’ of competences but omits the interactions between them. Although its criticisms and repercussions for success have not been detailed in any great depth here, this theory has been included as it represents a transition in communicative competence theory. As the next section will reveal, it represents a renewed attempt by theorists to describe more explicitly how knowledge is utilised in communication.

2.1.4 Bachman and Palmer’s model of language use and performance

This chapter’s introduction stressed that full understanding of second language learning is dependent not only upon the identification of an appropriate model or target, but also upon the identification of what needs to be learnt. Though previous theories have illustrated a shift from purely knowledge-oriented models of competence to models of communicative competence acknowledging the roles of sociolinguistic, discourse and strategic competence, a gap still remains. This gap pertains to how competence is realised in performance; to paraphrase Lewis (2012), to show ultimately what can be *done* with language knowledge. Despite assertions that performance cannot offer a true or complete reflection of learner competence, the final theory to be presented here, Bachman and Palmer (1996), outlines learner competences and their involvement in language use. Since the CEFR bases its descriptions on action-oriented approaches to “language use and learning” (CoE, 2001: 9), as will be seen in Section 2.3.3, a change towards language *in use* is needed if successful learner speech is to be more thoroughly understood.

Bachman and Palmer’s (1996) model of language use and performance by no means represents a departure from the theories discussed. They commend the move beyond sentence level grammar prompted by Hymes (1972) and state that it is “essentially an extension” of Canale and Swain’s (1980) theory of communicative competence (Bachman, 2007: 54). Similarly, it clearly shares parallels with Canale’s (1983) definitions of knowledge and skill due to its integration of Bachman’s (1990) notion of communicative language ability. By stating that operating communicatively in language involves both “knowledge of competence...and the capacity for implementing, or using this competence” (Bachman, 1990: 81), it is evident that skills, or one’s ‘capacity’ for

utilising language, are fundamental to the demonstration of language knowledge in performance. For second language learner success, it once again implies that capacity for executing language use can develop within and differ across individuals.

A key principle underpinning language use and performance in this theory is the interactive nature of language production, in particular, for the context of this study, how language use occurs within interactive situations in which meaning is produced and comprehended via exchanges and negotiations:

“In general, *language use* can be defined as the creation or interpretation of intended meanings in discourse by an individual, or as the dynamic and interactive negotiation of intended meanings between two or more individuals in a particular situation.”

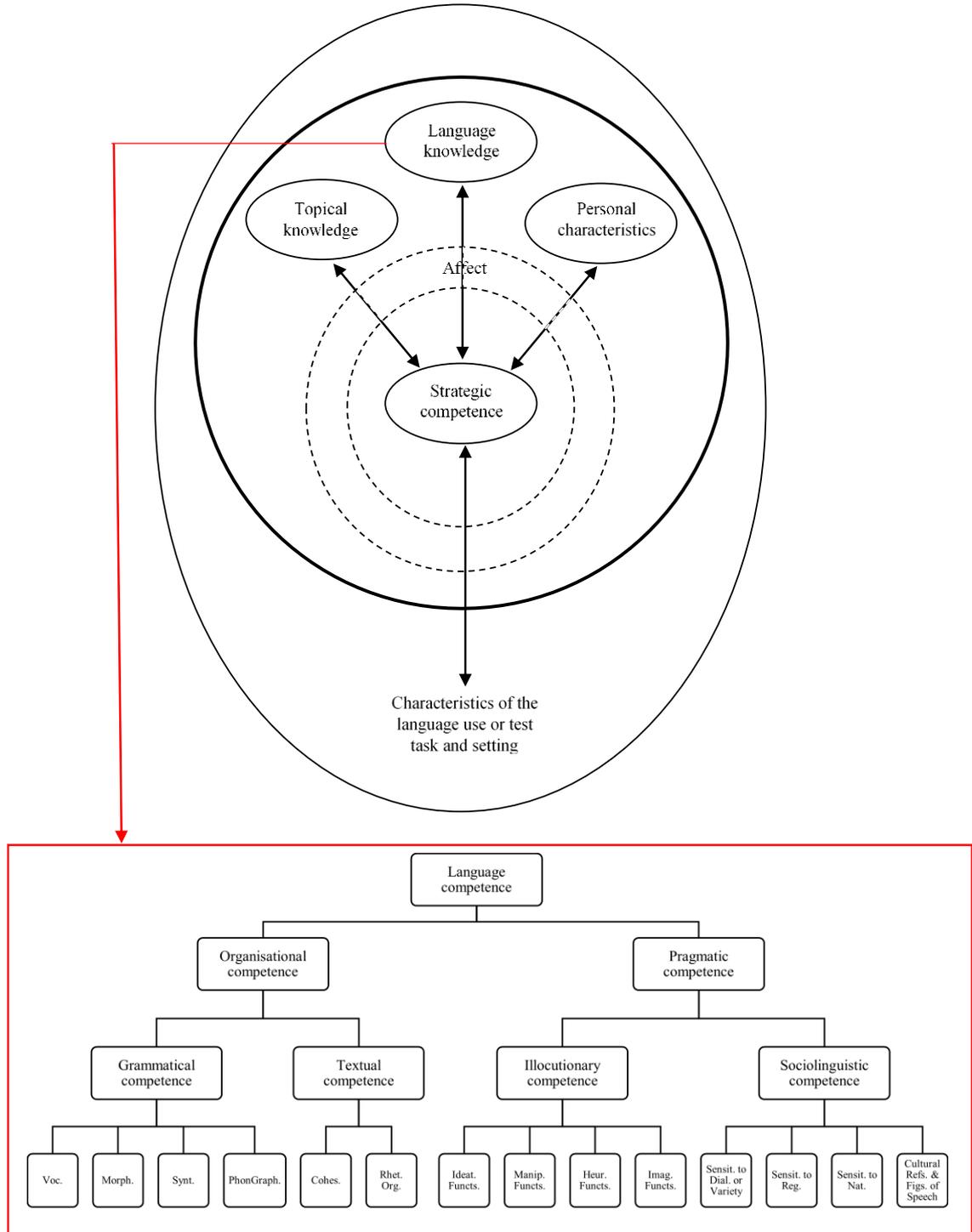
(Bachman & Palmer, 1996: 61-62)

Not assumed to be a definitive model of language ability or use but rather a model for test development (Bachman, 1990), it nevertheless resonates strongly with the rationale for choosing spoken language as a focus for examining learner success presented in Section 1.5. With speakers having to operate in real-time interactive settings, focussing solely on the construction of meaning at utterance-level will not suffice. Instead, speakers have to consider the norms dictated by the linguistic context and task, the effect upon the interlocutor and the development of utterances as part of the construction of discourse. As such, there exist many potential pitfalls for learners at the three levels investigated in this study and for second language learners in general.

As a theory useful for considering and planning test design, as well as other language uses, Bachman and Palmer’s model broadens the scope of language usage to incorporate characteristics of the speaker and the task and setting. As seen in Figure 4, this change was attributed to features other than language knowledge – personal characteristics, topical knowledge and affect – playing an important role in someone’s ability to communicate. For example, in a test situation, a learner’s display of their language knowledge may be impaired by factors affecting the individual on the day of the exam, or by a heightened sense of affect responsible for the potential reluctance to produce language. The figure demonstrates, therefore, the characteristics held by an individual (in the bold circle) – bearing some resemblance to the CEFR’s existential competence (CoE, 2001) – and how this relates to the external task or setting in which they interact. Although individual differences will not be elaborated upon, here, it is evident that factors such as mood, tiredness, personality, topic knowledge, and

willingness to attempt or adapt language use, to name but a few, will influence a speaker's overall performance.

Figure 4: Bachman and Palmer's (1996: 63) language use and performance with Bachman's (1990: 87) components of language ability



In terms of its composition, Bachman and Palmer's (1996) theory of language use and performance, as highlighted previously, rests ultimately on Bachman's (1990) theory

of language ability. They posit that the combination of two elements, language knowledge and strategic competence, bestow language users with “the ability, or capacity, to create and interpret discourse, either in responding to tasks on language tests or in non-test language use” (Bachman & Palmer, 1996: 67). In contrast to previous studies of communicative competence, these two factors are thus no longer seen as uniform in their impact. For instance, whilst Canale and Swain (1980) stress that grammatical competence is neither more nor less important than sociolinguistic or strategic competence, the latter in this theory is given a renewed role. In Bachman and Palmer’s model, strategic competence is made central to interactions between task and setting characteristics and an individual’s language knowledge. In a sense, strategic competence has been extended beyond the realm of simply compensating or accommodating language, to the heightened status of underpinning *all* language use (Bachman, 2007). With respect to performance and success, therefore, it is essential that learners’ competences in *goal setting*, “deciding what one is going to do”; *assessment*, “taking stock of what is needed, what one has to work with, and how well one has done”; and *planning*, “deciding how to use what one has” are crucial (Bachman & Palmer, 1996: 71-75); any deficit or gap in these areas, will clearly diminish a language user’s potential to impart meaning, cope with task demands, adhere to social and discourse norms, or adapt topic knowledge to the task at hand.

In their comprehensive discussion of language knowledge, another distinction with previous theories becomes evident. Previous theories kept grammar and sociolinguistic competences (discourse and pragmatics) relatively separate. Here, such competences are merged under the term ‘language knowledge’, which is split into two main components (see Fig. 4). Firstly, *organisational competence* concerns knowledge of formal structure contained within utterances or sentences which contribute to the production and comprehension of meaning at utterance and text level (Bachman & Palmer, 1996). Comprising grammatical competence and textual competence, they rely on a learner’s knowledge of vocabulary, morphology, syntax, phonology/graphology, cohesion and rhetoric organisation (a text’s “conceptual structure” and its effect on the language user (Bachman, 1990: 88)). Evident here is the move of discourse competence, as described by Canale and Swain (1980), away from sociolinguistic aspects of language to organisational aspects so as to illustrate how utterance level grammar blends with a language user’s knowledge of the building of conversation and the marking of connections across utterances.

Secondly, *pragmatic competence* refers to the “interpretation of discourse” when the formal structure of utterances is connected to the intended meanings of language users in accordance with the setting for language use (Bachman & Palmer, 1996: 69). Whereas illocutionary (functional) competence associates utterances and texts to the meanings implied by a language user, sociolinguistic competence refers to the interactions between language, the language user and the setting and context. Illocutionary competence is responsible for the additional meanings attached to utterances. For instance, the statement ‘*I’d like to know how much this vase is*’, rather than indicating desired knowledge, functions more as a polite, indirect request for information; likewise, a partner exclaiming ‘*I’m too tired to cook tonight*’ could be seen as a suggestion that they would prefer to order a takeaway or an instruction that the other person should cook. This type of competence can also be dependent upon prior knowledge or experience. Phonological aspects aside, an interlocutor listening to a tutor saying ‘*She spent one day on her assignment*’ would need prior knowledge of the student’s character and assignment mark to interpret this as a statement of fact, a statement of surprise or a statement of criticism; extending this example to involve a student as interlocutor, it could additionally act as warning for them not to do the same. Therefore, real word knowledge (ideational), learning through language (heuristic), humour and figurative language (imaginary), and getting other people to do something (manipulative) are all functions of language which can extend meaning beyond the message conveyed at utterance level. When combined with sociolinguistic aspects such as sensitivity to dialect or variety, register differences, naturalness of language and cultural references, it becomes easier to see how Bachman and Palmer’s (1996) model of language use and performance is more thorough than previous theories in terms of what, exactly, comprises communicative language ability and what makes it accurate *and* appropriate.

This theory is of significance to learner success in a number of ways. Through the use of arrows, a much more tangible attempt has been made to expose the types of knowledge required by L2 learners and the interactions occurring between them. It demonstrates how the form, or organisation, of language is of equal importance to the meanings inferred by the speaker or interpreted by the interlocutor as a result of pragmatic competence and the interactive setting. The notion that appropriacy as well as accuracy is integral to communication is thus maintained. Of particular interest to this study is the elevated function of strategic competence. Though in a study of success, it would be easy to confine this competence to compensation or communication breakdowns, the

implication is that learners require strategic competence in a broader sense. In terms of models of success therefore, the extent to which an NS can model such knowledge becomes questionable. As a final observation, turning language knowledge into successful language use obliges knowledge or experience of the L2 culture and its communicative conventions. As a tenet to intercultural communication, it is unclear whether a learner's L1 knowledge will provide adequate levels of sociolinguistic competence for them to produce and interpret language effectively.

Bachman and Palmer's (1996) model is by no means a perfect, definitive model. Based principally on language testing, its application to general language use may be more problematic. For instance, it might be relatively easy to document or evaluate aspects of pragmatic or sociolinguistic competence in language tests of a familiar design, but in wider, freer communicative settings, they may be harder to distinguish and therefore harder for the learner to exploit. Also, though Bachman and Palmer accentuate the interaction with the task or setting, gaps in teaching materials pertaining to spoken grammar (see Section 3.2) may actually interfere with a student's impressions of successful interactive speech. Finally, unlike in subsequent theories of communicative competence (see Celce-Murcia, 2008; Celce-Murcia & Dornyei, 1995), there is no explanation for how learners make use of formulaic language. One could question where such a factor should be placed in the model due to its range of uses and ultimately the bearings it has on learner success (see 3.3.2). Despite these drawbacks, however, discussions of success will base themselves around Bachman and Palmer's model. Though the CEFR has its own perspective on communicative competence, to be presented in Section 2.3.3, it is easy to see how can-do statements relate to the model presented in Figure 4. As will be shown in Chapter 5.2.1, the research tool used to collect successful language was, in fact, based on language exams so Bachman and Palmer's model is still of clear relevance to this study.

Communicative competence theory has widened the scope of success for this study. Though a look at the research questions presented in Section 1.6 reveals that grammatical competence is one way of examining B1, B2 and C1 learner success, it is clear that CEFR statements of ability will allude to textual, illocutionary and sociolinguistic competence. However, before the CEFR's view of competence is presented, it is necessary to look more closely at definitions of communication and the pragmatic and strategic skills learners need to learn so as to further understanding not

only of what learners need to develop, but also the challenges they face to being judged as successful language users in their speech.

2.2 Communication skills and the effect of pragmatic and strategic competences on success

After examining theories of communicative competence and language use, it is clear that competence in grammar, discourse, pragmatics, sociolinguistics and strategy usage are fundamental to learner success in learning a second language. Though Chapter 1 discussed success in terms of the models against which learners can be compared, Chapter 2 has so far concentrated on *what* they need to learn in order to be successful. Ending with Canale (1983) and Bachman and Palmer's (1996) theories of communicative competence and language use, introductions have been made as to how this knowledge can manifest itself in spoken performance. Put simply, this section will elaborate on the skills learners require if they are to become successful, *proficient* speakers of English. It is the writer's perspective that although proficiency appears in much linguistics literature, it is not always explicitly defined; the assumption possibly being that, due to its commonality, it no longer needs precise definition. However, since proficiency broadly denotes "*something like* 'the ability to make use of competence' [emphasis added]", it clearly shares similarities with the notion that performance is reliant upon the capacity to implement competence (Taylor, 1988: 166). Since proficiency can also be likened to the ability to "function in natural language situations" and to convey and comprehend messages (Spolsky et al., 1968: 8), this section of the literature review will centre on the definitions of and the abilities involved in spoken communication. After all, "the less one understands communication and communicative competence, the less certain is the understanding of...any purported measurement of communicative competence" (Canale, 1988: 68).

So far, the term 'communication' has been used rather generally without any explicit definition. However, a useful summary, allowing for its various characteristics and its fluid and variable nature to be distinguished, is provided by Canale (1983, citing the works of Breen & Candlin, 1980; Morrow, 1977; and Widdowson, 1978). He explains that communication:

- “is a form of social interaction, and is therefore normally acquired and used in social interaction;
- involves a high degree of unpredictability and creativity in form and message;
- takes place in discourse and sociocultural contexts which provide constraints on appropriate language use and also clues as to correct interpretations of utterances;
- is carried out under limiting psychological and other conditions such as memory constraints, fatigue and distraction;
- always has a purpose (for example, to establish social relations, to persuade, or to promise);
- involves authentic, as opposed to textbook-contrived language; and
- is judged as successful or not on the basis of actual outcomes. (For example, communication could be judged successful in the case of a non-native English speaker who was trying to find the train station in Toronto, uttered ‘How to go train’ to a passer-by, and was given directions to the train station.)”

(Canale, 1983: 3-4)

It is clear from this description that communication involves information-exchange in a rapidly evolving and unpredictable setting; it is as much adapted and extended by the listener(s) as it is by the speaker(s) (Canale, 1983). It is also mostly spoken, spontaneous, constructed in shared contexts, interactive, interpersonal, informal and expressive of social identities (Thornbury & Slade, 2006). Whereas language performance may be treated as a product of language, i.e the language created by a learner (Lewis, 2012), it is evident that communication is very much a process between speakers. However, as has been shown, for learners to make themselves fully understood, knowledge of the language alone will not suffice: “It is also useful if they are good communicators, that is, good at saying what they want to say in a way which the listener finds understandable” (Bygate, 1987: 22). Linking this to proficiency, as is argued by Field (2011: 82), success in spoken communication is thus not only reliant upon the features contained in speech, but upon the way they are “perceived” by the listener. Such a list of points, therefore, may be helpful in understanding the nature of communication, but it reveals little about how it is achieved or how spoken communication can be produced successfully. At this point, the importance of productive and interactive communication skills, of which Bygate (1987) provided a very detailed account, becomes pertinent. Though some writers highlight a distinction between production and communication skills (see Thornbury & Slade, 2006), here production will be treated as a sub-component of communication skills.

Production and interaction skills assist communication in different ways but they are both essential in managing spoken discourse between two or more parties under demanding processing conditions (see Levelt, 1989; 1999), most noticeably that of time constraints. Production skills, categorised into facilitation or compensation, are required during unplanned speech because of the limited time available for the articulation of ideas. Since it requires a great deal of spontaneity (De Bot, 1992; Hughes, 2011; Taylor, 2011), speech is often simplified using *facilitation* skills to reduce effort or gain more time by either simplifying structures, using ellipsis, utilising formulaic expressions or employing hesitations and fillers. At the same time, speech also “does not allow for reflection” (Field, 2011: 97), which means that *compensation* skills are necessitated to enable speakers to correct or reformulate their speech. Whilst written production, due to the time available, may allow language users to cross out and rewrite sentences after they have been created, speakers often have to simultaneously monitor and correct utterances as speech occurs. Spoken performance thus consists of self-correction or repairs, false-starts, repetition, unfinished utterances, ungrammaticality and rephrasing (Bygate, 1987; Thornbury & Slade, 2006).

Interaction skills, on the other hand, are significant in communication as they aid decision making in terms of the content of a message, its form, and its development with respect to a speaker’s intent or needs. Involved in preserving or sustaining relations with the other people engaged in the communication (Bygate, 1987), they comprise linguistic ‘*routines*’, the conventional patterns expected in particular types of discourse which govern the sequencing of informative and evaluative utterances, as well as the type of discourse the speakers engage in e.g. telephone calls, meetings, small talk (see also Weir, 2005a). *Negotiation skills*, finally, involve the management of interaction via topic selection, topic development, turn-taking, and the negotiation of meaning to ensure communication is clear and sufficiently explicit to all participants. With turn-taking and the maintenance of comprehension being essential features of communication, it is clear that such skills will be called upon frequently. Bygate’s (1987) summary of production and interaction skills are presented in Figures 5 and 6:

Figure 5: A break-down of production skills (Bygate, 1987)

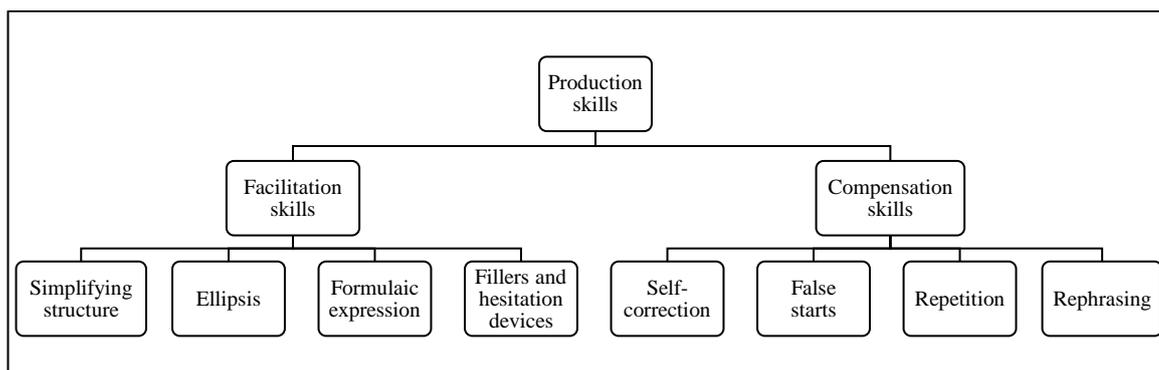
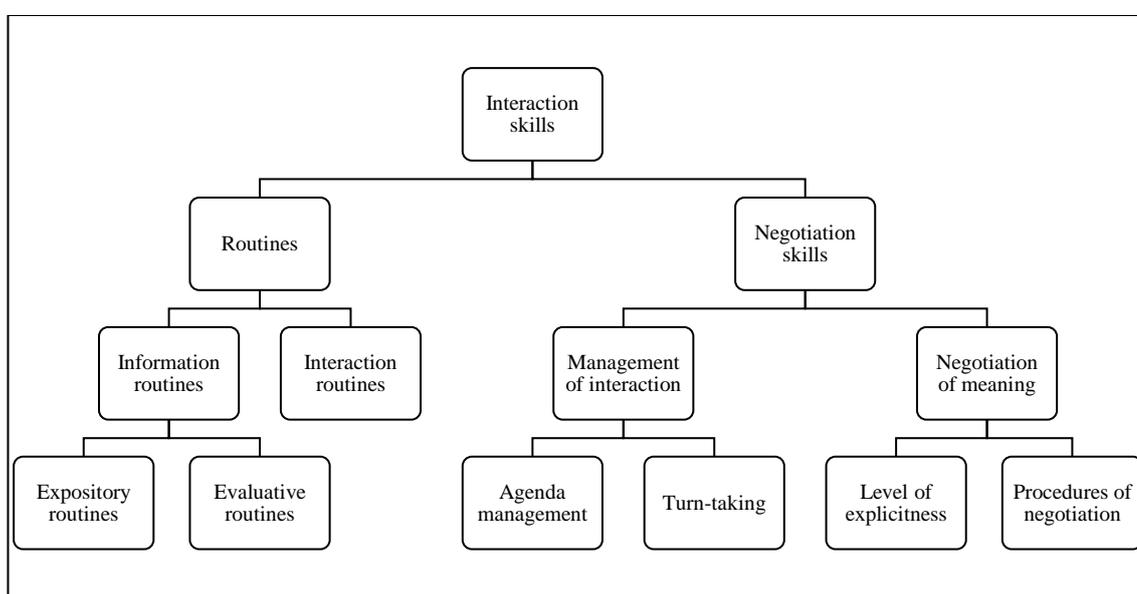


Figure 6: A break-down of interaction skills (Bygate, 1987)



In terms of learners' overall success, it is clear that communication skills are essential to their speech being seen as more, or less, proficient. Associated with pragmatic, strategic and discourse competence (Riggenbach, 1998), they help learners to manage the demands of unplanned, real-time spoken communication in everyday situations. As is noted by Field (2011: 86), communication skills themselves can act as an indicator of learner proficiency level due to the features receiving learners' attention:

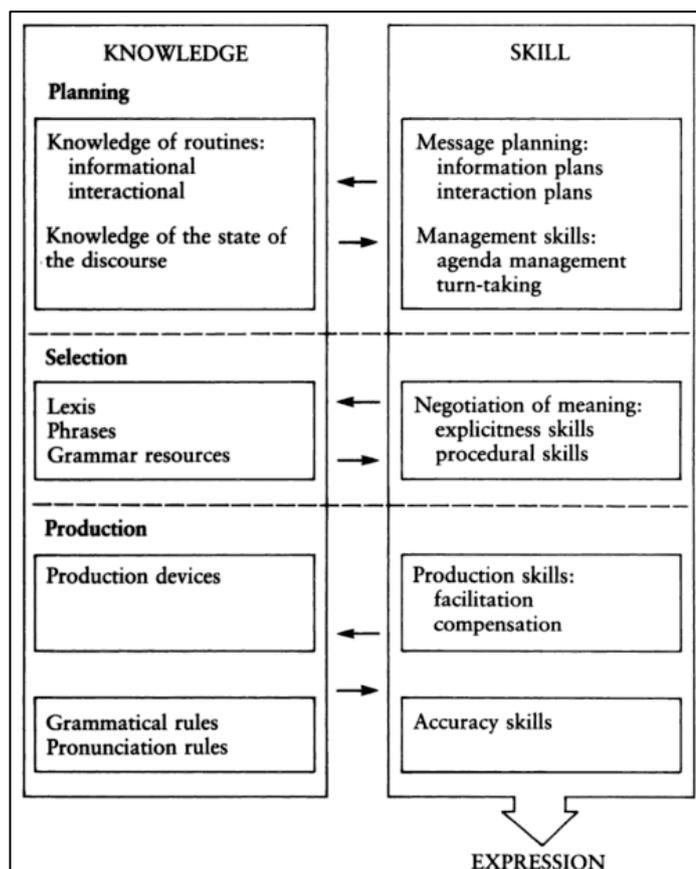
“...lower proficiency learners focus attention on linguistic features rather than pragmatic ones, comparing one or more of their syntax, lexis and pronunciation with what they perceive to be L2 norms. A mark of increased competence as an L2 speaker would thus be a gradual increment in the extent to which the speaker heeds the effectiveness with which the message has been conveyed.”

However, communication skills do pose challenges for learners with respect to i) L1 influences, and ii) the way such ‘skills’ may be perceived by others. Firstly, with regards to interaction routines, differences between the L1 and L2 speech patterns can be responsible for a lack of success during communication. For instance, turn-taking conventions in one culture could oblige speakers to speak only when the other speaker has completely finished their turn; alternatively, in other cultures, the use of overlaps and short turns may be considered the norm (Riggenbach, 1998). Rude or impolite discourse could therefore be counterbalanced by perceptions of unwillingness to talk and could influence the use of “micro skills” involved in claiming, maintaining and yielding turns (see Riggenbach, 1998: 57). Secondly, depending on the factors success is based on, communication skills may be misinterpreted as signs of diminished competence. Though productive skills are included by Bygate (1987) as essential to spoken communication, they are often associated with NSs due to assumptions that their linguistic knowledge is ‘complete’; conversely, interactive communication skills are associated with learners (Thornbury & Slade, 2006). It is debatable, therefore, whether some of the skills identified in Figures 5 and 6 would be considered a mark of higher learner proficiency as they are with NSs. A useful case in point refers to facilitation skills, specifically the use of hesitation and fillers. An extract from Viney and Viney (1996: 79, cited by Hughes, 2011; 77) explains that “hesitation strategies” such as fillers are disadvantageous because “If you use them too often you sound stupid”. Such a generalisation might dissuade students from using them at all despite them being unavoidable features of natural, spontaneous spoken language. Similarly, if someone were to hold a narrow view of proficiency, for instance focusing solely on use of grammar, rather than the discourse as a whole, the simplification of structures, ellipsis, self-correction and false starts may detract from learner impressions of success.

Relating this section to the rest of the literature review, understanding the processes occurring in communication and the skills that speakers employ is vital if speech is to be examined correctly. With respect to communicative competence, this section has shown that communication skills perform a range of duties for speakers engaging in spontaneous spoken conversation. Though Canale and Swain (1980) suggested that strategic competence related simply to the ability to employ language strategies to avoid or repair breakdowns in communication, the information presented here has demonstrated that they are responsible for much more than that. Rather than

simply allowing learners to resolve communication breakdowns, strategic competence allows speakers to *manage* their speech according to the norms of communication helping them deal with the demands of real-time speech. For success, therefore, strategies clearly help students exploit their knowledge and avoid communication gaps. They also help provide an important interface, as is demonstrated by Figure 7 below, allowing communicative competence to be utilised in performance.

Figure 7: Bygate's (1987: 50) summary of oral skills



2.3 Linking communication, communicative competence and learner success to the CEFR

Discussion in Sections 2.1 and 2.2 explored definitions of communicative competence, language use and communication in order to present previous literature relating to learner success in speech. However, as this thesis is examining success 'in accordance' with the CEFR, it is important that the Council of Europe's position is extended. The final section of this chapter therefore introduces the CEFR, along with its objectives and composition, so that its examination of communicative competence and

spoken proficiency can be fully related to this study's aims. In a sense, it serves a dual purpose. Firstly, after providing a general overview of the CEFR, it intends to relate previous definitions of communicative competence and communication to the CEFR's perspective; it will investigate the types of knowledge learners are said to possess and how this is realised in their speech. Secondly, it will expand on previous criticisms of the gaps that remain in describing learner proficiency so that their bearing on examining learner success can be fully appreciated. In doing so, it will indicate how this study answers calls from the CoE (Cambridge ESOL, 2011, 2011: 17) for researchers to "elaborate the contents of the CEFR...establishing which vocabulary and structures occur at a particular proficiency level in a given language" so as to make its application to a given context more meaningful.

2.3.1 A brief introduction to the CEFR and its relevance to the study of success

Currently translated into 39 languages (CoE, 2014) and in preparation between 1993 and 2000 (Goullier, 2006), the CEFR is a document outlining how language proficiency and abilities progress across a vast range of language learning contexts:

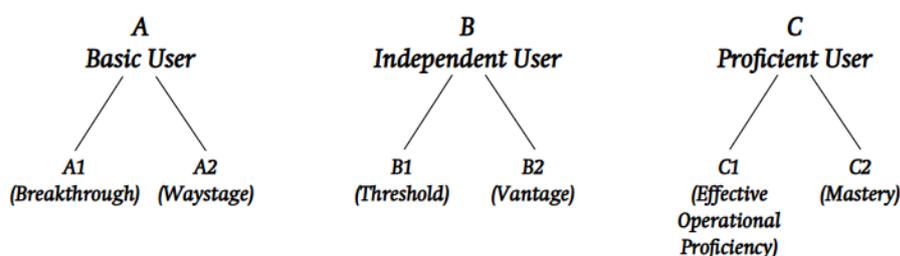
"It describes in a comprehensive way what language learners have to learn to do in order to use a language for communication and what knowledge and skills they have to develop so as to be able to act effectively."

(CoE, 2001: 1)

Documenting the language *activities* (receptive, productive, interactive, mediative), the language *domains* (public, general, educational, occupational) and the communicative *competences* (linguistic, sociolinguistic, pragmatic) of learners, the CEFR categorises language learning into six Common Reference Levels ranging from A1 Breakthrough to C2 Mastery (see Fig. 8). These levels are then each illustrated via a set of general and specific descriptors detailing the abilities of learners as they progress globally through the levels, or as they develop within a particular language use context. With regards to its usage, the CEFR acts as an aid to those involved in the learning, teaching, assessment and policy of language, and represents the culmination of over six decades of work by the CoE (2001), an organisation responsible for endorsing plurilingualism, linguistic diversity, mutual understanding, democratic citizenship and social cohesion across its 47 member states (CoE, 2015; Language Policy Division [LPD], 2006). In doing so, the

CEFR functions as a tool for improving unity between these members (CoE Committee of Ministers, 1982) and for establishing a common terminology to increase transparency, communication and reflection (CoE, 2001: xi). Likewise, in detailing the numerous competences learners possess and develop, it has also been found to be “extremely influential” in syllabus design, curricula planning and language examinations (Hulstijn, 2007: 663) as well as the planning of language learning programmes, language certification and self-directed learning (CoE, 2001). Intended to be non-language specific to widen its application and scope (CoE, 2001), the CEFR’s six levels illustrated with ability descriptors were attributed to its swift adoption (Alderson, 2007) and to its use as the “exclusive neutral reference” in the national, educational setting (Martyniuk & Noijons, 2007: 7).

Figure 8: CEFR Common Reference Levels (CoE, 2001: 23)



Before outlining the CEFR’s model of communicative competence, it is necessary to emphasise a fundamental distinction, of immediate relevance to the introductory chapter, that the document makes. This distinction is that the CEFR no longer equates native-like proficiency in a TL to the overall goal of second language learning. Although learners may still aspire to this level, the CEFR is novel in highlighting that the needs of individual learners may, in fact, be met at *different* competence levels. It states that “successful communication” can be achieved according to *varying* degrees of proficiency in multiple languages, and according to personal, ever-changing requirements (LPD, 2006: 5). Put simply, a learner may not demonstrate native-like ability not because they are deficient, but because such a level is not required for them to operate in the TL effectively and successfully. Also, second language learners are no longer ‘pigeon-holed’ into the broad non-native category highlighted in Chapter 1. Instead, by placing the learner at the heart of language policy in the evolving European climate (LPD, 2006), a TL no longer ‘belongs’ solely to its native speakers but to its users who operate in it at different levels for different purposes. As “social agent[s]” participating in intercultural

communications where languages are not viewed or learnt “in isolation” (University of Cambridge ESOL Examinations [CamESOL], 2011: 14). The CEFR ultimately sees learners as individuals interacting in settings specific to *their* daily lives (Jones & Saville, 2009). It additionally recognises the inimitable blend of competences and abilities which makes each multicompetent learner a “unique combination” (Saville-Troike, 2012: 8). In the CEFR, therefore, not only are learners viewed on an individual rather than a collective basis, but the pressure to emulate the NS is also lessened.

2.3.2 Language use, general competences and communicative competences

As Section 2.3.1 remarked, the CEFR views language users and learners as social agents: individuals achieving ‘tasks’ requiring the use of their strategic and general competences within specific contexts from which acts of speech acquire their “full meaning” (CoE, 2001: 9). In its action oriented view, language use in the CEFR:

“...comprises the actions performed by persons who as individuals and as social agents develop a range of competences, both general and in particular communicative language competences. They draw on the competences at their disposal in various contexts under various conditions and under various constraints to engage in language activities involving language processes to produce and/or receive texts in relation to themes in specific domains, activating those strategies which seem most appropriate for carrying out the tasks to be accomplished. The monitoring of these actions by the participants leads to the reinforcement or modification of their competences.”

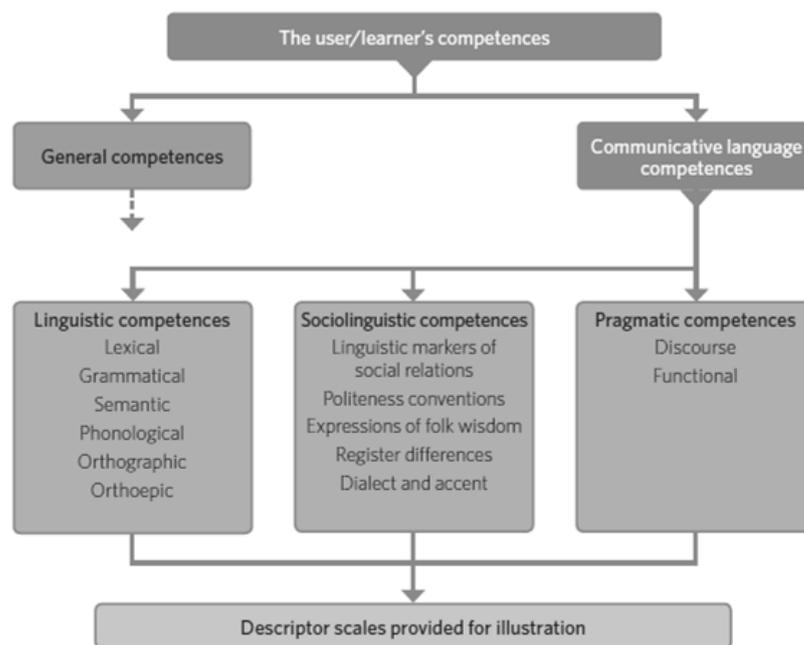
(CoE, 2001: 9)

For the CEFR, language use is not static, nor is it uniform. Language users, such as those in Canale and Swain’s (1980) theory, benefit from their unique set of competences and strategies which enable them not only to *achieve* linguistic and non-linguistic tasks, but also to *develop* in those competences. With competences subject to being at a user’s ‘disposal’, the above statement also demonstrates that while knowledge, skills and characteristics may be held in a learner’s competence (as defined by the CoE, 2001), they may nevertheless be unavailable.

The CEFR, similarly to Canale and Swain (1980), Canale (1983) and Bachman and Palmer (1996), views communicative competence as part of a language user’s wider, more general competence (see Figure 9). Comprising *knowledge*, *skills*, *existential competence* and the *ability to learn*, an almost cyclic relationship is created:

communicative competence is a part of general competence, but it is this general competence which rather vaguely “contribute[s] in one way or another” to learners’ communicative abilities (CoE, 2001: 101). *Declarative knowledge*, concerning world, sociocultural and intercultural knowledge, is a product of experience and formal learning. It relates to one’s language and culture as well as to language users’ knowledge of day-to-day life. In a sense, declarative knowledge is the CEFR’s equivalent to Bachman and Palmer’s (1996) treatment of topic knowledge and knowledge of different settings. *Skills*, pertaining to procedural knowledge or ‘know-how’, on the other hand concern the “ability to carry out procedures” (CoE, 2001: 11). Using similar terminology to Canale (1983), they relate to social, living, vocational and leisure skills which differ in their degree of mastery, and the ease, speed and confidence with which they can be performed. *Existential competence*, comparative to Bachman and Palmer’s (1996) personal characteristics, is composed of personal attributes unique to language users. Containing attitudes, motivations, values, beliefs, cognitive styles and personal factors, existential competence results from acculturation and a person’s readiness to interact socially with others. Finally, drawing on “various types of competence”, the *ability to learn* amalgamates the three previous aspects (CoE, 2001: 12). Helping language users to “deal with the unknown” (CoE, 2001: 12), it involves more than the capability to learn; it suggests that users will differ in their predisposition, or willingness, to seize or seek opportunities for exploiting learning potential, hopefully reducing the resistance or threat felt in learning a new language.

Figure 9: The CEFR's (Cambridge ESOL, 2011: 10) model of general and communicative competences



The CEFR's notion of general competence thus relies on more than knowledge that something *can* be done: it highlights the need to know *how* it can be done, the *willingness* to do it, and the *ability* to add to what is already known. As will be shown shortly, the CEFR is perhaps a little contradictory in this respect: the gap partly being addressed in this study of success relates to the CEFR's inability to explain how statements of what can be done at different proficiency levels can be fulfilled. Though it was never the CEFR's intention to be exhaustive, so as to encourage user reflection of their specific needs (Coe, 2001), this preliminary discussion of general competence illustrates some of the misgivings of previous authors. To keep the document non-prescriptive and flexible, for instance, it contains many boxes raising questions as to what CEFR users 'may wish to consider'. With one of these boxes for each of the four general competence components, users may be left with more questions than answers. This also reinforces criticisms levelled at the CEFR's vague definitions and terms (Alderson, 2007; North, 2007; Weir, 2005b). For discussions of success, however, such points, see Figure 10, further complicate understanding of a learner's general competence: if a language user's 'starting point' is difficult to pinpoint, the ability to monitor learner progress surely diminishes also.

Figure 10: An example 'users may wish to consider' box (CoE, 2001: 102)

Users of the Framework may wish to consider and where appropriate state:

- *what knowledge of the world the language learner will be assumed/required to possess;*
- *what new knowledge of the world, particularly in respect of the country in which the language is spoken the learner will need/be equipped to acquire in the course of language learning.*

Moving from general to communicative competence, it is clear that comparisons can be made with Canale and Swain (1980) and the CEFR's model as it contains three similar components: *linguistic*, *sociolinguistic* and *pragmatic* competences. Interesting to note here is the plurality of the term 'competence' as, in the CEFR's sense once again, competence refers not only to knowledge but to combinations of knowledge, skills and know-how in each of the three competences (CoE, 2001). *Linguistic* competences relate closely to preceding definitions. Comprising knowledge of and the ability to use lexical, formulaic, grammatical, morphological, syntactical, semantic, phonological, orthographical and orthoepic competence, lexical competences relates to the construction and formulation of well-formed, meaningful messages (CoE, 2001). Sociolinguistic competences, on the other hand, closely assimilate to previous models' social and cultural conventions, the particular linguistic customs which may need to be followed in different settings. *Pragmatic* competence, in place of Canale's (1983) discourse competence, concerns "the functional use of linguistic resources...the mastery of discourse, cohesion and coherence" (CoE, 2001; 13). Combining the arrangement of sentences into sequences (discourse competence) as well as the knowledge of and ability to construct interaction, it conforms closely to Bygate's (1987) discussion of 'routine' and 'interactive skills' to produce spoken conversation.

This model therefore raises several points for learner success. For instance, in stating that linguistic competence can be scaled, the CEFR argues that language is "never completely mastered by *any* of its users" [emphasis added] (CoE, 2001: 109). Perhaps in opposition to previous NS-learner comparisons, all language users are said to have something new to learn; individuals' skills and accessibility to the range and quality of knowledge also varies greatly. Similarly, as was highlighted in Chapter 1, no truly homogenous settings exist. To adapt language use accordingly in a variety of social contexts, learners thus need to develop their intercultural competence but this is not always a straightforward task. Despite having knowledge, skills and know-how, language

users may be uninformed of sociolinguistic norms in a target language setting: just as Paulston (1992) remarks, language users can be unaware of cultural rules until the point at which they are broken.

It is the CEFR's model of general and communicative competences which provide the platform for this study's examination of success. As previously identified, general competences are described but not illustrated. Similarly, communicative competences, the internalised yet changeable state of knowledge, skills and know-how are not fully exemplified. Though recognised and activated during performance, "observable behaviour" (CoE, 2001: 14), no specification is offered as to the interplay between knowledge, skills and know-how, nor whether performance fully reflects a language learner's competence. Though specific CEFR criticisms relating to this study's focus will be raised in the next section, it is easy to see how the CEFR is vulnerable to other criticisms relating to its length and its use of complex terminology (Alderson, 2007; Jones & Saville, 2009; Martyniuk & Noijons, 2007; Cambridge ESOL, 2011). Introduced in the opening pages, communicative competence is only fully explored after proficiency descriptors and scales for listening, reading, speaking and writing are presented. The fact that fuller discussion is 'tucked away' in the document's fifth chapter and presented in rather difficult vocabulary (Jones & Saville, 2009) could explain why many users have not reflected thoroughly upon the issues raised in the CEFR and why its levels and descriptors have been adopted so readily. Similarly, the use of specific, difficult terminology could explain why its accessibility to all users has been questioned. Finally, and most importantly for communicative competence, its aim of being comprehensive, non-prescriptive and encompassing of all languages and learning contexts results in a document which does not exemplify how such competences and sub-competences may be realised. In trying to be everything to all people, it may end up being nothing to anyone.

2.3.3 Linking the CEFR to success

The CEFR has faced a range of criticisms since its introduction: its misuse and inappropriate application in political and pedagogical contexts being a case in point (Alderson, 2007; Bonnet, 2007; Coste, 2007; Davidson & Fulcher, 2007; Figueras, 2007; Figueras, 2012; Fulcher, 2004). However, in relation to learner success, three of its most significant shortcomings include i) a lack of theoretical support from SLA theory, ii) the existence of vague definitions, vocabulary and descriptor scales, and iii) the lack of actual

language use in illustrating learner language (see Alderson, 2007; Davidson & Lynch, 2002; Fulcher, 2004; Hulstijn, 2007; Weir, 2005b). Establishing a ‘common’ basis for language learning, provision and policy has seemingly come at a cost for those seeking more explicit, tangible descriptions of learner proficiency. These criticisms will now be discussed in more detail and related to the notion of success for this study.

Firstly, though the CEFR has previously been praised for its empirical basis (North, 2007), its positive wording (Little, 2007), and its mapping of functional competence (Weir, 2005b), other writers have insisted that it is essentially “atheoretical” (Fulcher, 2004: 258). It was constructed using a combination of intuitive, qualitative and quantitative methods aiming to describe and scale the abilities demonstrated at different levels (CoE, 2001). The basis of such scaling, however, relied on teachers’ *perceptions* of proficiency and not on any L2 evidence (Figueras et al., 2005; Hulstijn, 2007). It also lacked any SLA validation due to the absence of an available and definitive theory of second language learning and competence (CoE, 2001; Figueras, 2012; Hulstijn, 2007). In a sense, the CEFR was empirical as it rested on observations made by teachers but it was not initially empirical according to actual research. Little (2007: 648) argues that this is perhaps to be expected “given the scope of what the CEFR attempts” but size and ‘scope’ can no longer act as a reasonable defence: the widespread application of the CEFR means that calls for a stronger empirical base are more relevant than ever. The examination of success in this study is thus reliant upon the accuracy of proficiency descriptions within the CEFR and the perceptions upon which they are based.

Secondly, the CoE (2201: 7-8) stipulates that the framework should be “open and flexible”; it must correspond to criteria requiring it to be multipurpose, flexible, open, dynamic, user-friendly, and non-dogmatic. However, in its aim to be adaptable, non-language specific and comprehensive (CoE, 2001), the CEFR’s wording and structure has generated frustration. In particular, criticisms have been directed towards its descriptors and can-do scales. Some warn of the apparent preference and priority given to scales rather than its descriptors (Alderson, 2007) which has resulted in some viewing the CEFR as *the* system and sequence of language learning (Fulcher, 2004) despite its non-hierarchical order (CoE, 2001; Little, 2007). In a similar vein, though the CEFR aims for “definitiveness” (CoE, 2001: 206), the wording of its descriptors and scales is said to be incoherent, inconsistent, ambiguous, non-distinct and open to interpretation (Alderson, 2007; Figueras, 2012; Jones, 2002; Weir, 2005b). Definitiveness means that descriptors should not be vague but what remains is a list of abstract statements which then have to

be applied or related to observable behaviour (Figueras, et al., 2005). Though the CEFR acknowledges previous scales' errors in their use of qualifiers (see CoE, 2001: 206), the language used in CEFR scales is still seen as ambiguous (see language highlighted in **bold** in Fig. 11). Beyond the vocabulary realising descriptors and scales, there exist more obstacles for CEFR users. For instance, it is unclear whether learners reach a particular level before or after they are able to satisfy the can-do statements, whether they can exhibit different proficiency levels on different scales, whether the relationship between quantity and quality of can-do achievement is relevant, and whether a proficiency level subsumes all or some of the abilities in lower levels (Hulstijn, 2007). Higher levels of proficiency, C1 and C2 are likewise sometimes neglected (Jones, 2002) (see Fig. 12). For success, this secondary group of shortcomings means that the expectations of those involved in language teaching or learning may influence or skew judgements of learner accomplishments in an L2. Not only does it become increasingly difficult to distinguish what makes B1, B2 and C1 learners different in their language use, but it becomes difficult to be certain that learners do fully represent the levels described in the CEFR.

Figure 11: CEFR Qualitative aspects (CoE, 2001: 29) with vague words in bold

	Range
C2	Shows great flexibility reformulating ideas in differing linguistic forms to convey finer shades of meaning precisely, to give emphasis, to differentiate and to eliminate ambiguity. Also has a good command of idiomatic expressions and colloquialisms.
C1	Has a good command of a broad range of language allowing him/her to select a formulation to express him/herself clearly in an appropriate style on a wide range of general, academic, professional or leisure topics without having to restrict what he/she wants to say.
B2+	
B2	Has sufficient range of language to be able to give clear descriptions, express viewpoints on most general topics, without much conspicuous searching for words, using some complex sentence forms to do so.
B1+	
B1	Has enough language to get by with sufficient vocabulary to express him/herself with some hesitation and circumlocutions on topics such as family, hobbies and interests, work, travel, and current events.

Figure 12: CEFR example of can-do scale. Spoken interaction (CoE, 2001: 79)

	GOAL-ORIENTED CO-OPERATION (e.g. Repairing a car, discussing a document, organising an event)
C2	<i>As B2</i>
C1	<i>As B2</i>
B2	<i>Can understand detailed instructions reliably. Can help along the progress of the work by inviting others to join in, say what they think, etc. Can outline an issue or a problem clearly, speculating about causes or consequences, and weighing advantages and disadvantages of different approaches.</i>
B1	<i>Can follow what is said, though he/she may occasionally have to ask for repetition or clarification if the other people's talk is rapid or extended. Can explain why something is a problem, discuss what to do next, compare and contrast alternatives. Can give brief comments on the views of others.</i>
	<i>Can generally follow what is said and, when necessary, can repeat back part of what someone has said to confirm mutual understanding. Can make his/her opinions and reactions understood as regards possible solutions or the question of what to do next, giving brief reasons and explanations. Can invite others to give their views on how to proceed.</i>
A2	<i>Can understand enough to manage simple, routine tasks without undue effort, asking very simply for repetition when he/she does not understand. Can discuss what to do next, making and responding to suggestions, asking for and giving directions.</i>
	<i>Can indicate when he/she is following and can be made to understand what is necessary, if the speaker takes the trouble. Can communicate in simple and routine tasks using simple phrases to ask for and provide things, to get simple information and to discuss what to do next.</i>
A1	<i>Can understand questions and instructions addressed carefully and slowly to him/her and follow short, simple directions. Can ask people for things, and give people things.</i>

The last criticism to be examined concerns the lack of actual language use to illustrate descriptors and scales. Of course, the CEFR never intended to be language specific so as to widen its application and use. With its descriptors and scales acting as a “concertina-like tool” (LPD, 2009:3) to map learner progress, the CEFR is said to inform users from any language background of the differing uses of language that a learner can deal with and the effectiveness of their language use (dubbed ‘quantity’ and ‘quality’ by de Jong, 2004). However, as Martyniuk and Noijons (2007: 6) report, users have complained about the document’s “lack of detail”. Using Figure 11 as an example, no indication of the breadth or depth of lexical knowledge is supplied to make the different descriptors more explicit (Weir, 2005). Ultimately, whilst it describes *what* abilities should be demonstrated across proficiency levels, it fails to exemplify *how* they can be achieved:

“[the CEFR] does not contain any guidance, even at a general level, of what might be simple in terms of structures, lexis or any other linguistic level. Therefore the CEFR would need to be supplemented with lists of grammatical structures and lexical items for each language to be tested, which could be referred to if terms like ‘simple’ are to have any meaning for item writers or item bank compilers.”

(Alderson et al., 2004:13)

Whilst other writers have also supported this notion (Figueras et al., 2005; Weir, 2005; Westhoff, 2007), for examinations of learner success it poses a fundamental challenge as a “barrier” is placed between learners’ accomplishments and their written descriptors (Kedde, 2004 cited by Weir, 2005; 43-44). This shortcoming also means that those using the CEFR, either as learners, teachers, assessors or curriculum planners, are left wondering what language should be evidenced at the six Common Reference Levels and what language would satisfy the criteria presented in its many can-do statements and scales. Likewise, despite the qualitative aspects given in the CEFR (see CoE, 2001: 37-38) regarding range, accuracy, fluency, interaction and coherence, users of the document are given no examples to demonstrate the differences and distinctions across levels. The authors were keen to adopt positive phrasing in the CEFR to emphasise what learners do know and can do rather than their inadequacies or deficiencies (Alderson, 2007; Figueras, 2012; Little, 2007; North, 2007) so as to allow its scales and descriptors to “serve as objectives” (CoE, 2001: 205). What remains with regards to success, however, is a document which seemingly stresses language development as a process of accomplishing tasks and satisfying criteria as opposed to the language employed to achieve it. For its contents to fully function as ‘objectives’, further detail is needed since “a learning outcome needs to state what the learner will have learnt and will be able to do at the end of a course of study” (Figueras, 2012: 481). Thus, with a lack of learner language, it remains unclear what lexical and grammatical structures learners should know at different levels. Only by filling this gap will CEFR users know how language use might develop from A1 to C2, and therefore, what learners at each of the levels can truly do.

2.4 Chapter conclusion

This preliminary literature review chapter has aimed to answer Spolsky et al.’s (1968: 79) question: “What does it mean to know a language?” Since *knowing* a language has been shown to be very different from *using* a language, it was necessary to revisit

communicative competence theories in this preliminary literature chapter to reveal not only the types of knowledge language users possess, but also to hypothesise how performance may or may not provide abstractions of it in communicative settings. With this study aiming to establish what makes B1, B2 and C1 students successful in the UCLanESB speaking exams, a clear understanding of communicative, productive and interactive skills, and the process of spoken production had to be exhibited. It has been shown that knowledge of linguistic, sociolinguistic, discourse and strategic competence is required if language learners are to become proficient speakers. Employing communicative skills and becoming more accurate, appropriate and fluent in their spoken expression are therefore dependent upon these competences but learning how to utilise them in English is a different matter. Similarly, the gaps within the CEFR pose significant challenges to users wishing to fully understand lexical and grammatical differences distinguishing learners at different levels. There remains a need to extend descriptors and scales with examples of real learner speech, if discussions of L2 success are to have a stronger foundation.

The next literature review chapter will build on discussion by focussing not on *what* learner language comprises, but rather *how* it is used in real communication. Focussing principally on findings from corpus linguistics research, specifically from spoken corpora, it aims to detail the true features of spoken English, and with regards to the research questions outlined in Section 1.6, the relevance of corpora in extending knowledge of learners' lexicogrammar and strategy usage in spoken English.

3. LITERATURE REVIEW (PART TWO): FINDINGS FROM CORPORA AND IMPLICATIONS FOR LEARNER SUCCESS IN SPEECH

3.0 Introduction

The previous chapter examined what it means to know and be able to use a second language. It highlighted the components of communicative competence and their interplay during language production. It also shed light on the weaknesses of current descriptions of learner language in the CEFR. This chapter instead reviews what has been learnt from inspections of real L2 performance, and specifically for this study, what previous research into learner speech has revealed. By introducing and analysing the findings from previous corpus linguistics research, it aims to pinpoint how evidence from real learner speech has extended understanding of its nature and the gaps that remain for discussions of success.

To achieve this aim, discussion first needs to be presented as to what corpus linguistics is. Although methodological information will be offered in Chapter 5, it would be difficult to understand the contributions it has made without introducing its aims, strengths, drawbacks and significance to this study. This chapter then proceeds to highlight the findings it has uncovered for spoken grammar and lexical competence. Finally, so as to draw together the two literature review chapters, a brief rationale will be offered as to how the study's RQs relate to the information presented.

3.1 An introduction to corpus linguistics and its significance for speech

3.1.1 What is corpus linguistics and what do corpora contain?

The terms corpus linguistics, *computer corpus linguistics* (Leech, 1992), and electronic text analysis (Adolphs, 2006) refer to the capture, storage and analysis of texts on computer for the study of language. A corpus, meaning body in Latin (Baker, 2010; McEnery & Wilson, 2001), comprises a principled collection of “naturally occurring” language (Hunston, 2002: 2). Representing pools of evidence ranging in size, such collections can be accessed by linguists for numerous analyses, giving rise to new insights into language use which subsequently inform teaching, material design and additional linguistic resources (Biber, Conrad & Reppen, 1998; Kennedy, 1998; McEnery and

Hardie, 2012; O’Keeffe, McCarthy & Carter, 2007; Sinclair, 2004; Stubbs, 2004). Though not embodying a *branch* of linguistics, but rather a *methodology* (Granger, 2002; Leech, 1992; McEnery & Wilson, 2001; Meyer, 2002; Stubbs, 2004), corpus linguistics is seen as a growing, fruitful field of research (O’Keeffe & McCarthy, 2010) which is able to increase the speed of analysis, reduce human error and uncover findings inaccessible to intuition alone (Adolphs, 2006; Baker, 2010; Biber et al, 1998; Hunston, 2002; Kennedy, 1998; McEnery & Hardie, 2012). In essence, it ultimately aims to answer one fundamental question: “how do people *really* use language?” (Baker, 2010: 102).

Corpora, as already indicated, contain ‘naturally occurring’ written, spoken or multimodal language. However, although the word ‘natural’ appears in much corpus linguistics literature, it is a word that is often not fully explained. Baker (2010: 94) and McEnery and Wilson (2001: 1) do attempt to clarify its meaning by remarking that corpus linguistics depends on “real-world” or “real life” instances of language. For language to be natural or real, it must therefore fulfil a practical role in people’s lives (see also Carter, 1998). Other writers have also asserted that corpora should contain ‘authentic’ (Sinclair, 1996 cited in Granger, 2002) or “used” language (Brazil, 1995: 24) in that it is employed not simply for the purpose of showing how a language system ‘works’. For instance, though a television script may be real in the sense that it presumably contains real words and real grammar, it would be unauthentic and unable to represent spontaneous spoken communication reliably. Whether written or spoken, a corpus, if aiming to be truly demonstrative of the language being explored, should therefore not base itself on artificial language but instead on genuine, authentic examples of *real* language use (Granger, 2002). For the purposes of this thesis, the term real will be chosen. As will be seen later in Section 5.2.1, the spoken language at the heart of this study may not be entirely natural, but it is real in that it was produced by real learners in a genuine spoken encounter.

An additional tenet to corpus content relates to the selection of texts for inclusion. Viewing a corpus as a straightforward ‘collection’ of texts is somewhat short-sighted and impractical. It can mislead people into thinking that corpus linguistics is a “mindless process of automatic description” when, in fact, studies seek to answer specific research questions (Kennedy, 1998: 2). For that reason, the selection of texts should be ‘principled’ (Sinclair, 2004) and based on sound, preselected criteria that will help the corpus to be *representative* of the chosen language (Biber et al, 1998) (see Section 5.2.4). A haphazard gathering of spoken or written texts may partially satisfy corpus definitions in that real language will be captured, but if not carefully chosen, subsequent findings may bear little

relevance to research questions or the language in question (Biber et al, 1998; Sinclair, 2004). To illustrate this point, a researcher aiming to investigate current-day newspaper reporting will fail to fully document the desired genre adequately if they include one newspaper brand or if either tabloids or broadsheets are omitted. Hence, representativeness is essential as it can provide “as accurate a picture as possible” of the language being observed (McEnery & Wilson, 2001: 30) but it must also be borne in mind that regardless of size, corpora can never be identical to the language under examination. Language is not finite; a finite series of grammatical rules may exist but the potential number of sentences yielded is infinite (McEnery & Wilson, 2001). A corpus can thus only provide evidence for the language contained within it (Hunston, 2002; McEnery & Hardie, 2012; Sinclair, 2004). The need for text selection to be principled and representative therefore, stems from the impossible task of including all examples or texts of a chosen language variety (except in the cases of fixed language variety accounts (Hunston, 2002) or finite text collections (Adolphs, 2006)). Corpora can facilitate research into real uses of language but only if they are carefully constructed.

3.1.2 The benefits and drawbacks of corpus linguistics

Corpus linguistics provides several advantages and disadvantages in research. They will first be presented here in a general context but then they will then be applied to this specific study more directly in Section 3.1.3.

First, corpora are relevant to the competence-performance debate. Generative grammarians, or rationalists, focus on the way that language is processed in the mind (Adolphs, 2006). Using artificial data and introspection, a native speaker exploits their own reflections of language in a bid to extrapolate the factors influencing internal and external cognition (McEnery & Wilson, 2001): it is an individual’s *competence*, the internalised knowledge of innate structures through which grammar is acquired that should be the sole focus of linguistic enquiry. Corpora, in the purely rationalist view, are deemed inadequate in this pursuit. Belonging to the empiricist school of linguistic enquiry, which concentrates on theory informed by language use observations (Adolphs, 2006; Leech, 1992), corpora instead reflect an individual’s *performance* in a language. “[T]he behavioural manifestation of language” (Leech, 1992: 107) is thus believed to not fully reflect a person’s competence in language. However, the sheer frequencies involved in corpora (e.g. the 100 million word British National Corpus (2015), the 4.5 billion word

Collins Corpus (2013), and the multi-billion word Cambridge English Corpus (2015)) suggest that items contained within them do not simply occur by chance alone (Stubbs, 2004). They instead show the “regular, patterned preferences for modes of expression” speakers call upon repeatedly in particular language contexts (McCarthy, 2006: 9). It is this recurrent patterning that is deemed to be partly representative of a speaker’s competence. This frequent patterning also reveals evidence which is pertinent to previous communicative competence theories as corpus research can reveal much about what is and is not ‘done’ in a language (see Section 2.1.1).

Another advantage is that corpora resolve some of the issues created by the sole use of introspection or intuition. Since corpus linguistics is often associated with quantitative analysis (Granger, 2002; Timmis, 2015), it can counter the potentially unreliable use of introspection (Baker, 2010). It specifically combats the unique, inaccessible and unreplicable source that is a researcher’s own mind (Adolphs, 2006) and it can help resolve issues surrounding bias in personal testimony which influence reliability and verification (Adolphs, 2006; Sinclair, 2004, Stubbs, 2004). Of course, this depends on the way corpora are used. *Corpus-based* approaches use corpora to “expound, test or exemplify theories and descriptions” originating from introspection (Tognini-Bonelli, 2001: 65); the researcher’s commitment to the corpus and the truth paid to its data can thus be questionable (McEnery, Xiao & Tono, 2006). *Corpus-driven* approaches, on the other hand, uphold the “integrity of the data” by ensuring descriptions are fully centred on the evidence provided in the corpus (Tognini-Bonelli, 2001: 84). Rather than using introspectively created rules as a basis for research, findings are instead said to emerge directly from the data. By adopting the latter approach, the absence of predetermined lines of enquiry can thus result in greater findings which are ‘truer’ to the data. The fact that corpus-driven studies are “descriptive” rather than “prescriptive” is therefore an obvious advantage as restrictions are not placed on the findings that could materialise (Jones & Waller, 2015: 9).

Finally, it must be highlighted that researcher intuition *does* still have a role to play in corpus linguistics. Without intuition, it is “doubtful” whether there would be any research questions to explore using corpora in the first place (McEnery & Hardie, 2012: 26). As corpora are built to be representative of language that best suits the research aims, their design and use would seemingly be made obsolete without an individual’s ‘feel’ for language and knowledge of the research gaps which remain unanswered; also, corpora are unable to reveal what is possible, correct or absent in a language (Baker, 2010;

Hunston, 2002) and on their own, they do very little other than provide evidence. Ultimately, someone needs to make sense of the evidence provided so corpora and intuition are not mutually exclusive (Baker, 2010; McEnery & Wilson, 2001). Instead, corpora should be seen as an aid to focus intuition (Stubbs, 2004); they not only provide evidence which may differ from introspective insights, but they also help researchers to make choices about which language to focus on. If combined, they can also balance each other's limitations in respect to intuition's tendency to focus on salient, rather than frequent aspects, and corpora's inattention to infrequent, rare or absent features (Baker, 2010; Hunston, 2002). Another area which benefits from corpora is patterning of "collocation, frequency, prosody and phraseology", an aspect difficult to pinpoint using intuition alone (Hunston, 2002: 20). Patterns are significant as they can present evidence that would challenge the purist competence-performance dichotomy. With intuition not considered to be fully accurate in this respect, it is clear that corpora supply a unique source of evidence that can uncover findings that would otherwise remain unknown, tentative or biased.

3.1.3 The significance of corpus linguistics for this study of spoken success

To sum up the introduction to corpus linguistics, effort should be made to emphasise why it is so key to this study of success. Though evidence from real language use can lead to new insights, it is generally acknowledged that greater focus has been lent to written rather than spoken corpora (Adolphs & Knight, 2010). As Guilquin, De Cock and Granger (2010: vii) highlight, in comparison with written corpora, "spoken corpora tend to lag behind". This in part is due to the time, expense and effort their construction exhausts (Thompson, 2004) but with the greater attention paid to NS rather than to learner language, it is unsurprising that gaps still remain for *learner speech*. However, though particularly large learner corpora now exist (the 3.7 million words International Corpus of Learner English – written (Hilgismann, n.d.), EVP/EGP – written, the million word Louvain International Database of Spoken Learner English [LINDSEI] – spoken (Gilquin, De Cock & Granger, 2010)), emphasis is still often placed on error identification and comparison. Clearly, whilst consensus is starting to promote comparison between learner groups as opposed to learners and NSs (Hunston, 2016), the risks of measuring success against NS norms and according to their perceived 'accuracy' still persist. Of course, comparing learners and NSs of a language is not an invalid course

of action in research (Aston, 2008), but in doing so, it can add to the feeling that learner language is in some way a “degenerate form of the target system” (Bley-Vroman, 1983: 4). This is why examining learner speech is so vital. As the previous chapter presented, since the CEFR does not provide any tangible, distinct descriptions of the learner language to be evidenced at different levels, corpus-driven research can be called upon to reveal what is characteristic of their speech, and what ultimately makes them successful as users of the target language. Furthermore, as the next section will demonstrate, the notion of accuracy, or the notions associated with dictating what should or should not happen in spoken language use, are not as clear-cut as they may initially seem.

3.2 Corpus findings for spoken grammar and their implications for success

Though speech may previously have been overlooked or not fully exploited in corpus research, there has been a growing availability and interest in the findings from spoken corpora (Cullen & Kuo, 2007; Leech, 2000). Studies and grammars such as the *Longman grammar of spoken and written English* (Biber et al., 1999) and the *Cambridge grammar of English* (Carter and McCarthy, 2006) have shown that features of spoken language should not be dismissed as ‘wrong’ simply because they do not ‘fit’ with the prescriptions of written grammar. Understanding the distinctions between written and spoken corpora is thus important as fully formed conclusions can lead to more positive appraisals of learners’ spoken achievement in a target language. The purpose of this section is therefore to document the benefits and characteristics of spoken grammar for learner success. Whilst Section 3.3 concentrates specifically on corpus findings regarding lexis, it is first important to explain the origins, relevance and make-up of a spoken grammar.

In its traditional sense, grammar has been criticised for being “rooted” (Carter & McCarthy, 1995: 141) or “too wedded” (Leech, 2000: 689) to the written form of language. As such, the accuracy or acceptability of spoken features has often been subject to the ability of written grammar to precisely and comprehensively explain speech. For this reason, non-conforming characteristics, such as those containing ellipsis *What you up to?* for example, have previously been labelled ‘wrong’ or ‘deviant’ from expected norms despite the suitable and inconspicuous nature of particular spoken features. However, as Carter and McCarthy (1995: 142) pronounce, “If our speakers are ‘wrong’, then most of us spend a lot of our time being ‘wrong’.” This sense of deviancy or inappropriacy thus heightens the restraints placed upon true descriptions of speech by

written grammar's inability to explain forms widely seen in spoken communication (McCarthy & Carter, 1995). In a sense, it prioritises what *should* be said at the expense of what is *commonly* said. Of course, this is not to say that written grammar cannot be applied at all to speech – some studies have in fact highlighted the 'sameness' between spoken and written grammars (see Leech, 2000) – but it has nonetheless prompted spoken corpus research to evidence that non-standard grammatical forms are still adequate and acceptable in spoken communication:

“External evidence points us toward a socially embedded grammar, one with criteria for acceptability based on adequate communicability in real contexts, among real participants. It is evidence that cannot simply be dismissed as “ungrammatical.””

(McCarthy, 2006; 32)

In spite of this view, there has previously been a noticeable pedagogical emphasis on written grammar in the classroom (Cullen & Kuo, 2007; McCarthy, 1995). Though this situation is changing, the argument for teaching learners spoken grammar is that by providing them with a choice of written or spoken features, they will be able to respond more flexibly to a wider range of language use contexts (McCarthy & Carter, 1995). For success in speech, if learners are not given opportunities to learn particular characteristics of speech, they could either transfer features from their L1s or they could transfer features from written grammar which could make them sound less natural or unnecessarily formal: e.g. in a conversation between friends, the utterance *in conclusion the decision has been made not to attend the get-together* would be grammatical in the written sense but entirely inappropriate for the context. However, writers have acknowledged that a gap exists between the findings from spoken corpora and practices in the language learning classroom (Cullen & Kuo, 2007). The question arises, therefore, as to whether learners utilising spoken grammar would be penalised for the use of speech deviating from written grammar despite an appropriate, more flowing communication. Denying them the opportunity to learn about spoken grammar could actually reduce their ability to be successful in speech since:

“In learning a second or foreign language, the goal of most learners is to use the language productively or receptively in communication. We learn languages in order to use them.”

(Leech, 2000: 686)

To enable teachers and learners to familiarise themselves with the distinct characteristics of spoken grammar, corpus research studies have produced a number of outputs which detail the characteristics of spoken grammar. For instance, McCarthy and McCarthy (2001) highlight ten considerations for the similarities and distinctions between written and spoken grammars that need to be acknowledged if it is to be understood fully and therefore applied to teaching (McCarthy and Carter, 2001: 2-21):

“1. Establishing core units of a spoken grammar”

Observation of transcripts of speech demonstrate that the identification of ‘well-formed’ units is not a simple task. Using labels such as main and subordinate clauses is problematic as units may exceed one speaker’s turn or clauses may be incomplete due to the co-construction of speech between speakers.

“2. Phrasal complexity”

The complexity of units in speech and writing may differ significantly. Though a complex structure *could* be used in speech, evidence from corpora suggests that it does not transpire in most cases due to the additional, or unnecessary, formality it lends to utterances.

“3. Tense, voice, aspect and interpersonal and textual meaning”

The interpersonal nature of speech influences the grammatical choices speakers make. As it is real-time, often face-to-face, speakers may change tense, aspect or voice in order to be less direct, tentative, polite or emphatic (when foregrounding or backgrounding details).

“4. Position of clause elements”

Language teaching is often facilitated via ‘strict’ descriptions of grammar but such rules do not apply to freer clause positioning. Dislocations relate to this criterion but in actual fact, they are not deviant or inaccurate; they simply reflect the unplanned nature of speech.

“5. Clause-complexes”

Some subordinate clauses, such as those containing *which*, *because* or *cos* can be ambiguous in terms of whether they satisfy subordinate clause definitions. Often performing an evaluative purpose, they more closely resemble main clauses.

“6. Unpleasing anomalies”

Unpleasing anomalies defy traditional grammatical description but due to their repeated occurrence across language, they cannot simply be dismissed. Containing features such as double negatives, they can be said to “go against the grain” (McCarthy and Carter, 2001: 14)

“7. Larger sequences”

Larger sequences involve stretches of language comprising several turns, sentences or paragraphs. In some cases, features like tense and aspect can evolve across those sequences but the main point made here is that a discourse-related approach to language is sometimes needed if it is to be fully appreciated.

“8. Comparative-criterion”

Stronger and weaker forms of the written-spoken grammar are evident in literature but assuming either stance could result in the overlooking of overlap or differences that they can both offer. Presenting spoken grammar to learners as a completely new form would also do them a “disservice” (McCarthy and Carter, 2001: 17) so it is important to hold a balanced view.

“9. Metalanguage”

Applying metalanguage from written grammar to describe features in speech can sometimes leave researchers looking for a more appropriate or precise terminology. Similarly, terms from written language bring with them their own connotations and metaphor which could influence descriptions of spoken grammar.

“10. Native and non-native users”

Just as Chapter 1 highlighted, the implications and limitations of a NS model for learners, the question of whose spoken grammar to present also becomes a pertinent matter. Learners aspire to what they consider to be standard norms but they may not always be appropriate for learners and their contexts.

Similarly, corpus research has also been able to illustrate how the nature of the medium and its typical features lead to changes in the way language is used. As Section 1.5 highlighted earlier, speech is typically a real-time, unplanned, face-to-face encounter. It commonly takes place in a shared, immediate context which requires learners to use their language efficiently and effectively so as to reduce the demands placed on speakers and listeners. It is interactive, it is often co-constructed by its participants, it reflects the shared, immediate context in which speakers find themselves in, it is often simplified and repetitive in nature, and it is expressive of emotion or attitude (Biber et al., 1999; Carter & McCarthy, 2006; Hughes, 2011; McCarthy, 1999). For these reasons, a range of features such as discourse markers, *you know*, ellipsis e.g. *You want to go out tonight?*, pauses and fillers, *erm I'm not sure*, hedging *it's kind of a big problem*, pronoun usage *it isn't as easy as she said*, deitic markers *that one over there*, and contractions *I don't wanna go*, are characteristic of speakers who do not need to include such a high degree of precision or detail in their utterances and who are creating language at the same time

as producing it (see Biber et al., 1999; McCarthy, 1998; McCarthy, 1999). If such findings are overlooked where learners are concerned, it is clear that misconceptions about what makes them successful may arise. For instance, a learner utilising fillers or pauses frequently may appear less fluent, the use of basic nouns, verbs or adjectives could make their vocabulary appear more simplistic, and any repetition or repair could be taken as a sign of a lack of control. However, since corpus research has revealed that such features occur frequently in spoken conversation, it would be a little unfair to expect learners to do more than is natural in their speech.

In sum of this review of spoken grammar, it is clear that neglecting to teach learners forms they may encounter or produce in speech may skew their ideas of what it means to be accurate and successful in speech but likewise if tutors and assessors do not acknowledge some of the main distinctions between writing and speech, their judgements of learners' success may also be swayed. As explained, the creation and awareness of spoken grammar can allow for a truer, more realistic depiction of the features of speech; it can also be advantageous for speakers wishing to sound more natural or for those needing to reduce the demands placed upon working memory as they construct meaning.

3.3 Corpus findings for lexical competence and their implications for spoken success

Section 3.2 concentrated on the conclusions drawn from corpora as to the relevance of spoken grammar in a study of success. It highlighted that whilst spoken and written grammars are not entirely contradictory of each other, recognition of the unique but typical features of speech needs to be gained if learners are to be adequately judged in their spoken performance. Since the RQs in the introduction (Section 1.6) focus heavily on the lexis produced by learners, it is necessary in this section to explore the findings of previous corpus research regarding i) the vocabulary learners are said to display, and ii) the significance of formulaic language for learner success in speech. A key principle underpinning this section, and thesis as a whole, is that language “consists of grammaticalised lexis, not lexicalised grammar” (Lewis, 2012: vi). In this sense, words are no longer seen as building blocks attached to a preconceived grammar; instead it is the lexis itself which is responsible for constructing meaning (ibid.). Hence, it is necessary in this section to devote attention to the words and chunks which comprise learners' lexical competence before a full rationale for the RQs to be answered can be given. Though the English Profile studies (CUP, 2015a, 2015b) have contributed significantly

to such aims, they will not be discussed explicitly due to their written language foundations.

3.3.1 Vocabulary size

The “basic dimension of lexical competence” is vocabulary size (Meara, 1996: 37). Despite the numerous ways in which vocabulary size can be defined (see Goulden, Nation and Read, 1990; Meara, 1996; Nation, 2001), here it will be discussed in the rather loose sense of knowing individual words (Lewis, 2012) and in the sense of knowing word families containing the inflections and derivations stemming from one word root (Nation, 2001; Schmitt, 2008). Learned for its communicative purpose (Laufer & Nation, 1995), vocabulary aids the construction and comprehension of meaning, enhances acquisition of new vocabulary and extends knowledge of the world. For learners, not only is the learning of vocabulary fundamental to second language competence and mastery (Laufer, 1998; Lewis, 2012; Schmitt, 2008; Stær, 2008), but it is also extrinsically linked to success as, disregarding other factors, learners with larger receptive or productive vocabularies are seen as more proficient (Meara, 1996; Stær, 2008; Taylor, 2011). For instance, it has occasionally been claimed that the “striking difference” between NSs and learners is the amount of vocabulary they know (Laufer, 1998: 255) as vocabulary size can ‘give them away’ despite an otherwise “impeccable” language level (Hasselgren, 1994: 250 cited by Götz, 2013: 65). Facilitative of student performance in reading, writing, listening and speaking (Chujo, 2004), it is therefore clear that research should aim to discover what or how much vocabulary learners require in order to be successful. In addition, since English is rich in its vocabulary (Götz, 2013), it is necessary to determine how many words are ‘known’ by NSs and how many of these are actually needed by learners in order to *use* the language (Nation, 2001).

Corpora have made significant gains in pinpointing the vocabulary size of NSs using frequency (see Laufer & Nation, 1999; Nation, 2001) and coverage: the percentage of “known words in a piece of discourse” (Van Zeeland & Schmitt, 2013: 457). With frequency studies often grouping words into bands of 1000, researchers have been able to calculate how many words are needed by speakers of a language to surpass various lexical coverage ‘thresholds’ (Laufer & Nation, 1999). It has been calculated that an adult NS will know approximately 17,000-20,000 word families (Goulden, Nation & Read, 1990; Nation, 2001; Nation & Waring, 1997). This immediately infers that learners

require a vast number of words if they are to satisfy any long-term learning goals (Nation, 2001) but when word families are explored in terms of the individual words they contain, such a task becomes even more daunting as vocabulary of only 6000 word families can necessitate a knowledge of over 28,000 word forms alone (Schmitt, 2008; Stær, 2008). Evidently, such a target will not be achievable nor indeed necessary for most learners. However, whatever the figure set, coverage is equally important as it clarifies how much of a particular language medium can be satisfied with a particular number of words. For instance, in speech:

“the speaker needs to have appropriate words at disposal to articulate speech fluently, while the listener needs to recognize words at an adequate speed to understand the speaker.”

(Zhang & Lu, 2014: 285)

The need to know all words in a language therefore diminishes as knowing only the words useful for communication becomes the more appropriate and time effective objective in language learning. Fortunately, there is a broad consensus that a vocabulary of 2000 word families will satisfy most language demands (Götz, 2013; Laufer & Nation, 1999; McCarthy, 1999; Nation, 2001; Nation & Chung, 2009; Nation & Waring, 1997; Schonell, Meddleton & Shaw, 1956 cited by Adolphs & Schmitt, 2003; Stær, 2008; Thornbury & Slade, 2006). Though many studies have focused on written language, estimates of 2000-3000 word families have been suggested for “everyday conversations” (Götz, 2013: 64) as they can reach coverages of up to 95% in speech (Adolphs & Schmitt, 2003; Schonell, Meddleton & Shaw, 1956 cited by Adolphs & Schmitt, 2003). The first 2000 words in English are thus said to encompass the “heavy duty” vocabulary (McCarthy, 1999: 4) which, due to their high frequency and coverage, can provide a strong foundation for meeting individual learners’ needs (Thornbury & Slade, 2006). However, one noticeable gap in this research concerns how this figure compares to learner speech. Both Schonell et al., (1956) and Adolphs and Schmitt’s (2003) studies based themselves on NS speech so claims as to the coverage provided by the 2000 most frequent words in English have not been substantiated for learner speech nor speech at different proficiency levels. The CEFR, as mentioned, offers no elaboration as to changes in B1, B2 and C1 vocabulary sizes other than the rather ambiguous descriptors in Figure 13 (CoE, 2001: 28-29). Similarly, though Laufer and Nation (1999) indicate that the usage of frequent words reduces as proficiency increases no confirmation has been offered from

a learner corpus of speech. It is difficult, therefore, to determine whether the first 2000 words in English do provide such high coverages in speech and whether this figure changes across CEFR levels. Ultimately, it is difficult to pinpoint how learners can be successful in their speech if the vocabulary sizes forming the basis of their lexical competence remain unknown.

Figure 13: CEFR scales for vocabulary range (CoE, 2001: 112)

	VOCABULARY RANGE
C2	<i>Has a good command of a very broad lexical repertoire including idiomatic expressions and colloquialisms; shows awareness of connotative levels of meaning.</i>
C1	<i>Has a good command of a broad lexical repertoire allowing gaps to be readily overcome with circumlocutions; little obvious searching for expressions or avoidance strategies. Good command of idiomatic expressions and colloquialisms.</i>
B2	<i>Has a good range of vocabulary for matters connected to his/her field and most general topics. Can vary formulation to avoid frequent repetition, but lexical gaps can still cause hesitation and circumlocution.</i>
B1	<i>Has a sufficient vocabulary to express him/herself with some circumlocutions on most topics pertinent to his/her everyday life such as family, hobbies and interests, work, travel, and current events.</i>
A2	<i>Has sufficient vocabulary to conduct routine, everyday transactions involving familiar situations and topics.</i>
	<i>Has a sufficient vocabulary for the expression of basic communicative needs. Has a sufficient vocabulary for coping with simple survival needs.</i>
A1	<i>Has a basic vocabulary repertoire of isolated words and phrases related to particular concrete situations.</i>

3.3.2 Formulaic language: Collocation and lexical chunks

As demonstrated, the previous section treated vocabulary size in the ‘loose sense’ of knowing individual words. Indeed, both in the research and pedagogy of lexis, individual words were the main focus of attention until rather recently as they embodied “the central units to be acquired” (McCarthy, 2006: 8). Such an outlook, though, fails to appreciate the links generated in meaning and form when individual words co-occur and the impact they have for language descriptions when they recur frequently. However, corpus research has been able to highlight the lexicogrammatical nature of language which acknowledges the “growing awareness that much of the systematicity of language is lexically-driven” (Schmitt et al., 2004: 55). It has led to a re-evaluation of lexis accentuating the formulaic pairing and grouping of words into collocations and multi-word units [MWUs] as opposed to their treatment in isolation. It has also led to changes in the way speech for NSs *and* learners is understood. Since an expansive body of existing

research and literature has explored formulaic language, discussion here will first briefly explain the principal theories underpinning such a shift before the focus is directed specifically to collocation, lexical chunks and their implications for success.

The majority of research concurs that considerable portions of discourse comprise word patterns rather than grammatically ordered individual words (Lewis, 2012; McCarthy, 2010; Wray, 2002). For collocations and MWUs, the fundamental concept is that a single meaning can be attached to expressions extending beyond the use of one individual word (Schmitt, 2000). However, this is not a new idea. Chomsky's (1965) postulation that a person's competence accounts for an infinite, rule-based ability to generate meaning has been strongly challenged by those asserting that such bespoke or ad-hoc generations would instead signify a lack of "nativelike control" (Pawley & Syder, 1983: 193) or would only be needed "if all else fails" (Becker, 1975: 28). As McCarthy (2010) declares, it is highly unlikely that speakers construct from scratch each utterance in their discourse. In fact, it has been said that theories such as Chomsky's focus on what is *possible* rather than what is *natural* (Hoey, 2005). One key theory supporting this idea is Sinclair's (1991 cited by Erman & Warren, 2000) distinction between the *idiom principle* and the *open choice principle* during language production. Related to the concept that discourse is either composed of pre-fabricated language which is repeated or of original language, the idiom principle argues that:

"a language user has available to him or her, a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments"

(Sinclair, 1991 cited by Erman & Warren, 2000: 30)

It is Sinclair's belief, therefore, that language is mostly created using common, multi-word patterns which are likely to generate meaning as a whole rather than by combining words independent of each other. Involving the use of collocation, fixed expressions, and conversational routines in language production (Adolphs & Carter, 2013), Sinclair's idiom principle, similarly to Becker's (1975) perspective, holds that the selection of patterns always precedes the open-choice principle which is activated only when necessary. In a similar vein, Hoey's (2005) more recent theory of lexical priming explains that words are primed to co-occur: one word immediately prompts the use of a particular 'target words' in relation to the context surrounding their usage. Working both receptively, for contexts in which a speaker will not partake, and productively, where they

can actively participate, it is experience with the target language that results in an increase of, rejections of or additions to an existing priming. With such ‘experience’ having implications for “social, physical, discoursal, generic and interpersonal context” users of a language will build up a bank of associations that are expected to occur within particular situations (Hoey, 2005: 11). The significance for this discussion of formulaic language is that yet again, the dominant role of lexis over grammar is highlighted. Hoey’s (2005: 12) argument is that learners should be provided with real language data to help them formulate primings as they constitute the “driving force behind language use, language structure and language change.”

The relevance of corpus linguistics in illustrating the formulaicity of language has been that it has revealed how central collocations and MWUs are to language. As Section 3.1.2 highlighted, intuition is unable to reliably draw conclusions of patterning and frequency in language but corpus linguistics has allowed further exploration in terms of type, occurrence and usage (Adolphs & Carter, 2013; Ellis et al., 2008; Hunston, 2010). With respect to collocations, definitions vary from those emphasising their methodological influence to those highlighting their relevance for language. In the former, collocation signifies two or more words in adjacency or proximity of each other (McEnery & Hardie, 2010: 123). They co-occur frequently enough that their occurrences cannot be attributed to chance alone (Greaves & Warren, 2010: 212), but they surpass a threshold of possibility and statistical significance (see Section 5.3.2) that confirms their status as a meaningful collocation (Hunston, 2002; McEnery & Wilson, 2001). Collocations are also considered fundamental to the creation of meaning (Adolphs & Carter, 2013; Lewis, 2012) despite being unconnected to pragmatic function (Nattinger & DeCarrico, 1992). Specifically:

“Collocations are not absolute or 100 percent deterministic, but are the probabilistic outcomes of repeated combinations created and experienced by language users.”

(McCarthy, 2006: 8)

In this sense, collocations relate to tendencies rather than exclusive, fixed bonds between words (Wray, 2002). In contrast to the previous section on vocabulary size, they infer that a much greater body of lexical items is stored within a language user’s mental lexicon. Meaning is not only contained within single words, but across them too; for example the

compound *white noise*, the idiom *out of the blue*, and the phrasal verb *to miss out* being cases in point.

On the other hand, lexical chunks, a form of MWU, tend to perform a more functional role in language (Schmitt, 2000); they also differ from collocations in terms of length and role (see De Cock, 1998; Granger, 1998; Nattinger & DeCarrico, 1992). In terms of length, whilst collocations can primarily be associated with word pairings (see Adolphs & Carter, 2013: 23) e.g. *fussy eater*, *tall man*, *commit fraud*, *stark reminder*, it is clear that lexical chunks extend “far beyond” this level (Schmitt, 2000: 400). They can be short, long or “anything in between” (Schmitt & Carter, 2004: 3) but they are typically expected to be between 2 and 4 words long (McCarthy, 2010). With regards to function, although they are relevant for conveying meaning or “referential” topic-related information e.g. *to be in a film* or *favourite restaurant is*, they are considered to be of higher importance in realising pragmatic and discourse functions (De Cock, 1998: 69). They cannot be assumed however to be syntactically complete e.g. *in the*, *top of*, *the end of* (Adolphs & Carter, 2013; Biber et al., 1999; De Cock, 1998), nor are they restricted to one sole purpose (Erman & Warren, 2000). For this study, therefore, they will be defined as:

“a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of the use, rather than being subject to generation or analysis by the language grammar.”

(Wray, 2002: 9)

Though a multitude of terms exists (*ibid.*), the definition provided being that of ‘formulaic sequences’, the term lexical chunks has been chosen as it is considered the most common (Schmitt, 2000) and because it shares similarities with lexical *bundles* in that their occurrence is specifically related to a given register, in this study’s case, learner spoken interaction (Biber et al., 1999).

As Wray’s definition above indicates, lexical chunks are clearly of particular importance for this study into successful speech due to the pragmatic advantages they can offer both speakers *and* listeners as they comprehend and construct meaning in real-time communication (Adolphs & Carter, 2013; Conklin & Schmitt, 2008; Ellis et al., 2008; Götz, 2013; Lewis, 2012; McCarthy, 2010). Firstly, communication requires the purposeful exchange of information (see Section 2.2). Lexical chunks can both facilitate the *transaction* of communicative content (Conklin & Schmitt, 2008), as well as enable

speakers and listeners to communicate on an *interactive* level. Secondly, as they are retrieved as ‘ready-made’, whole units, they can have considerable advantages for fluency during real-time speech as they lessen the demands on producing meaning spontaneously (Conklin & Schmitt, 2008; Erman & Warren, 2000). Similarly, a number of studies have also concluded that some less fixed chunks allow a certain degree of flexibility due to ‘slots’ which are sometimes contained. Whereas collocations only allow “limited substitution” (Paquot & Granger, 2012: 136), e.g. it is possible to *bake a cake* but not *bake a chicken*, lexical chunks (exclusive of idioms) can be amended according to the speaker’s needs at that time. Chunks requiring the addition of a word(s) according to the open choice principle e.g. *a lot of time*, *a lot of ___*, *a few ___ ago*, can therefore provide a degree of flexibility which can be exploited by the speaker (Erman & Warren, 2000). For learner language in particular, lexical chunks have significant implications for perceptions of complexity, accuracy, fluency and assessments of learners’ holistic proficiencies (Boers et al., 2006; Paquot & Granger, 2012). If a learner were to compose each utterance word-by-word, there would be a higher probability that they could “misfire” thus affecting accuracy (Field, 2011: 81); it would also increase the effort that is required and affect the overall fluency with which they speak (Brand & Götz, 2013). Similarly, they are also beneficial for beginners who can become familiar with their meaning and use regardless of the words they contain (Schmitt & Carter, Field, 2011; Thornbury & Salde, 2006).

Considered once again to make learners appear more “native-like” (Granger, 1998: 145), if that is their target, the learning and use of lexical chunks is therefore deemed to be a “goal” for learners (Spöttl & McCarthy, 2004: 191). Though language teaching may have taken time to react and may have been fraught with decisions about which chunks to teach, materials and approaches to teaching vocabulary are catching up. Researchers have previously stressed that the English Language Teaching profession should meet demands to teach vocabulary according to formulae rather than in isolation (Erman & Warren, 2000; Granger, 1998; Nattinger & DeCarrico, 1992; Lewis, 2012) but if learners are to truly benefit and reap the benefits that lexical chunks can have on success, this change in perspective needs to be widespread. For learners to be successful and efficient in their speech, the extension of the mental lexicon to store knowledge of individual words, collocations and lexical chunks “should not be seen as an optional extra...[as they] are extremely frequent, are necessary in discourse and are fundamental to successful interaction” (Adolphs & Carter, 2013: 36). Given that over 55% of NS

language hinges on the idiom principle and that formulaic language is more frequent in speech than writing (Biber et al., 1999; Erman & Warren, 2000), treating vocabulary as a collection of individual words seemingly overlooks a fundamental feature of language production in general. Furthermore, with direct comparisons of NSs and learners revealing that the latter produce fewer lexical chunks, it is clear that reduced usage could hinder learner's abilities to be seen as successful in their speech.

3.4 A brief rationale of the research questions

By way of concluding this chapter, brief justification of the research questions posed will be offered. In doing so, information from this and the previous literature review chapter will be applied to learner success in speech.

RQ1) What percentage of the words used by successful B1, B2 and C1 learners come from the first 1000 and second 1000 most frequent words in English?

Section 3.3.1 established that a basic vocabulary of 2000 word families in English serves the majority of language users' needs. With coverages calculated as reaching approximately 95% for NSs, it is clear that words beyond this figure are not used in comparable frequencies. However, most studies have focussed on reading or writing and the two studies involving speech centred on NSs. It is necessary therefore to ascertain how many words in learner speech originate from this 2000 word vocabulary. By repeating this analysis at B1, B2 and C1, it is hoped that CEFRL descriptors for vocabulary range can be further developed according to the similarities or differences discovered.

RQ2a) What were the 20 most frequent words at B1, B2 and C1 and their notable collocations, colligations and functions?

RQ2b) What were the 20 most frequent keywords at B1, B2 and C1 and their notable collocations, colligations and functions?

RQ2c) What were the most frequent 3- and 4-word chunks at B1, B2 and C1 and their notable collocations and functions?

In corpus linguistics research, a variety of "routine procedures", including word frequency, keyword and chunk lists, can be undertaken in order to uncover facts about

the language under inspection (Kennedy, 1998: 244). In addition to extracting common or significant features, comparison with other corpora can distinguish meaningful and characteristic traits of the language under inspection. As a continuation of RQ1, RQ2a-c have been asked so that learners' vocabularies at B1, B2 and C1 can be further described. Research into collocations and chunks has been included to see whether learners do utilise them, thus potentially promoting their successful spoken language use. The notion of colligation, the patterning and linkage via grammar (Scott & Tribble, 2006), has also been included as many of the routines in language can be realised through them. Though the EVP (CUP, 2015a) and EGP (CUP, 2015b) have already revealed substantial insights into learner language, the vocabulary and grammatical patterns they employ in their language, both have been based on written learner language. This study therefore aims to investigate B1, B2 and C1 language and its differences from a spoken perspective.

RQ3) What Common European Framework of Reference for Languages (CEFR) indicators are present in terms of spoken interaction, spoken production and strategies at B1, B2 and C1 and how are they realised?

Chapter 2 revealed how very little information is offered as to the language used by learners in satisfying CEFR can-do statements. Information is neither provided as to the statements typically fulfilled by learners in their spoken interaction or the differences that may exist across levels. Since the CEFR wishes its can-do statements to act as objectives for language learners, further detail needs to be added as to how they can be realised in speech. Finally, the notion of strategy is implicated in this RQ. Though not discussed in great detail, strategy or skill in using a language very much relies on the lexis learners use. As the following definition from Hulstijn (2007) demonstrates, all five RQs are, in fact, related in terms of knowledge evidenced, the frequent items displayed and the abilities exhibited in speech. Though written with the NS in mind, the following definition can clearly be applied to the language learner also:

“Skill refers to the ability to process phonetic, lexical, and grammatical information receptively and productively, accurately and online. The core of language proficiency restricts this knowledge and skill to frequent lexical items and frequent grammatical constructions that may occur in any communicative situation, common to all adult native speakers regardless of age, educational level, or literacy.”

(Hulstijn, 2007: 665)

3.5 Chapter conclusion

This chapter has introduced corpus linguistics and some of the valuable insights it has produced for research into speech. Not only has research of this type elicited a vast array of recurrent language patterns which can illustrate (to an extent) a speaker's competence, but it has revealed that understanding of speech's grammar, vocabulary and formulaicity, has been strengthened by its findings. For learners, it has also been argued that a lack of adequate appreciation for what is acceptable and common in speech, rather than just in writing, can lead to inaccurate or unfair assumptions regarding what distinguishes them as successful speakers. This potential misgiving needs to be borne in mind if research is to adequately report on the features that lead to positive appraisals of oral proficiency.

4. PILOT STUDY

4.0 Introduction

The focus of this chapter is to outline methodological decisions taken in obtaining and analysing this study's pilot data. Pilot studies present an opportunity to "test...and...finalise" proposed research methodology and data analysis on a small scale before they can be applied to data collection in the remaining study (Mackey & Gass, 2005: 43). They also allow shortcomings in research design to be rectified so as to maximise data and their usefulness, as well as to reduce frustration and additional work later in the process (Dornyei, 2007). In relation to corpus techniques, pilots are of particular worth as they can identify strengths and weaknesses in design and analysis, but more importantly, they can confirm whether the corpus is capable of fulfilling the specific aims of the research (Barnbrook, 1996).

Undertaking a pilot study was therefore essential in this research process for various reasons. Primarily, it helped to test the appropriacy and adequacy of the analytical methods used in answering the amended research questions (see Section 4.4); it also helped to identify issues and challenges influencing the initial construction of the corpus. As corpus compilation requires a great deal of careful preparation, decisions made not only need to take the practicalities of accessing data into account, but they must correspond to the purposes of the research as well as the questions to be asked. Due to this, Sinclair's (2004) guiding principles, encompassing corpus type, compiler, representativeness, balance, sampling and size, were used as a preliminary basis for corpus construction; judgements made were also evaluated in relation to both *ethical responsibilities*, the processes of accessing and recording exam data, and the *aims* of the refined pilot study. Additional justification related to previous corpus research experience. With the researcher possessing no prior experience of compiling a corpus, the pilot would allow for familiarity to be gained in i) corpus design decisions, ii) recording data, iii) transcribing data and iv) analysing data via the use of corpus software. Had experience not been gained in advance of constructing the main study corpus, clear flaws or shortcomings may have been overlooked or the process of compiling the corpus may have been hindered or delayed.

Now that the aims of the pilot study have been highlighted, the objective of this chapter is to report on the corpus compilation procedure, the questions asked, the methods

of analysis, and the results of the piloting stage. It begins by commenting on the decisions made prior to and during the construction of the C1 pilot corpus, it continues by outlining the corpus' demographics and it presents the pilot's amended research questions. After the analytical methodology is explained, results and discussions are given for each individual question. A conclusion is then offered as to the pilot study's implications for success and its relevance to the main study's methodological procedure. Due to the similarities in process and rationale, the present chapter and the following methodology chapter detailing the construction of the B1, B2 and C1 UCLanESB corpus will present some overlap. To reduce repetition, descriptions and justifications relating only to the pilot study will be offered in this chapter; greater depth of detail will be then given in the methodology chapter. References to the UCLanESB speaking exams will also be made during this chapter, but the rationale for their use will only be offered in Chapter 5.

4.1 Pilot corpus type, level and research population

In preparation for the main study, a learner corpus of spoken exam data at one of the three levels was to be compiled. Learner corpora are useful in that they can provide data to aid understanding of the mechanisms and processes involved in SLA as well as factors affecting the learning and teaching of foreign languages (Granger, 2002). Specifically, they are appealing and useful in enhancing descriptions of "actual language in use" (Adolphs, 2006: 97) and, as the name suggests, they consist of a collection of language that is produced by the individuals *learning* the language (Baker, 2010; Hunston, 2002; Litosseliti, 2010). For testing in particular, corpora can also reveal "what people actually do with language" (Barker, 2010: 683). In this pilot study, a learner corpus, reduced in size compared to the final corpus, was hoped to test the design and analytical procedures to be utilised later in the research; it also intended to uncover some preliminary insights into successful spoken language in the C1 UCLanESB exams. The data collected in the pilot study, if successful, would also be incorporated into the resulting B1, B2 and C1 corpus.

The first decision concerned the level to be used in the pilot study. To help test the methodology and analysis, the chosen level had to allow the research questions to be fully evaluated as well as research feasibilities to be fully tested. The C1 level was selected as the focus of the pilot study as it would presumably reveal more about the practicalities of performing such research than the B1 and B2 levels. As the most

advanced level (in this study), it was assumed that it would consist of more words than the other levels: more precise approximations of the time needed for collecting, transcribing and analysing data could be achieved, more informed estimations of the number of exams to be incorporated could be provided and, with more potential lexis, methods for probing data and identifying areas for further investigation could be practised. With regards to the qualitative study of data for research question 3, the C1 level would also provide a more simplified opportunity to practise identifying operational themes in the exam data that would reveal the occurrence of relevant can-do statements. Having fewer can-do statements for spoken production, interaction and strategy usage than the other levels (B1 = 59, B2 = 43, C1 = 26) (see CoE, 2001) meant techniques could be practised on a less-overwhelming scale.

The next factor concerned the research population: the candidates taking the tests whose exam data would form the basis of the preliminary examination into C1 success. For the corpus to be truly representative, the data collected had to match the research purpose well (Biber et al., 1998; Sinclair, 2004). This meant that the candidates' nationalities, L1s, ages, and C1 exam scores needed to be contemplated and controlled. Although UCLanESB speaking exams are taken in the UK, Greece, China and Spain, the data had to be representative of more diverse UCLan context (see Appendix 5). Choosing a monolingual setting could have resulted in a corpus study into success applicable only to that group of learners, and not to the multilingual population of candidates at UCLan. The C1 data was therefore collected from learners ranging in L1 at the Foundation Campus. This simultaneously helped to control the factor of age as all students at the university are 18 years of age or over. Had Greek speaking exam data been used, the age groups would have varied considerably. Including child learner data could have had a detrimental effect on the application of conclusions about successful language (if mixed with adult learner language) to the UCLan context (for discussions regarding the effects of age on SLA, see Cook, 2008; Dornyei & Skehan, 2003). Further complications would have arisen due to the procedures in place for collecting data ethically as additional safeguards would have been needed for younger participants (see British Association for Applied Linguistics [BAAL], 2006; UCLan's Ethical Principles, 2012a). The most significant factor relating to corpus representativeness, however, involved ensuring that the exam data coincided with the study's purpose. As a study of success, it was clear that failed exams could not be used; similarly, exam pairs or triads in which one or more candidates failed would also have to be excluded from the C1 pilot corpus. Pinpointing

other criteria for maintaining representativeness of success was not as straightforward. A UCLanESB speaking exam candidate can secure a pass by scoring an global mark of 2.5 (see Appendix 3). This borderline pass mark, though, would not have been sufficient in demonstrating that the candidates exhibited a solid pass performance. Conversely, a global mark of 5, the maximum mark, would have been equally problematic; a score this high could have indicated that the student demonstrated a speaking ability beyond the tested level. For this reason, exams in which both or all candidates secured a global mark of 3.5 or 4 were used in the C1 pilot corpus. These two scores were believed to represent a solid C1 ability which was not so low that it would have been insufficient and which was not so high that it would have been potentially representative of a C2 exam.

The final characteristic of the C1 pilot corpus requiring careful consideration concerned the number of exams to be collected and the estimated size of the corpus. It was felt that C1 exams would be rather easy to collect as, on average between 2013-2015, 28% of all speaking exams held at the Preston campus tested this level (see Appendix 5). What was unclear, however, was how many potential exams would remain once exam score and informed consent had been checked. The number of usable exams collected per exam session was therefore unknown. Another unspecified factor involved the number of words that would be yielded per exam. With limited resources and time, the B1, B2 and C1 corpus was always intended to be a small corpus, as is expected of specialised spoken corpora (Baker, 2010), but it was difficult to predict just how small it would be. Previous calculations (see Adolphs & Knight, 2010; O’Keeffe et al., 2007) were thus used to estimate how many words would be produced in a 15-minute C1 exam. Although based on native speaker data, it was surmised that a maximum of 2500 words could be captured per C1 exam. In order to obtain a sufficient amount of data for the C1 pilot study, the researcher aimed to collect 25,000 words, which equated to 10 exams. If this number of exams failed to reach the 25,000 word target, more exams would be added to the pilot corpus.

4.2 Procedure for constructing the C1 pilot corpus

The first stage in constructing the C1 pilot corpus involved ensuring that ethical procedures as detailed in the UCLan (n.d. - a; 2012a) and BAAL (2004) guidelines were adhered to. Although a detailed discussion of this process can be found in Section 5.2.9, a summarised description of obtaining informed consent for the pilot study will be given

here. The first step involved obtaining informed consent from all research participants: the Foundation Campus directors, teachers and students. In corpus linguistics literature, clear advice is given about obtaining written permission to use data from those whose data will be captured in transcripts or recordings (Adolphs & Knight, 2010; Thompson, 2004). This is important as in addition to students' language being recorded and transcribed, interlocutor language will also be captured. However, it is good practice to make sure that other individuals who may be included or affected by the research are aware that it is being conducted and are also contacted for consent (Dornyei, 2007). For this reason, permission to conduct the research and contact teachers was first obtained from the director of studies (DoS); teachers were then consulted and finally, students were contacted. As C1 exam data was gathered from genuine exam sessions at the Foundation Campus, it was vital that no adverse effects arose due to the research process. After discussions with the DoS, the agreed procedure was to inform students a week before the exam day via email that research was being carried out and then to distribute consent forms before students took their exams (see Appendices 6 and 8 for email and consent form). Following the exams being conducted, copies of audio and visual recordings were obtained and then checked against exam scores and consent forms; it was also possible that candidates could have withdrawn in the seven-day period following the exams taking place but no one chose to do so.

Once usable exams had been identified, they were transcribed and examined for the total number of words collected. However, decisions first had to be made regarding the transcription system to be used. Transcription of spoken language use is notoriously complex: not only does transcription often take a very long time, but also representing speech orthographically results in an incomplete impression of the original communication as paralinguistic features are often lost (see Adolphs & Knight, 2010; De Cock, 2010; O'Keeffe et al., 2007; Thompson, 2004; Timmis, 2015). Transcription must also suit the purposes of the research, so whilst multimodal corpora are available and enable the orthographic, prosodic and phonetic features to be recorded (Adolphs & Knight, 2010; Thompson, 2004), the researcher opted for what De Cock (2010: 124) terms a "broad" transcription: information relating to *what* was said is favoured over *how* it was said. Suitable for pedagogic purposes (Timmis, 2015) and optimal for increasing readability, such broad transcription can also facilitate studies restricted by time or budget (Crowdy, 1994). The CANCODE transcription conventions (Adolphs & Carter, 2013) were thus adopted. Initially, a group of transcribers trained in using CANCODE assisted

with the volume of transcription to be undertaken but it was soon discovered that discrepancies and inconsistencies in transcripts occurred frequently. Proofreading transcriptions and maintaining consistency therefore took much longer than was anticipated. Despite it delaying the completion of transcription, it was decided that the researcher alone would listen to and transcribe all of the remaining exams: any time saved by using a team of transcribers was ultimately spent reading exams thoroughly to check that they were correct. In total, each exam took approximately two hours to transcribe using the CANCODE system.

After collecting, transcribing and proofreading the successful speaking exams, analysis of the data took place. Before detailing the analytical methods used however, the final C1 pilot corpus demographics and the research questions used in the pilot study need to be presented.

4.3 Pilot study demographics

The final C1 pilot study corpus demographics are shown below in Table 1. Although initial estimations predicted that ten exams would achieve the target of 25,000 words, they did not suffice. Ten exams (including the examiner) comprised 19,730 words (15,742 words without the examiner) so another two exams were added. Containing 12 exams, the final pilot corpus provided a final wordcount (including the examiner) of 23,359 tokens (individual words). Despite still not reaching the 25,000 word target, time constraints meant that no other C1 exams could be added: waiting for another exam session was not feasible at that stage. With regards to learner nationality, of the 25 candidates who participated in the selected exams, 15 (60%) were L1 Chinese. Ideally, a greater mix of L1s had been hoped for. However, upon receiving data from UCLan's International Office (see Appendix 5), it was found that Chinese students comprised 54% of student nationalities within the School of Languages, Literature, and International Studies. Though nationalities across the university would differ, this cross-section would arguably be more representative of students who would typically take UCLanESB exams. After reducing the available number of exams according to exam score and consent, being selective with nationality was not practical: *very* few exams would have remained. The Chinese majority had the potential to skew findings and their applicability to all learner nationalities at UCLan, but the majority of international students do come from China. Although not entirely ideal for the research purpose, it could be argued that the C1 pilot

corpus still corresponded with the UCLan international student demographic. The C1 pilot corpus therefore had the following characteristics:

Table 1: Pilot study corpus demographics

CEFR level examined	C1
Examination venue	Foundation Campus: University of Central Lancashire.
Total word count including examiner	23,359 tokens
Total word count excluding examiner	18,280 tokens
Total number of texts (exams) used	12
Total number of exam candidates	25 (13 males (52%), 12 females (48%))
Candidate L1	15 Chinese (60%) 5 Arabic (20%) 3 Not given (12%) 1 Kurdish (4%) 1 Spanish (4%)
Candidate nationality	15 Chinese (60%) 3 Nigerian (12%) 1 Omani (4%) 1 Emirati (4%) 1 Saudi (4%) 1 Egyptian (4%) 1 Iraqi (4%) 1 Qatari (4%) 1 Columbian (4%)
Average age	22 years
Average time spent learning English	10 years
Average time in UK	9 months

4.4 Pilot study research questions and brief rationale

Before explaining the analytical methods applied, the research questions used in the pilot study will now be outlined. As the research questions in Section 1.6 contain wording relating to all three CEFR levels, they were adapted in order to relate solely to C1, the level selected for the pilot study. The research questions [RQs] used were as follows:

RQ1) What percentage of the words used by successful C1 learners come from the first 1000 and second 1000 most frequent words in English?

RQ2a) What were the 20 most frequent words at C1 (across the entire exam and in Parts A, B, and C) and their notable collocations, colligations or functions?

RQ2b) What were the 20 most frequent keywords at C1 (across the entire exam and in Parts A, B, and C) and their notable collocations, colligations or functions?

RQ2c) What were the most frequent 3- and 4-word chunks at C1 (across the entire exam and in Parts A, B, and C) and their notable collocations or functions?

RQ3) What Common European Framework of Reference for Languages (CEFR) indicators are present in terms of spoken interaction, spoken production and strategies at C1?

It is important to briefly explain the choice of questions used in the pilot study. As stated in the literature review, vocabulary plays an essential role in the construction, comprehension and negotiation of meaning whether it is used for speaking, writing, listening, reading or any other skill (Chujo, 2004; Nation, 1993). Quite simply, vocabulary is the “basic dimension of lexical competence” (Meara, 1996: 37). For this reason, the above questions were used to explore the overall and specific composition of C1 learners’ vocabularies.

RQ1 aimed to ascertain the composition, rather than the overall size, of the vocabularies used by C1 candidates. Since it is unfeasible and unnecessary for learners to achieve a vocabulary size similar to NSs, estimated to be in excess of 30,000 words (see Taylor, 2011), this question focuses entirely on the K1 and K2 word bands to confirm previous studies’ findings that the majority, approximately 80%, of language comprises lexis from these bands. In particular, it aimed to calculate the proportion of vocabulary originating from the 2000 most frequent words in English to see whether the coverage provided to C1 learners was comparable with previous findings. Although investigating C1 vocabulary size in more breadth would have been interesting, it is evident that vocabulary produced in one spoken task would not have been representative of the learners’ full vocabularies. Findings from the research study would also provide a basis for comparison following the incorporation of the B1 and B2 levels.

Once the vocabulary profiles had been produced, RQs 2a-c were posed to allow a closer examination of the individual and chunked lexis required by successful C1 students. One reason for this was to counterbalance a clear drawback to RQ1: K1 and K2 words may afford C1 learners a degree of success, but as the most frequent words in English, it is inevitable that they will also be used by unsuccessful learners. RQs 2a-c would therefore allow a greater degree of detail in terms of i) which words are used

repeatedly by learners, ii) which words appear significantly more frequently in this corpus than more general corpora, and iii) how they are used with regards to collocation, colligation and function. Put simply, these questions sought to pinpoint what it was about the most frequent and key lexis, that made C1 learners successful. As an additional note, any lexis revealed in the pilot would also provide a useful starting point for subsequent research and comparison into the B1 and B2 levels in the main study.

The rationale behind the final pilot study question arose from the CEFR's lack of illustrative learner language use. Readers of the CEFR may be well informed as to what C1 speakers can do, but, as has been highlighted on several occasions, the language used to realise such actions is not specified. The purposes of this research question were two-fold. Firstly, in order to be considered successful, C1 learners must be able to demonstrate certain abilities. The marking criteria in Appendix 2 details the qualities examiners look for during such demonstrations but little knowledge exists as to how the language produced in UCLanESB exams relates to the CEFR's many spoken can-do statements. This is not a study into test design but it would be difficult to regard a learner as successful or unsuccessful if the abilities they should demonstrate remain unknown. As a result, RQ3 aimed to establish which CEFR can-do statements for spoken interaction, production and strategies were relevant and evidenced by C1 learners. This would provide a platform for investigating how such abilities were realised in the language of learners later in the main study. Secondly, this question's rationale was also attributed to a methodological aim. Quite simply, practice and experience had to be gained in identifying and analysing the language of relevant can-do statements so that the process could be made more streamlined for the remaining two levels. Since some of the language in the CEFR can-do statements is repeated and, at times, a little vague, practice would also need to be sought in making the statements more operationalised so that categories would not overlap.

4.5 Methods used for pilot corpus analysis

4.5.1 Analysis used for research question 1

To detect which words comprised the majority of student test language, the researcher had to create a lexical frequency profile (Laufer & Nation, 1995) using the C1 pilot corpus. Such profiles identify the words which arise in learner language but more

importantly, they identify the frequency band to which the words correspond. Put simply, the Lexical Frequency Profile is a “measure of how vocabulary size is reflected in use” (Laufer & Nation, 1995: 307). It is helpful as it allows comparisons to be made with previous text coverage findings such as those outlined in Section 3.3.1. To produce the lexical frequency profile, uncoded text files consisting solely of candidate language use were uploaded into Cobb’s (n.d.-b) *Compleat Lexical Tutor* online software: examiner questions and comments were removed as the sole focus was successful language use by learners. The files were then computed using the ‘vocabulary profile’ function with the BNC-20 used as a measure. The resulting output demonstrated how many *tokens* (“instance[s] of particular wordform[s] in a text” (McEnery & Hardie, 2010: 50)) and *types* (“a particular, unique wordform” (McEnery & Hardie, 2010: 50)) were present in the pilot corpus data along with information about their distribution across the 20 BNC word families. Not only did analysis consist of identifying vocabulary coverage for the first 2000 most frequent words, but token-type ratios were also examined. Token-type ratios, although a more basic statistic, assist in providing a mean score for the repetition of individual word types; they allow texts of similar length to be subjected to preliminary comparisons for lexical variation (Barnbrook, 1996; McEnery & Hardie, 2010). All analysis was also repeated for sub-sections of the test (Parts A, B and C) to establish whether vocabulary coverage differed in any of the three sections at C1.

4.5.2 Analysis used for research questions 2a-c

In order to provide answers for research questions 2a-c, a combination of corpus software tools were used. Question 2a was answered by using Michael Scott’s (2015) *Wordsmith Tools*, a renowned and seemingly well-used suite which allows a variety of corpus analysis procedures to be performed. Firstly, frequency lists, also referred to as wordlists, were created for all 12 exams using coded files. A wordlist was compiled for language use across the entire exam and the three individual sub-sections once again. These were then manually compared to frequency lists for i) the BNC spoken corpus (consisting of 10 million words) (British National Corpus, 2004) and ii) the CANCODE corpus (containing 5 million words) (Carter & McCarthy, 2006) to identify words which seemed to vary considerably in their frequency positioning. These two corpora were chosen as they are both examples of general corpora and they were both compiled using a variety of spoken texts; despite containing native-speaker data, very few specialised or

spoken corpus studies are freely available. *Wordsmith Tools 6.0* (Scott, 2015) were also used to ascertain keyness – “words which are significantly more [or less] frequent in one corpus or another” (Hunston, 2002: 68) – within the C1 pilot corpus in order to answer question 2b. Again, keyness lists were generated for the entire exam as well as the three sub-sections. The reference frequency list utilised for comparison was compiled using BNC World, the second and improved version of the BNC produced in October 2000 (Burnard, 2000). In order to collect and synthesise data for research question 2c, Lawrence Anthony’s *AntConc* (2014) software was used. While similar to *Wordsmith Tools*, the researcher felt that the n-gram (lexical chunk) tools were simpler to use, although admittedly, this is not the strongest of reasons (Gries, 2010). For all frequency and keyness analyses, the use of concordance lines was also incorporated to ascertain how lexical items were used by the C1 exam candidates. Such concordance lines would be valuable in ascertaining the collocations, colligations and functions of certain lexis.

4.5.3 Analysis used for research question 3

Using the C1 pilot corpus as a basis, this part of the pilot study required a change in analytical approach. Moving beyond quantitative corpus analysis, the researcher had to employ qualitative research analysis (see Cohen et al., 2011; Dornyei, 2007) to identify can-do statement occurrence within the test data. To begin with, all productive, interactive and strategic CEFR can-do statements at C1 level were listed; they were then assessed for relevance to the UCLanESB exams as some domains, for instance transactions of goods and services, did not apply. A total number of 26 possible statements were reduced to 7 relevant statements following this process. After ascertaining which statements were relevant, they then had to be made operationalisable to ensure that they did not overlap; overlapping categories in qualitative research can present a serious flaw in validity and can complicate the process (Miles, Huberman & Saldaña, 2014). Finally, exam scripts were then manually examined for can-do occurrence. Examination transcripts containing both learner and examiner speech were first of all read to improve familiarity with the texts and then analysed according to each of the 7 statements. Language deemed to satisfy the relevant spoken abilities suggested by the CEFR was then highlighted on printed copies of the transcripts and then inputted into *NVIVO* (QSR, 2012); can-do occurrence was then counted quantitatively using the statements as codes, and a record of the language they contained was also kept. It is important to note, however, that this approach

may have been exposed to ‘interpretive analysis’ as “the research outcome is ultimately the product of the researcher’s subjective interpretation of the data” (Dornyei, 2007: 38). Had other researchers been involved, can-do statements could have been placed in differing categories. Therefore, selected language deemed to satisfy a descriptor or statement may, to some extent, have been affected by the researcher’s personal interpretations of exam performance.

4.6 Pilot study findings and discussion

4.6.1 RQ1: What percentage of the words used by successful C1 learners come from the first 1000 and second 1000 most frequent words in English?

Table 2 displays the vocabulary profile for the entire C1 pilot corpus. The K-1 band represents words in the first 1000 most frequent words whilst K-2 gives figures for the second 1000 most frequent words.

Table 2: Pilot study C1 Vocabulary Profile - Entire exam

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token %
K-1 Words :	557 (60.61)	806 (64.38)	15598 (91.76)	91.76
K-2 Words :	201 (21.87)	237 (18.93)	863 (5.08)	96.84
K-3 Words :	63 (6.86)	70 (5.59)	178 (1.05)	97.89
K-4 Words :	39 (4.24)	44 (3.51)	86 (0.51)	98.40
K-5 Words :	23 (2.50)	23 (1.84)	41 (0.24)	98.64
K-6 Words :	11 (1.20)	11 (0.88)	24 (0.14)	98.78
K-7 Words :	6 (0.65)	6 (0.48)	11 (0.06)	98.84
K-8 Words :	8 (0.87)	9 (0.72)	11 (0.06)	98.90
K-9 Words :	2 (0.22)	3 (0.24)	3 (0.02)	98.92
K-10 Words :	3 (0.33)	3 (0.24)	5 (0.03)	98.95
K-11 Words :	2 (0.22)	2 (0.16)	6 (0.04)	98.99
K-12 Words :	2 (0.22)	2 (0.16)	2 (0.01)	99.00
K-13 Words :				
K-14 Words :				
K-15 Words :				
K-16 Words :				
K-17 Words :				
K-18 Words :				
K-19 Words :	2 (0.22)	2 (0.16)	2 (0.01)	99.01
K-20 Words :				
Off-List:	??	76 (6.07)	169 (0.99)	100.00
Total (unrounded)	919+?	1252 (100)	16999 (100)	100.00

As can be seen, the vast majority, 91.76%, of the total running words within the C1 data originated from the first 1000 word frequency band (K-1). A further 5.08% came from the K-2 frequency band revealing that for this pilot study at least, successful spoken language at C1 mostly makes use of the first 2000 most frequent English words. A total token-coverage percentage of 96.84% for words in these frequency bands provides the first answer to Research Question 1 in that less than one in 20 words came from a different word family band. As is demonstrated in Extract 1, very few words from beyond the K-2 band (highlighted in **bold**) are needed for learners to convey their message (word in **blue** = K-1, words in **red** = K-2):

Extract 1: Pilot study, C1, Exam 2, Part A

<\$0> Okay brilliant erm so what are the major transport problems in your country?

<\$3M> Okay erm **immediate** transport problems in my country would be the fact that <\$=> erm the erm <\$G?> </\$=> it would be er like the transportation agency or should I say like erm the people er like that **handle** transport are not very **strict**. Erm young people like er people like four years older younger than me like sixteen years olds or fifteen year olds are allowed to drive. Basically it's not allowed in the er law in my country but even if you're fifteen or thirteen they could drive around in a car and if like a policeman should stop you or a road safety person should stop you you could **bribe** them like really low amounts like anybody could afford it and they will let you go so it causes a lot of er **accidents** it causes people driving **recklessly** and everything and er yeah they are not **strict** at all with the laws so it causes a lot of problem and also the roads are really bad and the government **hardly** does anything about it just like a few **states** try to **fix** the roads and make them good most of the other **states** are really bad which cause other **accidents** as well.

Further analysis also revealed that the coverage of word *families*, at first, seemed to correspond to figures in previous vocabulary profile studies which calculated that K-1 and K-2 families granted 80% coverage (Cobb, n.d.-a; Francis & Kubera, 1982 cited by Nation & Waring, 1997; Laufer & Nation, 1999; McCarthy, 1999; Nation, 2001; Nation & Waring, 1997). Of 919 word families in the C1 data, 758 came from the top two bands. For this data, the first 2000 most frequent word families (combining K-1 and K-2 figures in Table 2's second column) provided a coverage of 82.4%. Whilst useful for a preliminary overview of successful C1 learner language in the UCLanESB speaking tests, these studies are not entirely applicable due to their focus on written language; figures for spoken language thus had to be taken into account. Although little quantitative research had identified the impact of vocabulary breadth on spoken language, two studies were of relevance. In a study of informal spoken language use, Schonell, Meddleton and Shaw (1956 cited by Adolphs & Schmitt, 2003) were able to indicate that the first 2000 most frequent word families resulted in a drastically increased coverage in comparison to written language. Whereas words of this frequency constituted 80% of written texts, Schonell et al. (ibid) calculated that a coverage of 96% was achieved with the same vocabulary bands. In an attempt to generalise and confirm these findings with more up-to-date corpora, Adolphs and Schmitt's (2003) study also found that coverage for the two most frequent word family bands in the CANCODE corpus stood at a higher rate of 95%. When comparing these statistics with the C1 pilot study constituting a coverage of 82.4%, it is evident that a stark disparity exists, but it is one that could be explained.

One justification for this distinction involves the people at the centre of the research. Adolphs and Schmitt's study focussed on casual, informal speech from a variety

of settings in the UK and Ireland; this pilot study focussed specifically on non-native learners. Although the CEFR (CoE, 2001: 28) states that C1 learners should be able to demonstrate a “broad range of language” to express themselves, their vocabulary breadth would not equate to that of a native speaker; a case in point is the suggestion that knowledge of word families in the K-1 to K-6 bands would be equivalent to knowledge of 28 000 word types (Schmitt, 2008; Stær, 2008), the same range in the pilot study data revealed usage of only 1191 types. This clearly, is open to question but it nevertheless demonstrates that to be successful in the C1 speaking exams, a native-like vocabulary is not needed. Another possible cause for reduced K-1 and K-2 coverage might be found in the analysis of the type-token ration (the total number of tokens divided by the total number of types). Table 3’s figures reveal that there were approximately 14 tokens per word type. Whilst this information is not available for Schonell et al.’s, nor Adolphs and Schmitt’s studies, one could conjecture that fewer word families are evidenced by C1 candidates because they repeat vocabulary rather often: perhaps they stick to comfortable, more familiar lexical “teddy bears” or they use repetition as a type of filler, planning device or communication strategy similar to the way native speakers do (Götz, 2013: 30). Alternatively, with native speakers expected to have a vocabulary of approximately 20,000 words (Nation & Waring, 1997), C1 students will have a smaller vocabulary to call upon. Nevertheless, a coverage of just over 80% with K-1 and K-2 words allowed the C1 students in the pilot study to succeed.

The final aspect to be considered in answering the first pilot study question related to whether K-1 and K-2 coverage remained uniform throughout the three exam sections or whether it fluctuated. It was found that comparisons of Parts A, B and C in the C1 exam did not present any stark findings. Table 3 presents a summary of the types, tokens and cumulative coverage for parts A, B and C. It also contains data demonstrating the number of tokens per type.

Table 3: Pilot study K1 and K2 words in Parts A-C

Part of exam	Band	Families (%)	Total % coverage in word families	Types (%)	Tokens (%)	Cumm. token	Total tokens per text	Different types	Type-token ratio	Tokens per type								
A	K-1	322 (75.23)	90.40%	399 (76.29)	2998 (91.91)	91.91	3262	523	0.16	6.24								
	K-2	65 (15.19)		70 (13.38)	145 (4.45)	96.36												
B	K-1	388 (70.04)	86.60%	517 (72.01)	6583 (91.44)	91.44					7199	718	0.1	10.03				
	K-2	92 (16.61)		108 (15.04)	385 (5.35)	96.79												
C	K-1	390 (67.59)	86.30%	523 (69.00)	6018 (92.02)	92.02									6540	758	0.12	8.63
	K-2	108 (18.72)		127 (16.75)	333 (5.09)	97.11												

In all of the three sections of the exam, there is little deviation from the 96% figure for total running words originating from the two most frequent bands. However, slight variation can be seen for total word family coverage: whereas K-1 and K-2 family coverage for the entire exam stood at 82.4%, percentages rose slightly. For section A in particular, only one word in ten came from a different frequency banding. In relation to token-type ratio, again, there is some slight divergence. The entire exam produced a token-type ratio of approximately 14 words which suggests that lexical range may not be as broad due to repetition. In Part A, however, this figure falls to 6 tokens per type, meaning that in this part, repetition is not as high. Alternatively, in Parts B and C, average word-type repetition rises to 9 and 10. This could be interpreted in three ways. First of all, students may once again be relying on familiar vocabulary within Part B to convey their opinions during the discussion task; apart from the writing on the prompt sheet, there is little help with vocabulary other than what the students already know and can produce. Secondly, although more analysis would be required, repetition in Part B could represent one way that shared meaning is co-constructed (see Bygate, 1987). By repeating vocabulary within the pair, students may be able to maintain a sustained dialogue whilst help from the interlocutor is unavailable. Finally, higher average repetition in Part B could be a result of exam design. In Parts A and C, students are provided with questions from the interlocutor. These will contain lexis that the student might understand, but not have readily available in their productive vocabulary. The supply of additional vocabulary may thus give students a slightly larger range in vocabulary than had they been left to their

own devices: a broader range with less repetition would be a probable result in these sections.

4.6.2 RQ2a: What were the 20 most frequent words at C1 and their notable collocations, colligations and functions?

The 20 most frequent words within the C1 pilot data are shown below. C1 frequency data is first shown in Table 4 for the entire exam and then in Table 5 for Parts A, B and C, respectively. Before looking at the data more closely, it is relevant to quickly point out that the frequency lists need to be viewed with the exam structure in mind (see Appendix 1). For instance, Part A contains the smallest frequencies but it is the shortest exam section in terms of duration. For candidate pairs, Part A lasts two minutes, whereas parts B and C both last five minutes; for groups of three students, Part A lasts four minutes, whilst parts B and C last 7 minutes. The frequencies of different word types is therefore expected to and indeed does increase as the exams develop, so taking the frequency *and* the position of words in the list into consideration, rather than frequency alone, provided more insight.

Table 4: 20 most frequent words in pilot study - Entire exam

N	Word	Freq.	%	Cumm. Freq.	Texts	%
1	THE	890	5.31	5.31	12	100
2	ER	661	3.94	9.25	12	100
3	I	630	3.76	13.01	12	100
4	AND	487	2.90	15.91	12	100
5	TO	476	2.84	18.75	12	100
6	ERM	400	2.39	21.14	12	100
7	IS	330	1.97	23.11	12	100
8	IN	313	1.87	24.98	12	100
9	THINK	285	1.70	26.68	12	100
10	YEAH	280	1.67	28.35	12	100
11	YOU	277	1.65	30.00	12	100
12	A	248	1.48	31.48	12	100
13	LIKE	241	1.44	32.92	12	100
14	OF	225	1.34	34.26	12	100
15	SO	207	1.23	35.49	12	100
16	IT	181	1.08	36.57	12	100
17	IT'S	172	1.03	37.60	12	100
18	THEY	169	1.01	38.61	11	91.7
19	BECAUSE	144	0.86	39.47	12	100
20	FOR	133	0.79	40.26	12	100

Table 5: Pilot study frequency lists for Parts A, B and C

N	Word	Part A				Part B				Part C					
		Freq.	% of sub-corpus	Cum. %	% Range	Word	Freq.	% of sub-corpus	Cum. %	% Range	Word	Freq.	% of sub-corpus	Cum. %	% Range
1	ER	182	5.66	5.66	91.67	THE	412	5.78	5.78	100	THE	326	5.07	5.07	100
2	I	168	5.23	10.89	100	I	258	3.62	9.40	100	ER	233	3.63	8.70	100
3	THE	152	4.73	15.62	100	ER	246	3.45	12.85	100	I	204	3.18	11.88	100
4	TO	109	3.39	19.01	100	TO	190	2.67	15.52	100	AND	196	3.05	14.93	100
5	AND	106	3.30	22.31	100	AND	185	2.60	18.12	100	TO	177	2.76	17.69	100
6	ERM	88	2.74	25.05	100	ERM	165	2.32	20.44	100	ERM	147	2.29	19.98	91.67
7	LIKE	69	2.15	27.20	100	YEAH	164	2.30	22.74	100	IN	136	2.12	22.10	100
8	IN	66	2.05	29.25	100	YOU	159	2.23	24.97	100	IS	127	1.98	24.08	100
9	MY	56	1.74	30.99	100	IS	158	2.22	27.19	100	THINK	111	1.73	25.81	100
10	IS	45	1.40	32.39	83.33	THINK	139	1.95	29.14	100	THEY	104	1.62	27.43	91.67
11	A	42	1.31	33.70	91.67	A	112	1.57	30.71	100	SO	96	1.49	28.92	100
12	SO	42	1.31	35.01	91.67	IN	111	1.56	32.27	100	A	94	1.46	30.38	100
13	YEAH	41	1.28	36.29	100	OF	98	1.38	33.65	100	YOU	93	1.45	31.83	100
14	GO	40	1.24	37.53	91.67	LIKE	90	1.26	34.91	91.67	OF	87	1.35	33.18	100
15	OF	40	1.24	38.77	66.67	IT	88	1.24	36.15	100	LIKE	82	1.28	34.46	91.67
16	IT'S	38	1.18	39.95	100	HOTEL	81	1.14	37.29	4	YEAH	75	1.17	35.63	91.67
17	THINK	35	1.09	41.04	83.33	SO	69	0.97	38.26	12	PEOPLE	69	1.07	36.70	91.67
18	IT	29	0.90	41.94	91.67	IT'S	67	0.94	39.20	12	IT'S	67	1.04	37.74	100
19	OR	28	0.87	42.81	75.00	WE	66	0.93	40.13	9	COUNTRY	66	1.03	38.77	83.33
20	HAVE	26	0.81	43.62	91.67	FOR	65	0.91	41.04	10	IT	64	1.00	39.77	91.67

Before focusing on particular words in this list, a brief summary of what the data shows will be given. An initial look at the C1 pilot frequency lists reveals that there is very little variation in the first six positions across exam parts: *the, er, I to, end, and* and *erm*. Confirmed by other corpus studies (McCarthy, 1999; Moon, 2010), these words

were to be expected to be highly placed since grammatical and functional words, such as pronouns and prepositions, typically rank high in investigations of frequency, especially in conversation (Biber et al., 1999). To delve deeper into the data, as is suggested by Hunston (2002) and Kennedy (1998), a comparison of the test data against general corpora data from the spoken BNC (Leech, Rayson & Wilson, 2001) and CANCODE was carried out to identify similarities and differences. Whilst some words occupied very similar positions in the C1 pilot and top 20 NS frequency lists (*the, I, and, to, is, in, yeah, of, they, it's, it*), others varied (*er, erm, you, a, that*). Other words also ranked in high positions within the top 20 in the learner data (*think, like, so, because*) or solely in the BNC or CANCODE data (*n't, we, do, was, have, know*). Words which varied clearly identified potential avenues for further investigation that would be borne in mind for the main study. Due to limited discussion, here only the words *er, erm, and you* will be examined.

Er and erm

Fillers such as *er* and *erm* have previously generated debate as to whether they constitute involuntary symptoms of speech, non-linguistic signals for maintaining or finishing turns, or words with their own lexical properties (Clark & Fox Tree, 2002). Nevertheless, it is clear that they occupy very high positions in all three exam parts with percentage proportions peaking slightly in Part A (see Fig. 3-5) whilst in the BNC and CANCODE lists, they occupy lower positions. This is unsurprising given i) the demands of real-time spoken communication for learners and NSs alike, and ii) the nature of the language contained in the corpus.

In real-time communication, fillers “frequently precede important lexical choices” made by speakers during conversation (Carter & McCarthy, 2006: 172). Therefore C1 learners searching their vocabularies for a suitable word or phrase may need to use *er* and *erm*, amongst other fillers to ‘buy time’ during these moments. Since NSs of English will possess larger vocabularies than C1 students, and since proceduralisation is still occurring (see Taylor, 2011), learners may need to use these fillers more frequently which could explain the higher ranking of *er* and *erm* in the C1 pilot corpus in comparison with the BNC and CANCODE lists. Similarly, real-time speaking demands also require learners to make quick decisions regarding the content of their responses or statements. This brings discussion to the latter topic of the corpus’ contents.

Smith and Clark (1993) suggest that *er* and *erm* are commonly used after questions when speakers are unable to provide an answer, are slow or are uncertain. With this information in mind, the fact that the corpus contained exams in which learners are continually asked questions could not be ignored. The decision was thus taken to examine the way that students attended to questions in all three stages of the test. The twelve exam transcripts were inspected to determine how many student responses starting with *er* or *erm* followed questions (see Table 6). It was found that of 275 questions, 71 (26%) posed by examiners and 13 (5%) posed by learners received a response beginning with one of the two fillers; 45 (63%) of these occurred in Part A. This provides *some* support for Smith and Clark's (1993) claims as the data indicates that 84, approximately a third of all question responses started this way. It demonstrates more clearly, however, that Part A does act as a 'warm-up' section: nearly two-thirds of question responses began with *er* or *erm* whereas usage in Parts B and C dropped. This initial data indicated that to be successful, students should not be dissuaded from using such fillers as these words can act as stalling devices when responding to questions. Likewise, learners should not be deemed unsuccessful when a raised appearance of fillers could be evidence of candidates 'settling into' the exam or, in more general contexts, be 'getting used to' the dialogue.

Table 6: Usage of *er* and *erm* after questions

Total number of questions asked			Responses beginning with <i>er</i> or <i>erm</i>		
By candidates	By examiner	Combined total	Following candidate questions	Following examiner questions	Following examiner questions <u>only</u> in Part A
103	175	275	13	71	45

Although this preliminary analysis has not explored all uses of *er* and *erm*, it would be reasonable to argue that these fillers are not always a sign of poor fluency. Instead, they potentially represent a type of conscious or unconscious strategy that students use to fill silence while they search for either content or linguistic features in order to respond to questions. It must be acknowledged, however, that the higher ranking of *er* and *erm* in the C1 pilot data, when compared with the NS corpora, may be explained by the nature of the language captured. With students continually being asked questions in exams, there is a greater likelihood that *er* or *erm* will be employed in this way.

You

From Table 5, it is evident that usage of *you* fluctuated across Parts A, B and C. Occupying third position in the BNC and CANCODE frequency lists but only 11th position in the entire C1 pilot exam data, *you* clearly warranted further investigation. Indeed, closer analysis of concordance data revealed subtle differences in the way *you* was utilised by learners.

Using the CANCODE corpus, Carter and McCarthy (2006) identify three main uses. As a singular, plural or generic personal pronoun, it allows speakers to express generalisations and demonstrate interactivity to others involved in the communication. When used with *know*, it can also act as a discourse marker to acknowledge others' utterances or to fill silences during hesitations; a technique evidenced on 35 occasions but owing to space will be discussed in the main study. Indeed, making generalisations was found to constitute the most frequent function of *you* in the C1 data across Parts A, B and C. Used in the sense of "people in general" (Carter & McCarthy, 2006: 120), there were 11 generalisations using *you* either as a subject or object in Part A (44% of the exam section), 88 (55%) in Part B, and 55 (59%) in Part C. In Part A's utterances, *you* was often used to express statements of ability or permission (see Extracts 2 and 3):

Extract 2: Pilot study, C1, Exam 4, Part A

<\$7F> *Erm I think it must be erm traffic jam because you know <\$/=> China have erm lot erm </\$/=> the population is very large and er a lot of people er get one or more than one cars and erm er especially in the morning or the afternoon erm the traffic is very bad*
+
<\$0> *Mhm.*
<\$7F> + *and maybe <\$/=> you cannot </\$/=> erm you can go just one metre er in ten minutes.*

Extract 3: Pilot study, C1, Exam 9, Part A

<\$0> *Chips and chicken. Right good okay. Right erm and er <\$25F> why do you think people like to eat out in restaurants? Why do people like going to restaurants do you think?*
<\$25F> *Well I think it's easier faster and er you can just er have it as a hangout and a meal at a time +*
<\$0> *Mhm.*
<\$25F> + *you can enjoy it with friends or family +*

Parts B and C similarly evidenced utterances of this type but were combined with conditional structures using *when* and *if* (See Extracts 4, 5, 6 and 7) much more frequently than in Part A (Part A = 1 instance; Part B = 11 instances; Part C = 12 instances). This

demonstrates that while the meaning of *you* remains constant throughout the exam, additional evidence could confirm that the structures surrounding it become more complex in the level-check and probe sections. Successful C1 students may therefore be distinguished by the ability to flexibly adapt and combine structures to convey more complex messages, an ability that may or may not be demonstrated by B1 and B2 learners.

Extract 4: Pilot study, C1, Exam 7, Part B

<\$13M> *And thinking thinking about it's the location maybe because if **you** choose a good location er you can tra= tra= travel convenient and erm you can see a lot of er special points yeah...*

Extract 5: Pilot study, C1, Exam 12, Part B

<\$24M> *+ between university and study time and maybe when **you** work hard like **you** can't sleep very well but the problem is that insufficient sleep cause **you** stress.*

Interestingly, the use of *you* in questions also revealed some variance across the exam sections. Although *you* was clearly most often used in generalisations, it similarly helped learners to ask for clarification, repetition and opinion. Parts A and C, showed some similarity in this respect. As Extracts 6-10 show, requests for clarification or repetition, not always correctly formed, were posed by learners to the interlocutor on five occasions in Part A and ten occasions in Part C:

Extract 6: Pilot study, C1, Exam 4

<\$7F> *Transport <\$07> erm </\$07> **you** meaning the er traffic?*

Extract 7: Pilot study, C1, Exam 5

<\$10F> *Er so **you** mean cultural awareness?*

Extract 8: Pilot study, C1, Exam 7

<\$13M> *Yeah so can **you** say again?*

Extract 9: Pilot study, C1, Exam 8

<\$15F> <\$=> *Er I think the plane is the </\$=> er sorry can **you** explain the topic?*

Extract 10: Pilot study, C1, Exam 6

<\$12F> *Erm could **you** repeat that?*

In Part B, however, the use of *you* in questions posed by learners exhibited a different function: that of seeking opinions. In fact, using *you* in requests for clarification or repetition was not found at all in the C1 pilot data in this section. *You* in Part B therefore adopted a very different role. In addition to using *you* for generalisations and as a discourse marker, *you* was employed on 34 occasions to form a question that would help

the interaction continue and that would elicit another learner's opinion on a subject. Some exams displayed these questions throughout the level-check discussion task but others used them to 'kick-start' the conversation at the beginning. What was unexpected, however, was the variety of questions that were asked and the flexibility demonstrated in forming yes/no questions and wh- questions as shown in Figure 14 below. Being successful at C1 therefore necessitates an ability not only to give opinions and thoughts, but also to adequately seek them whilst simultaneously maintaining the communication, a criterion that will be discussed in Section 4.6.5. In terms of implications for language teaching and materials design, such model questions could be included in textbooks or practice materials, albeit not all at once, to demonstrate the range of question forms that learners may encounter in successful C1 spoken interaction. Whilst a grammatical syllabus would introduce a range of pronouns for similar questions e.g. *has he/she been to...? What does he think about...?* to display changes in auxiliary word forms, the learner language in this pilot study could suggest that particular emphasis should be placed on 'you' and 'I', as a common feature of spoken interaction, before other forms are presented.

Figure 14: Uses of *you* in questions. Part B, pilot study.

Question forms using <i>you</i>	
• How important is it to <i>you</i> ...?	• Have <i>you</i> been to...?
• How do <i>you</i> feel about...?	• What do <i>you</i> reckon about...?
• How important do <i>you</i> think it is?	• How do <i>you</i> think?
• What do <i>you</i> think about...?	• How do <i>you</i> think about...?
• What do <i>you</i> think?	• How about <i>you</i> ?
• Do <i>you</i> agree with me?	• Which do <i>you</i> think is...?
• Do <i>you</i> like...?	• Do <i>you</i> have a ... in mind?
• Where are <i>you</i> from?	• Who do <i>you</i> think is...?
• Do <i>you</i> know...?	• What else do <i>you</i> think about...?

To sum up preliminary insights into the first pilot study question, comparison with a reference corpus, the BNC spoken, has uncovered some findings into the frequent words which were common in the C1 pilot corpus and words which differed considerably. It is clear that words, such as *think*, *like*, *so* and *because* will provide key insights in the next research question and, indeed, in the main study. RQ2a in the pilot was able to establish that fillers like *er* and *erm*, often considered features of less successful speech, vary according to task type. They also act in learner speech as a characteristic of real-time

processing when learners are thinking of what to say, rather than just a sign of them thinking about how they are going to say it. *You* similarly exhibited a degree of fluctuation across the C1 exam; though used predominantly in statements of generalisations, it also enabled learners to create questions for clarification and interaction purposes. Though only two words have been explored in depth, they have shown the level of detail that words in the main study require, and the subtle differences that can be identified.

4.6.3 RQ2b: What were the 20 most frequent keywords at and their notable collocations, colligations and functions?

Tables 7-8 display the keywords across the entire C1 exam and in Parts A, B and C, respectively. As well as identifying words of particular importance, and establishing the ‘aboutness’ of the texts, it was anticipated that the keywords would reveal items requiring further exploration (see Bondi, 2010; Hunston, 2002; Scott & Tribble, 2006). The lists may also confirm some of the words earmarked for closer analysis following RQ2a’s findings. Indeed, this was the case.

Table 7: Top 20 keywords in pilot study - Entire exam

N	Key word	Freq.	%	Texts	RC. Freq.	RC. %	Keyness
1	ER	661	3.94	12	90,254	0.09	3,714.99
2	ERM	400	2.39	12	63,095	0.06	2,129.77
3	YEAH	280	1.67	12	83,012	0.08	1,149.25
4	THINK	285	1.70	12	88,700	0.09	1,143.61
5	I	630	3.76	12	732,523	0.74	1,055.51
6	LIKE	241	1.44	12	147,936	0.15	663.74
7	TOURISM	55	0.33	7	1,461		483.59
8	HOTEL	82	0.49	4	10,911	0.01	462.24
9	MAYBE	76	0.45	12	10,023	0.01	429.71
10	IT'S	172	1.03	12	126,792	0.13	417.28
11	DUBAI	31	0.18	2	141		376.50
12	COUNTRY	92	0.55	12	27,959	0.03	372.36
13	BECAUSE	144	0.86	12	100,659	0.10	362.62
14	SO	207	1.23	12	239,549	0.24	344.94
15	MM	90	0.54	9	34,736	0.03	323.77
16	UM	32	0.19	7	651		297.97
17	PEOPLE	131	0.78	12	116,196	0.12	275.69
18	IMPORTANT	81	0.48	12	38,721	0.04	259.23
19	REALLY	86	0.51	8	46,477	0.05	255.98
20	YOU	277	1.65	12	588,503	0.59	215.12

Table 8: Top 20 pilot study keywords in Parts A, B and C

N	Key word	Part A					Part B					Part C									
		Freq.	%	Texts	RC. Freq.	Keyness	Freq.	%	Texts	RC. Freq.	Keyness	Freq.	%	Texts	RC. Freq.	Keyness					
1	ER	182	5.66	11	90,254	0.09	1,156.31	ER	246	3.45	12	90,254	0.09	1,318.79	ER	233	3.63	12	90,254	0.09	1,271.91
2	ERM	88	2.74	12	63,095	0.06	492.92	ERM	165	2.32	12	63,095	0.06	869.45	ERM	147	2.29	11	63,095	0.06	771.19
3	I	168	5.23	12	732,523	0.74	376.42	YEAH	164	2.30	12	83,012	0.08	775.13	TOURISM	50	0.78	6	1,461		526.16
4	LIKE	69	2.15	12	147,936	0.15	241.23	THINK	139	1.95	12	88,700	0.09	594.74	THINK	111	1.73	12	88,700	0.09	449.09
5	MY	56	1.74	12	146,775	0.15	174.79	HOTEL	81	1.14	4	10,911	0.01	591.7	COUNTRY	66	1.03	10	27,959	0.03	347.13
6	RESTAURANT	20	0.62	8	3,410		168.28	I	258	3.62	12	732,523	0.74	416.73	I	204	3.18	12	732,523	0.74	286.72
7	CHINESE	20	0.62	7	4,153		160.46	DUBAI	23	0.32	1	141		306.15	YEAH	75	1.17	11	83,012	0.08	257.16
8	YEAH	41	1.28	12	83,012	0.08	147.41	LIKE	90	1.26	11	147,936	0.15	227.1	LIKE	82	1.28	11	147,936	0.15	208.45
9	GO	40	1.24	11	87,257	0.09	138.24	IMPORTANT	53	0.74	12	38,721	0.04	212.55	SO	96	1.49	12	239,549	0.24	190.4
10	FOOD	23	0.72	8	18,675	0.02	122.8	YOU	159	2.23	12	588,503	0.59	190.38	MAYBE	32	0.50	10	10,023	0.01	186.99
11	THINK	35	1.09	10	88,700	0.09	111.21	MAYBE	32	0.45	10	10,023	0.01	180.49	PEOPLE	69	1.07	11	116,196	0.12	183.73
12	UM	10	0.31	2	651		103.20	IT'S	67	0.94	12	126,792	0.13	152.39	MM	44	0.68	8	34,736	0.03	178.61
13	IT'S	38	1.18	12	126,792	0.13	101.82	BECAUSE	60	0.84	12	100,659	0.10	149.04	BECAUSE	63	0.98	11	100,659	0.10	173.64
14	EAT	15	0.47	7	7,280		95.16	UM	14	0.20	4	651		131.55	IT'S	67	1.04	12	126,792	0.13	164.55
15	RELAX	10	0.31	4	1,731		83.81	REALLY	41	0.58	7	46,477	0.05	130.71	THEY	104	1.62	11	376,289	0.38	143.98
16	LAUGHS	8	0.25	5	588		80.63	LOCATION	20	0.28	4	3,977		130.58	REALLY	36	0.56	6	46,477	0.05	113.02
17	FRIENDS	15	0.47	8	14,577	0.01	74.82	OKAY	24	0.34	10	12,098	0.01	113.18	DUBAI	8	0.12	1	141		92.03
18	SO	42	1.31	11	239,549	0.24	73.9	EMIRATES	9	0.13	1	142		103.61	CULTURE	18	0.28	7	8,481		90.83
19	TRAFFIC	11	0.34	3	6,468		65.63	MM	30	0.42	8	34,736	0.03	94.44	MY	49	0.76	11	146,775	0.15	82.18
20	MAYBE	12	0.37	7	10,023	0.01	63.37	VERY	50	0.70	11	119,611	0.12	93.76	LOT	24	0.37	9	27,912	0.03	79.93

RQ2a revealed that some of the 20 most frequent C1 words did not appear in the 20 most frequent BNC and CANCODE words: namely *think*, *like*, *so* and *because*. The keyword lists in Tables 7 and 8 confirm that in nearly every instance (except *because* in Part A, and *so* in Part B), these words were used significantly more frequently than in the BNC World reference corpus and thus warranted more attention. The words identified, clearly had a strong exam connection, so for this reason, *think* and *so* were selected for further analysis.

Think

Overall, *think* ranked 4th in the entire exam C1 keyword list. Part A comprised only 35 uses (1.09% of the section's corpus) in comparison with Part B, 139 uses (1.95%) and Part C, 111 uses (1.73%). Once again, a closer look at the data, in particular the occurrence of *I think*, suggested that successful students required Part A as a 'settling in' period. *I think* was found to be an extremely common structure in the C1 pilot data unsurprisingly due to the nature of the language observed. Using exam data to examine success meant that learners would be asked for opinions throughout the dialogue; it was, in fact, discovered that *I think* constituted 87% of all uses of *think*. Returning to the 'settling in' phase, Part A demonstrated that *I think* was preceded by *er* or *erm* on 12 (34%) of occasions. Such usage in Parts B and C, while displaying comparable frequencies, indicated that *er I think* and *erm I think* occupied lower proportions: Part B included 9 (6.5%) instances and Part C consisted of 16 (14.4%) instances. Successful students may consciously or unconsciously use such structures to therefore demonstrate that they are going to answer the question or to 'buy' valuable thinking time, especially in Part A.

Upon examining the use of *I think*, another distinction arose, this time in Part B of the exam. The previous figures relating to its very frequent usage suggested that throughout the exam, *I* was the only personal pronoun commonly used with the verb *think*. When using the clusters function in WordSmith Tools, which automatically computes phrases of three words in length (Scott, 2015), results showed that *you* was also rather numerous. RQ2a uncovered the more interactive nature of Part B in the various question forms implemented; RQ2b has confirmed this by showing that the cluster *do you think* arose 24 times, and *what do you* arose 16 times. Being successful at C1 rests not only on

the ability to express personal opinions, but also on the ability to seek opinions from others, sometimes in seemingly simplistic ways.

So

The use of *so* by C1 learners was widespread in their performances. Appearing 207 times, it constituted 1.23% of the total C1 pilot corpus and was 14th in the keyword list (see Table 7). In order to uncover its usage, once again Carter and McCarthy's (2006) *Cambridge Grammar of English* was consulted. Concordance lines for all instances of *so* were inspected in order to identify the functions evidenced.

Successful language at C1, with respect to the word *so*, demonstrated a degree of flexibility. Table 9 displays evidence for its use as an adverb of degree, a substitute, a subordinator for expressing result, purpose and consequence and as a discourse marker.

Table 9: Uses of *so* in the C1 pilot corpus

Categories	Description (Carter & McCarthy, 2006: 140-144 & 734)	Freq.	%	Combined Freq.	Combined %
Adverb of degree	Modifier of adjective or adverb meaning 'to this extent'	0	0.00	13	6.28
	Modifier of adjective or adverb meaning 'very' or 'really'	10	4.83		
	So + adjective/adverb + that	0	0.00		
	Premodifier of quantifiers e.g. so many...	3	1.45		
Substitute	So + modal and auxiliary verbs to mean 'also'	0	0.00	8	3.86
	Clausal substitute with 'assume', 'be afraid', 'believe', 'hope', 'imagine', 'reckon' and 'think'	2	0.97		
	Expression of uncertainty: e.g. 'I guess so', 'I suppose so'	4	1.93		
	Indicator of previous knowledge, e.g. 'mm, so I understand'	2	0.97		
Subordinating conjunction	To introduce results, consequences or purposes	98	47.34	98	47.34
Discourse marker	Marker to summarise another speaker	7	3.38	12	5.80
	For openings e.g. so, how's John getting on?	5	2.42		
Incorrect use of 'so'	N/A	6	2.90	6	2.90
And so on	N/A	3	1.45	3	1.45
Miscellaneous	N/A	67	32.37	67	32.37

Immediately, it can be seen that there was a marked use of *so* as a subordinating conjunction; demonstrated on 98 occasions, it occupied a proportion of 47%. In this sense, *so* was utilised not only to add extra detail such as positive and negative factors, but also

to occasionally express hypothetical scenarios. To be seen in RQ3, the ability to integrate subthemes and develop points is integral at C1 (CoE, 2001) and, as shown in the examples below, can be achieved rather simply by using *so*.

Extract 11: Pilot study, C1, Exam 9

<\$25F> *Erm I guess another problem in Cairo is the smoke. We've got lots of factories <=\$=> which are </\$=> really they are at the boundaries but they are so close to the tow= to the city itself so it it made that black cloud so they're trying to fix it but it's not yet fixed by putting filters on the um pipes and stuff and er try to reduce they carbon footprint.*

The second-most numerous function of *so* was its use as a discourse marker. Initially, using Carter and McCarthy's descriptions, such usage of *so* only stood at 5.80%. However, there were several utterances in which *so* acted as a discourse marker despite not conforming strictly to the existing descriptions. From the miscellaneous category, two additional uses of *so* as a discourse marker were identified. Firstly, Carter and McCarthy's (2006) descriptions state that *so* can help in summarising another speaker's utterances, as is demonstrated in Extract 11. Alternatively, it was found that students would summarise *their own* utterances: much like an essay would offer a conclusion, some C1 learners used *so* to 'round off' or reiterate their main points (see Extract 12). Secondly, there is initial evidence to suggest that *so* sometimes acted as a signal that a speaker had finished their turn. *So* or *so yeah* (see Extracts 12 and 13) when used in final position might signal to the other learners that the speaker would not add anything further, an observation requiring further investigation in the main study. Both of these uses could be interpreted as discourse markers for indicating the end of a turn. If taken as such, the proportion of *so* as a discourse marker actually rises to 38.17%. As will be seen in RQ3, the ability to take and pass turn is a requisite of C1 speech; the use of *so* to do this might therefore indicate that such an ability can be achieved more subtly than some interlocutors or examiners judging success might expect.

Extract 12: Pilot study, C1, Exam 1

<\$1F> *It's not the most important cos you waste money going like if you are going or coming to your country I would I waste money for tickets <\$G?> everything so + <\$2M> So you are not the type of person who travels a lot? <\$1F> + yeah.*

Extract 13: Pilot study, C1, Exam 1

<\$2M> *So my best foreign food is er Italian food which I enjoy the most is pasta most type of pasta so yeah.*

To sum up this research question, keyword analysis presented lexis which was typical of exam based data. The exploration of *think* and *so* revealed that C1 learners were successful in their speech as they could express their own opinions but also seek those of others. The conjunction *so*, often used as a subordinating conjunction, allowed learners not only to extend their responses and include sub-ideas, but it also functioned as a discourse marker to allow them to ‘round off’ their responses as per the CEFR can-do statements.

4.6.4 RQ2c: What were the most frequent 3- and 4-word chunks at C1 and their notable collocations and functions?

Table 10: 20 most frequent 3-word chunks in pilot study - Entire exam

Rank	Frequency	Range	Chunk	Spoken BNC frequency	Spoken BNC Chunk
1	38	12	I THINK IT	7015	I DON'T KNOW
2	37	10	IN MY COUNTRY	5728	A LOT OF
3	31	10	A LOT OF	3836	I DON'T THINK
4	30	11	I DON'T	3241	ONE OF THE
5	28	10	SO I THINK	2462	I MEAN I
6	27	10	THINK IT'S	2441	DO YOU WANT
7	25	4	OF THE HOTEL	2345	AND I THINK
8	24	11	I THINK THE	2217	BE ABLE TO
9	22	10	DO YOU THINK	2209	YOU WANT TO
10	21	10	IT'S VERY	2172	AT THE MOMENT
11	20	8	ER I THINK	2115	WHAT DO YOU
12	20	9	ERM I THINK	2112	GOING TO BE
13	18	5	I THINK ERM	2097	THE END OF
14	17	7	I THINK THAT	2071	DO YOU THINK
15	17	7	IT'S A	1990	I THINK IT'S
16	17	10	IT'S NOT	1973	IT WAS A
17	17	8	WHAT DO YOU	1952	YOU HAVE TO
18	16	7	I THINK ER	1934	A BIT OF
19	16	8	YEAH I THINK	1866	I THINK IT
20	15	9	DON'T HAVE	1780	I THINK THAT

Table 11: 20 most frequent 4-word chunks in pilot study - Entire exam

Frequency	Range	Chunk	Spoken BNC frequency	Spoken BNC chunk
27	10	I THINK IT'S	1163	THE END OF THE
14	8	WHAT DO YOU THINK	1103	AT THE END OF
11	4	LOCATION OF THE HOTEL	1031	THANK YOU VERY MUCH
10	3	THE LOCATION OF THE	868	I DON'T KNOW WHAT
9	6	I DON'T KNOW	667	HAVE A LOOK AT
8	4	DO YOU THINK ABOUT	628	I THINK IT WAS
7	3	IN MY COUNTRY IS	625	DO YOU WANT TO
7	3	IT'S IT'S	601	A LOT OF PEOPLE
6	5	A LOT OF PEOPLE	600	I DON'T KNOW WHETHER
6	6	I AGREE WITH YOU	589	IF YOU WANT TO
6	5	THINK IT'S VERY	572	TO BE ABLE TO
6	2	TOURISM IN MY COUNTRY	562	A BIT OF A
6	5	YEAH I AGREE WITH	550	KNOW WHAT I MEAN
5	1	A FOUR STAR HOTEL	545	YOU KNOW WHAT I
5	4	A LOT OF THINGS	528	I DON'T WANT TO
5	3	COMFORT OF THE HOTEL	507	WHAT DO YOU THINK
5	3	DON'T WANT TO	501	I WOULD LIKE TO
5	4	ERM I THINK IT	475	ARE YOU GOING TO
5	4	I THINK MOST OF	474	NO NO NO NO
5	3	I THINK THAT THE	471	I THOUGHT IT WAS

At first glance, the C1 chunks presented in Tables 10 and 11 exhibit several parallels with the Spoken BNC data. Clearly, the raw frequencies differ drastically but with only 12 exams collected, the pilot data has shown that chunks such as *I think it*, *a lot of*, *what do you think*, and *I think that* are used by successful C1 learners and NSs alike. The lists also display three- and four-word chunks which do not appear in the 20 most frequent Spoken BNC chunks; their usage may demonstrate unique features of the language examined which, if investigated more thoroughly, may reveal key insights. Although there is also evidence to suggest that learners employed topic related chunks, e.g. *location of the hotel* and *tourism in my country*, it is clear that their seemingly high rankings were a result of individual exams' topic choice which may have skewed results. For instance, *location of the hotel* is the third most common four-word chunk but it arose in only four exams. Discussion here, will specifically focus on the use of *don't* in 3- and 4- word chunks, specifically, *I don't*, *don't have*, and *don't know*. Though a range of chunks could have been chosen, these n-grams were thought to display subtle differences in usage at C1 which could be investigated in the main study.

Initially in RQ2a, *don't* did not clearly stand out as an item requiring further investigation. Whereas *n't* appeared 11th in the Spoken BNC list, suggesting that *don't* could be one of the structures linked to that listing, *don't* only appeared 49th in the UCLanESB most frequent words occurring a total of 73 occasions. However, in Tables 10 and 11 above, it forms part of four 3- and 4-word chunks. The fact that two of the chunks contain different verbs and one of the chunks omits a subject implied that its usage differed in the pilot study.

BNC chunk lists indicate that *I don't know* was the most frequent lexical chunk in NS data for 3-word chunks; this was also true of the CANCODE corpus (Carter & McCarthy, 2007). UCLanESB 4-word chunk data revealed that *I don't know* was used only on 9 occasions (12%) in 6 exams, a frequency much reduced in comparison. It was therefore necessary to uncover which other words collocated with it to see how successful C1 students utilised it. Analysis of *I don't* found that it collocated with 10 different verbs: *believe, go, have, know, like, think, trust, understand, want, watch*. These verbs immediately demonstrated that *don't* offered learners some flexibility. When used with verbs such as *go, like, and want*, learners seemed to be giving details of actions, hobbies or routines (see Extracts 14 and 15); when used with *believe, know, think, trust, and understand*, learners seemed to be indicating personal opinion, belief or stance. A simple word like *don't* can therefore combine easily with other verbs to compose not only new structures, but also potential new chunks for expressing routines and opinions that may appear in other learner speech in the main study.

Extract 14: Exam 1

<\$2M> <\$=> *I used to </\$=> I used to see the rating of it like **I don't** go through the wording itself I just see erm like if going to state it out of ten what do you think about it er so yeah so.*

Extract 15: Exam 7

<\$14M> *Well in my opinion I prefer the mm location of the hotel and the review of the hotel. Erm er **I don't** know others but generally every time so I just have my holiday and travel on site I just before that I check the website on the internet and the things I check is the location because I need to make sure that it's really close to er my destination or I **I don't** want to spend er half hour on road to er drive or walk to the hotel.*

The second chunk with *don't* identified in Table 10, *don't have*, pinpointed another distinction. Though its frequency was rather low, a key difference appeared not in the collocations following the chunk, but in the subjects preceding it. The emphasis on

expressing personal beliefs or describing personal habits was replaced with the use of *don't* with *have* as a lexical, and modal verb to express problems, obligation, and facts affecting not only the speaker, but wider society too. Whereas *I don't have* appeared only twice out of 15 occurrences of the chunk, 7, nearly half of all instances, contained *people don't have* or *they [people] don't have*:

Extract 16: Exam 3

<\$5F> *Yeah and also need to give them money because if you live in a poor countryside you will plant the tree or some corn and they **don't have** money to get education and they will never go out to the mountain.*

Extract 17: Exam 12

<\$23M> <\$O36> *Yes </\$O36> because fast food is popular in every country because it's convenient and people **don't have** to wait so people prefer to take fast food yeah.*

This change in focus provided initial evidence that chunks containing *don't* allowed C1 learners to discuss more complex subject areas, as is expected in the CEFR, with relative simplicity. As a grammatical rather than lexical item, *don't* is clearly not predisposed to any particular subject matter; the observation that it could facilitate discussion of more complex structures rests instead on the premise that learners are not focussing on themselves but on other groups of people and, rather often, the problems and challenges they face.

The final analysis performed used data from the 4-word chunk list. Whilst the drastically reduced frequency of *I don't know* in the C1 pilot corpus was indicated in discussion of 3-word chunks, it still appeared fifth in Table 11. Investigation here not only aimed to provide observations of how successful C1 learners used it, but instead, it aimed to identify possible analyses that could be conducted of B1, B2 and C1 speech in the main study. Since the subject and collocating verb were already contained within the chunk, closer inspection focused on the colligational patterns created and functions performed. Carter and McCarthy (2006) remark that verbs requiring a direct object, such as *know*, are followed either by a noun phrase, pronoun or clause. Using these three categories, subtle differences were uncovered. The most common colligation of *I don't know* found in its 9 uses involved its combination with *how*, *why*, or *if* clauses to express a lack of knowledge or uncertainty (see Extracts 18 and 19). Alternatively, only one instance contained a noun phrase. This finding may provide an interesting basis for analysis of B1, B2 and C1 speech. Though admittedly frequencies were low, any increase

or decrease in *don't know* with a noun phrase or clause could offer greater illustration of the abilities of successful students at the three levels. During this analysis, an alternative usage of the chunk *I don't know* also became apparent. Although one speaker did quite simply use it as a basic statement of uncertainty, other speakers used it as a hedging device between utterances to reduce their strength. In Extracts 18 and 19 it can be seen that the hedging device was used in the middle of utterances to show that the speaker was not entirely certain that they were correct or that the example they were presenting was fully relevant. Such a strategy can be essential in speech where individuals continuously assess the inclusion and construction of new and shared knowledge and the “common ground” between listener and speaker (Carter & McCarthy, 2006: 835). Performing an interpersonal role, chunks like *I don't know* can help students be successful in encounters where speaker relationships have to be maintained.

Extract 18: Exam 11

<\$21M> *Two points agreeing and disagreeing historical points I don't know I think it's er a matter of culture.*

Extract 19: Exam 12

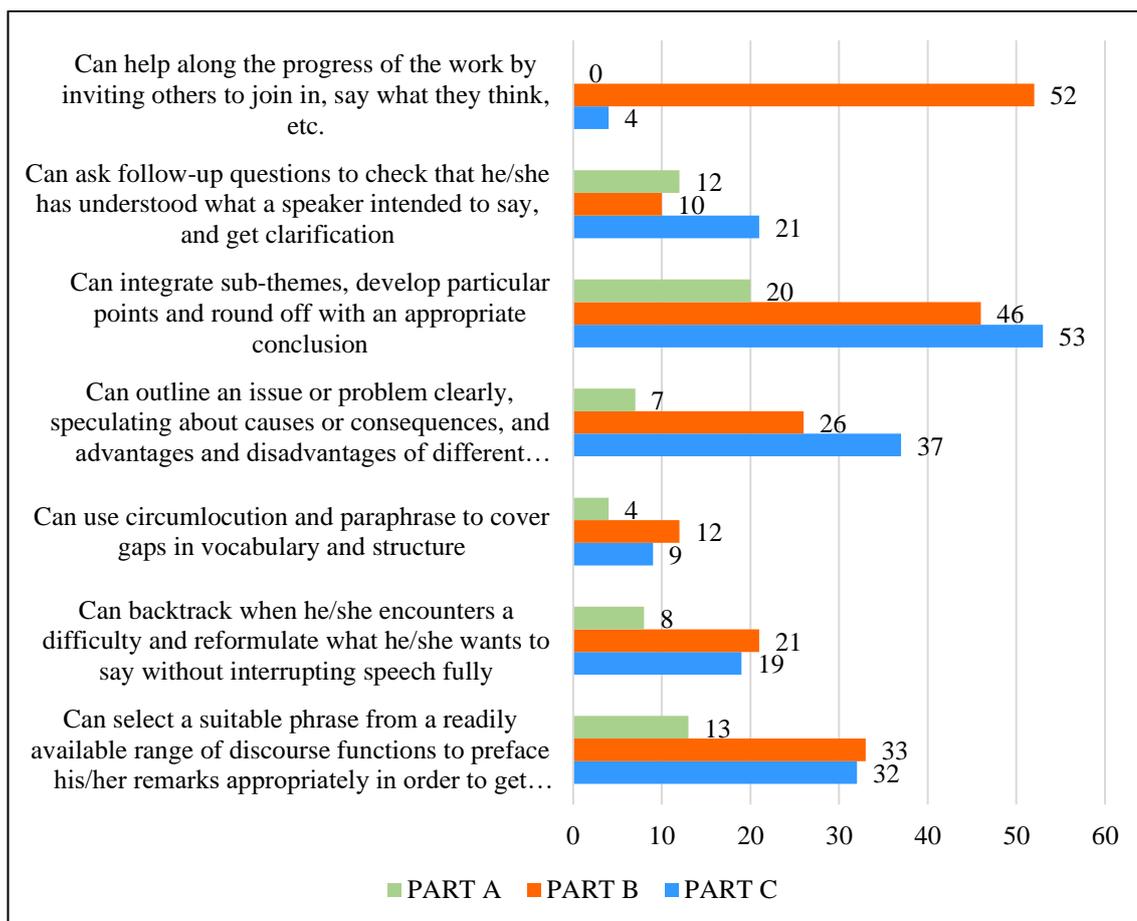
<\$24M> *Here at university I don't know they don't give you like much options about like good food or may= maybe we don't know all the options.*

In terms of success, this pilot data has demonstrated that C1 students use a range of 3- and 4- word chunks in spite of their low frequency. They also perform a range of functions. Closer examination of *don't*, for instance, has shown that lexical chunks not only satisfied content-related needs, but they also allowed successful C1 learners to fulfil more pragmatic duties which are important in face-to-face interaction. Chunks of a multifunctional nature should therefore be presented to learners so that the depth rather than breadth of lexical knowledge can be enhanced and so that knowledge of previously learnt lexis can be exploited fully. The perceived benefits of lexical chunks on memory demands, proceduralisation and accuracy during real-time speech may also help to satisfy CEFR calls for C1 students to evidence the integration of sub-themes and explanations “spontaneously” and “effortlessly” (CoE, 2001: 28) in comparison with learners at other levels.

4.6.5 RQ3: What Common European Framework of Reference for Languages (CEFR) indicators are present in terms of spoken interaction, spoken production and strategies at C1?

Assessing Figure 15's results holistically, successful C1 students at some stage, demonstrated all seven of the relevant can-do statements. Of course differing in degree, and according to exam section, the pilot study exams revealed that successful learners at C1 do conform to some of the CEFR's horizontal scales for spoken interaction, production and strategy usage. Learner speech was analysed according to CEFR statements deemed to be relevant to the speech produced in the UCLanESB exams. One selected, exams were read and examined closely for language satisfying such abilities (see Section 5.3.5 for more detail). Of a total 439 examples of can-do statement language, the most frequent ability evidenced involved the integration of sub-themes and concluding statements: success at C1 thus rests on the ability not just to offer responses or opinions, but instead to explain them and develop them. The least frequent statement exhibited related to the use of circumlocution and paraphrase: less than 6% of highlighted language was used in this respect. This finding could lead to two suppositions. First, C1 students, who may or may not have encountered such language, might not know or remember chunks that assist them when a key piece of vocabulary is unknown. The other alternative is that successful learners at C1 simply do not need to use circumlocution or paraphrase often: their current vocabulary may serve them adequately during discussions of familiar or 'complex' subjects as is expected by the CEFR. Both assumptions would prove difficult to corroborate in this present study but the latter suggestion might be a useful analysis point when B1 and B2 language is assessed in the main study.

Figure 15: Pilot study can-do statement occurrence in parts A, B and C at C1



As previously mentioned, can-do statement occurrence varied considerably according to exam section. Judgements of C1 learner language are therefore subject to the task completed by learners. For instance, if an interlocutor were to base their assessment of learners on their ability to invite others to contribute to discussion, Part A of the exam alone would clearly not offer a fair basis upon which to draw conclusions. For this reason, overall discussion of C1 success needed to account for the distribution of can-do statements across the pilot exams. Using the descriptions of Parts A, B and C offered by Jones, Waller and Golebiewska (2013: 32), the pilot data was thus able to elicit subtle changes in learner language throughout the exams. The following brief discussions therefore serve as a foundation for comparison in the main study. Although CEFR can-do statements evolve and even disappear across the levels, discovering the nature of the language used in C1 tests would hopefully allow comparisons of successful language use later in the main study.

Part A. “The interlocutor asks for mainly general personal information about the candidates (question-answer form). Candidates answer in turn. This stage lasts for approximately [two] minutes.”

Unsurprisingly, there were no attempts to progress discussion in Part A as only the interlocutor was responsible for asking questions. Similarly, little expansion of ideas was offered in this part possibly due to more basic question subject-matter: only 17% of attempts to integrate sub-themes and 10% of outlining a problem were found. Part A also elicited the lowest strategy usage in all three exam parts. Whilst notably shorter than Parts B and C, less than 1 in 5 CEFR C1 *strategies* occurred in the opening section. In fact, only 16% of circumlocution/paraphrase, 17% of backtracking and 17% of suitable phrase selection took place here. Successful language use is therefore not ‘littered’ with evidence of strategic C1 abilities as described by the CEFR can-do statements. Acting as the warm-up, a reduced need for strategies could be explained by the familiar, rather than complex, subject matter with which learners should be able to contend.

Part B. “Candidates engage in an interactive discussion based on two written statements. The interlocutor does not take part. This stage lasts for approximately [five] minutes.”

The students' responsibility to initiate and maintain conversation saw a dramatic increase in the first CEFR statement: 93% of attempts to progress discussion or invite other speakers into the conversation appeared here. Language of this type seemed not only to represent attempts to continue discussion: the use of questions to elicit responses from other learners sometimes appeared to be a delaying tactic for speakers wishing to have more time to think about their answer. In a sense, some questions appeared to ‘pass the buck’ when a speaker needed more time, did not have anything to say, or perhaps was unsure of their understanding of the written statements. With regards to *productive* can-do statements, students attempted to integrate sub-themes (the third statement in Figure 15) more often than outlining an issue (statement four). This could insinuate that students preferred to ‘add on’ ideas to fill the allotted time rather than outlining a problem which could demand more complex control of lexis and grammar. In complete contrast to Part A, Part B generated the highest strategy usage of all of the three exam stages. 43% of strategies took place here but circumlocution and paraphrase were surpassed by reformulations and suitable phrase selection. Successful C1 learners thus require the ability to attend to features which may reduce clarity for listeners.

Part C. “Candidates discuss questions related to the topic in Part B both together and with the examiner. This stage lasts for approximately five minutes.”

The use of prompts or questions to develop or invite discussion drastically fell again after Part B. The interlocutor’s involvement seemed to retake control of the conversation and could be taken as a sign that students relinquished responsibility for directing the conversation: compared to 52 instances in Part B, the first statement in Figure 15 elicited only four attempts in Part C. The use of questions instead assumed a different role. As already demonstrated in Section 4.6.2, students relied on questions for clarification purposes more often in Part C than in any other part of the test. As the ‘level-check’ section, the inclusion of more complex vocabulary and concepts by the interlocutor required successful learners to clarify or simplify some terms. The level-check seemed also to elicit extended responses from successful C1 learners since over half of attempts to outline a problem, advantage or disadvantage occurred in this section. Finally, strategy use, similarly to Part B was high with approximately 40% of strategies found in this stage of the exam. To be successful, learners mostly need to be able to reformulate utterances with little disruption to the communication and use suitable phrases to ‘buy time’, gain and maintain their turns.

Although summarised very briefly here, the findings for this research question do seem to suggest that the C1 exam did evidence some of the criteria suggested by the CEFR document. The different exam sections do influence which criteria are evidenced and more complex stages (parts B and C) do require students to utilise strategies in order to succeed. Not only would tuition of such strategies help future exam candidates, but further understanding of how they are implemented through language could help examiners to recognise their occurrence.

4.7 Chapter conclusion

The pilot data has shown that C1 students relied heavily on the first most frequent 2000 word families. In terms of individual words (tokens), only approximately 5% derived from other frequency bands. In relation to word families, approximately 20%, 1 in 20 words, also did not originate from the two most frequent bands. This provides good support for the claims made in Section 3.3.1. In order for students to be successful in the

C1 exam, learners should be taught the first 2000 words in English. Due to their flexibility and the manner in which they can be recycled, it be sensible to teach these words first before focussing on other vocabulary (as was suggested by Cobb (n.d.-a) and Nation (2001)). As classroom time is limited, students may benefit from being taught vocabulary learning strategies (Nation, 2001) so that they can progress further in their vocabulary development. Vocabulary from other bands constituted 1% or less of total tokens in the C1 pilot corpus so focussing on such lexis in the classroom may not be as beneficial. With regards to individual and chunked lexis, some of the pilot corpus' most frequent and key words and chunks have revealed subtle, yet significant differences in the way that successful learners use them to achieve a variety of functions. Importantly, being successful at C1 does not necessitate an unnatural or unrealistic use of overly complex structures or lexis. Some seemingly basic words, such as *you*, *think* and *so*, allowed learners to verbalise generalisations, express consequences, make clarification requests and elicit opinions to various degrees across the three exam sections; evidence of fillers, discourse markers and stalling phrases also demonstrated an ability to successfully manage speech while learners simultaneously participated in real-time conversation. The key to success may therefore lie not only in teaching students the common functions of individual words, but the *range* of functions they can facilitate. As mentioned briefly in Section 1.5 and discussed in Section 3.3.1 being successful at C1 may hinge on the capacity to exploit vocabulary that is already known to its full potential. Finally, C1 success has been shown to correspond to some of the CEFR's can-do statements. While the language used to realise them will be elaborated upon in Chapter 6 (main study findings), the pilot study has provided evidence of a variety of productive, interactive and strategic language use, and its occurrence, throughout the exams.

4.8 Changes to methodology following the pilot study

The pilot study enabled a preliminary exploration of successful language use at one of the three selected CEFR Common Reference Levels. Whereas some methodological decisions proved fruitful, others required improvement so that analytical techniques and findings could be made more robust for the main study. Greater detail of the amendments will be offered in Chapter 5, but justification for their inclusion will be outlined briefly below:

Table 12: Amendments to the main study

Affected research question(s)	Amendment	Rationale
RQ2a RQ2b RQ2c	Choice of corpus for comparison	The use of the Spoken BNC and CANCODE frequency, keyword and chunk lists proved useful. However, general consensus acknowledges that NS general corpora are not suitable for learner language comparisons. The BNC and CANCODE lists were still used in the main study but they were supplemented with lists generated using LINDSEI (Gilquin, De Cock & Granger, 2010).
RQ2a RQ2b RQ2c	Normalised frequencies	The pilot study relied on the position of words and chunks in lists. Due to the stark difference in corpora size, it was necessary to include normalised frequencies so as to make comparisons in the main study more insightful and meaningful. Normalised frequencies provide figures per 10,000 or million words so that corpora of different sizes can be compared; they ultimately reveal how often a word can be expected in a set amount of running text (McEnery & Wilson, 2012).
RQ2a RQ2b RQ2c	Concordance lines	Although concordance lines were used throughout the study, they were not included as evidence in the discussion. Where necessary, concordance lines for UCLanESB data, as well as Brigham Young University British National Corpus (n.d.) and LINDSEI, will be added to support findings.
RQ2a RQ2b RQ2c	Inclusion of statistics for: Dispersion Keyness Collocation Significance	The pilot study used raw frequencies and percentages to demonstrate the occurrence and proportions of word usages and functions. To add more depth, where necessary, the <i>dispersion</i> of words throughout the main study corpus would be given using Juilland's D to ascertain whether their appearance in lists were due to particular tests 'skewing' figures or whether they occurred across all tests. In terms of keyword data, the keyness figure computed using log-likelihood in <i>WordSmith Tools</i> was not commented upon or explained so this will also be justified in the main study. Similarly, when using concordance lines to investigate the usage of individual words or chunks, collocational discussions were strengthened by the inclusion of Mutual Information scores, scores representing the strength of collocations within the texts, and t-scores, representing the evidence existing for a particular collocation. Finally, where averages as to the frequency of words or chunks were calculated, it was necessary to strengthen comparisons to discover if an item was statistically more or less <i>significant</i> . This was achieved using t-tests where applicable.
RQ2b	Additional keyword analysis using BNC World and LINDSEI	Keywords give the 'aboutness' of texts. While some topic-related lexis was identified, not much was revealed about what learners discussed. In a study by Jones, Waller and Golebiewska (2013) into B2 learner speech, keywords were calculated using <i>Compleat Lexical Tutor</i> , thus revealing more about the topics or activities that learners mentioned. Although RQ2b worked well using BNC World in <i>WordSmith Tools</i> ,

		alternative analysis of keywords using LINDSEI, a more comparable corpus, was incorporated to discover what topics were discussed.
RQ3	Triangulation	The process of refining relevant CEFR can-do statements in the pilot study resulted in only 7 statements out of a possible 26 and was completed subjectively. It was necessary to gain opinions from other examiners familiar with the UCLanESB tests as to which can-do statements could have been evidenced during the exams. This would assist in resolving another problem encountered during the pilot. A reduction of 26 to 7 C1 statements seemed rather drastic. Gaining the views of other examiners would potentially increase or simply confirm the number of can-do statements to be used in qualitative analysis. Similarly, checking selections of language which satisfied the statements also needed confirming to once again reduce subjectivity in analysis. Judgements from examiners were sought once again.

5. METHODOLOGY

5.0 Introduction

The previous chapter presented the preliminary methodology adopted, the findings gathered and the changes identified during the pilot study process. The purpose of this chapter is to detail the finalised procedures for collecting and analysing the data that answered the RQs outlined in Section 1.6 of the introduction. It is organised as follows. First, demographics for the UCLanESB corpus will be presented before decisions underpinning its construction are justified. Consisting of information about the research tool for collecting learner speech, corpus design considerations and ethical issues, the first half of the chapter aims specifically to satisfy calls for explicit corpus documentation which permits others to evaluate the strength and validity of the findings drawn (see McEnery, Xiao & Tono, 2006; Sinclair, 2004; Tognini-Bonelli, 2001). The chapter then culminates with discussion of the analytical processes implemented both in examining the UCLanESB corpus and in associating the language within it to CEFR can-do statements for B1, B2 and C1 learner speech.

5.1 UCLanESB corpus demographics

Sections 5.1.1-5.1.6 display information about the UCLanESB corpus' composition. Whilst figures are given as to the sub-corpora comprising each proficiency level, Section 5.1.1 also gives information about the corpus as a whole. This is then followed by information about learner nationality, exam topic, exam duration and tokens per exam section at all three levels. Though information about the corpus itself is of course vital, this section incorporates characteristics of the learners involved so as to enhance generalisations of findings to wider populations, a feature which has been deemed superficial in other corpus descriptions (Gass & Selinker, 2001). Whilst consensus appears to increasingly dictate that learner corpora should be designed to enable the analysis of individual learner groups (McCarthy, 2016) this was not feasible due to the limited size of the UCLanESB corpus.

5.1.1 B1, B2, C1 sub-corpora and combined UCLanESB corpus information

Table 13: B1 sub-corpus

CEFR level examined	B1
Examination venue(s)	University of Central Lancashire and independent language school
Total word count including examiner	20,414 tokens
Total word count excluding examiner	14,527 tokens
Total number of texts (exams) used	15
Total number of exam candidates	30 (12 males (40%), 18 females (60%))
Average age	24 years
Average time spent learning English	6 years, 3 months
Average time in UK	9 months

Table 14: B2 sub-corpus

CEFR level examined	B2
Examination venue(s)	University of Central Lancashire
Total word count including examiner	23,585 tokens
Total word count excluding examiner	18,821 tokens
Total number of texts (exams) used	15
Total number of exam candidates	32 (19 males (59%),13 females (41%))
Average age	24 years
Average time spent learning English	7 years, 8 months
Average time in UK	15 months

Table 15: C1 sub-corpus

CEFR level examined	C1
Examination venue(s)	University of Central Lancashire
Total word count including examiner	26,579 tokens
Total word count excluding examiner	20,633 tokens
Total number of texts (exams) used	15
Total number of exam candidates	31 (16 males (52%),15 females (48%))
Average age	24 years
Average time spent learning English	10 years, 9 months
Average time in UK	21 months

Table 16: UCLanESB corpus composition

CEFR level examined	B1, B2 and C1
Examination venue(s)	University of Central Lancashire and independent language school
Total word count including examiner	70,578 tokens
Total word count excluding examiner	53,981 tokens
Total number of texts (exams) used	45
Total number of exam candidates	93 (47 males (51%), 46 females (49%))
Average age	24 years
Average time spent learning English	8 years, 3 months
Average time in UK	15 months

5.1.2 Candidate L1

Table 17: UCLanESB learner L1

L1	B1	%	B2	%	C1	%	Total	%
Chinese	15	50.00	20	62.50	15	48.39	50	53.76
Arabic	9	30.00	11	34.38	9	29.03	29	31.18
Korean	4	13.33	0	0.00	0	0.00	4	4.30
Not given	0	0.00	0	0.00	3	9.68	3	3.23
Kurdish	0	0.00	1	3.13	1	3.23	2	2.15
Spanish	0	0.00	0	0.00	2	6.45	2	2.15
Hausa	0	0.00	0	0.00	1	3.23	1	1.08
Italian	1	3.33	0	0.00	0	0.00	1	1.08
Turkish	1	3.33	0	0.00	0	0.00	1	1.08
Total	30	100.00	32	100.00	31	100.00	93	100.00

5.1.3 Candidate nationality

Table 18: UCLanESB learner nationality

Nationality	B1	%	B2	%	C1	%	TOTAL	%
Chinese	15	50.00	20	62.50	15	48.39	50	53.76
Saudi	5	16.67	3	9.38	2	6.45	10	10.75
Qatar	1	3.33	2	6.25	3	9.68	6	6.45
Nigerian	0	0.00	0	0.00	4	12.90	4	4.30
Republic of Korea	4	13.33	0	0.00	0	0.00	4	4.30
Iraqi	0	0.00	2	6.25	1	3.23	3	3.23
UAE	0	0.00	2	6.25	1	3.23	3	3.23
Unanswered	1	3.33	2	6.25	0	0.00	3	3.23
Columbian	0	0.00	0	0.00	2	6.45	2	2.15
Egyptian	0	0.00	0	0.00	2	6.45	2	2.15
Libyan	2	6.67	0	0.00	0	0.00	2	2.15
Omani	0	0.00	1	3.13	1	3.23	2	2.15
Italian	1	3.33	0	0.00	0	0.00	1	1.08
Turkish	1	3.33	0	0.00	0	0.00	1	1.08
Total	30	100.00	32	100.00	31	100.00	93	100.00

5.1.4 Exam topics per level

Table 19: Topics discussed at B1, B2 and C1

B1	B2	C1
Cinema (5)	Cultures and traditions (2)	Travel and tourism (7)
Games and sports (2)	Outdoor hobbies (2)	Cities (1)
Learning (2)	Advertising (1)	Community (1)
Friends (1)	Education technology (1)	Environment (1)
Homes (1)	Future (1)	History (1)
Memories (1)	Homes (1)	Language learning (1)
Personality (1)	Jobs (1)	Staying healthy (1)
Travel and tourism (1)	Lifestyles (1)	The world around us (1)
Work (1)	Modern technology (1)	Transport (1)
	Music (1)	
	Success and luck (1)	
	Weather (1)	
	Work and training (1)	

5.1.5 Exam duration per level

Table 20: Exam length at B1, B2 and C1

Level	Stated exam duration (three candidates)	Actual average duration
B1	10 minutes (13 minutes)	13 minutes 12 seconds
B2	12 minutes (17 minutes 30 seconds)	12 minutes 27 seconds
C1	14 minutes (19 minutes)	15 minutes

5.1.6 Tokens per exam section

Table 21: Tokens in Parts A, B and C at B1-C1

Level	Part A	Part B	Part C	Part A average	Part B average	Part C average
B1	2796	5953	5813	186	397	388
B2	3924	8012	6902	262	534	460
C1	4032	8563	8065	269	571	538

*Combined exam section totals may not correspond to overall totals in Tables 13-15

5.2 Corpus construction rationale

Planning the construction of a corpus is a complex task which requires very careful consideration if it is to provide a sound foundation for the investigations to be made of it and successfully describe actual language use (Biber et al., 1998; Kennedy, 1998; Meyer, 2002). To be truly useful and “valid” (Stubbs, 2004: 107), preparation must

be undertaken to facilitate the final corpus' ability to reflect the research purpose, the language under investigation and the resources available to the researcher (Sinclair, 2004). However, with corpus linguistics deemed "inexact" (Sinclair, 2004: 81) and "problematic" (Biber et al., 1998: 246) methodological perfection is an unrealistic goal. "[C]ompilers must make the best corpus they can in the circumstances" (Sinclair, 2004: 81) since no single agreed approach exists when corpus techniques are employed (Baker, 2010; McEnery & Hardie, 2012). With some corpus features seen as 'ideals' rather than feasible targets, attention turns here to the preliminary considerations made prior to the corpus' construction so that its final composition portrayed in Section 5.2 can be justified. Using Sinclair's (2004) principles as a basis, decisions regarding representativeness, balance, size and sampling, amongst other factors, will be explored. In a slight amendment to Sinclair's principles, this section will also incorporate discussion of corpus content, i.e. the research tool selected for collecting data, recording and transcription, and ethical standards. Overall, it intends to present the deliberations carried out and the rationale for the final decisions made.

5.2.1 Research tool for collecting learner language

The pilot study chapter detailed the composition of a preliminary C1 corpus based on learner language from UCLanESB speaking exams. What remains to be discussed, however, were the reasons for using test data as a foundation for a study of success. Evidently, language exams are intrinsically linked with success: candidates taking them wish to achieve a particular mark and avoid failure. Whilst test research has seen a relatively recent introduction of corpus linguistic tools (Barker, 2010; Callies & Götz, 2015; Hunston, 2002), it must be emphasised that this study bore no resemblance to 'traditional' testing research into factors such as reliability, validity, authenticity, and interactiveness (see Bachman, 1990; Bachman & Palmer, 1996; Callies & Götz, 2015; McNamara, 2004 for common test research aims). The use of tests in this study therefore functioned as a tool for collecting successful spoken language, rather than a tool for determining the success of the tests themselves. It is for this reason that the term *learner* corpus, rather than *test* corpus, has been used in this thesis. Since the decision to use speaking tests as a basis for this learner corpus inevitably influenced the nature of the language collected, it is necessary to give a summary of their design and the rationale for their inclusion.

The exams at the heart of this study, the UCLanESB exams, were developed in 2006 and extended in 2011 (UCLanEB, n.d.-a). They currently assess language proficiency at the B1, B2, C1 and C2 levels according to the skills of listening, reading, use of English, writing and speaking. Although used within an academic setting, and unlike EAP or ESP tailored examinations, it is important to note that the UCLanESB exams adopt a similar format to the Cambridge ESOL examinations which were principally designed to satisfy educational demands but were “not intentionally linked to a specific domain of language use” (Taylor, 2011: 18). This “overall approach” (Spolsky, 1968: 91) means that no elaboration is offered as to the precise elements ‘mastered’ by the learner, but instead, it satisfies the requirement of aiding judgements of learners’ general, future language abilities (Bachman & Palmer, 1996; Barker, 2010; McNamara, 2004; Taylor, 2011; Underhill, 1987). Intending, therefore, to evaluate or give an overview of a candidate’s potential, overall proficiency, the UCLanESB exams were developed with reference to the CEFR, and are labelled according to the levels tested: B1, B2, C1 and C2 (UCLanEB, n.d.-a).

The B1, B2 and C1 speaking exams, a component of the UCLanESB exam set, follow an interactive structure. Involving the integration of skills in face-to-face communication, interactive tests aim to ascertain a learner’s knowledge of lexical, grammatical and phonological structure (see Douglas, 1997 cited in Hughes, 2011; Weir, 2005a). The UCLanESB speaking tests therefore adopt this style to assess learners’ communicative abilities in more natural situations via the use of the Oral Proficiency Interview [OPI]. Though OPIs typically consist of four stages, the B1, B2 and C1 exams consist of three: a warm-up, a level-check and a probe stage. The warm-up stage does little more than settle the candidates into the exam: it can give the assessor a preliminary sense of their level but it typically puts candidates ‘at ease’ and allows them to familiarise themselves with the interlocutor and exam structure. The *level-check* phase aims to establish the learners’ “highest sustainable level of speaking” (Johnson & Tyler, 1998: 29). Via pre-planned tasks or questions, language is elicited which may or may not satisfy “linguistic[ally] targeted criteria” (Brown, 2004: 168). In a sense, the level-check stage can help to confirm the ‘floor’ of learner performance. To end the OPI, UCLanESB opts for the *probe* section and omits the *wind-down* period. Once the floor of a candidate’s level has been surmised, the ‘ceiling’ is found using a range of questions, posed by the interlocutor. These endeavour to test the limits of a candidate’s language ability (Brown,

2004; Johnson & Tyler, 1998). For more detail about exam structure at each level, see Appendix 1.

Following completion of the speaking exams, pre-approved examiners, who must be native speakers or speakers equivalent to C2 level in the CEFR (UCLanEB, n.d.-a), are responsible for assessing students according to criteria on marking scales at each level. Individual marks in the areas of grammar, vocabulary, pronunciation, discourse management and interactive ability are given to each candidate and then amalgamated into one global score; although fluency is not explicitly assessed, hesitations are alluded to in the interactive ability column (see Appendix 2). A global score of 2.5 equates to a borderline pass, whereas a score of 3 equates to a firm pass. Once exam sessions are completed, the Chief Examiner then second-marks a selected of exams from each assessor to ensure that new and experienced assessors are correctly grading candidates. Here, there is the potential for problems relating to the comparative fallacy to arise since, learner success is ultimately judged by NSs once again. However, personal experience has revealed that non-native raters do not necessarily evaluate learners less strictly; sometimes, learner expectations can be higher as they are based on their own accomplishments. Hence, controlling learner assessments would pose challenges with or without a native speaker.

The use of UCLanESB speaking exams to collect learner speech clearly facilitated collection of interactions. That being said, speaking exams following the OPI design face considerable criticism. As a direct test in which an interlocutor is present, resulting speech is said to be strongly comparable to spontaneous, real-life, face-to-face communication (Fulcher, 2003; Hughes, 2011; Weir, 2005a). However, concern exists as to whether such speaking tasks truly and adequately reflect the natural, realistic conventions of conversation such as turn-taking, topic spontaneity, topic negotiation, or listener interest (Chalhoub-Deville, 2003; Fulcher, 2003; Johnson & Tyler, 1998; Young & He, 1998). As has been argued, OPIs are instead deemed to represent their own genre of communication (Chalhoub-Deville, 2003; Fulcher, 2003; Johnson & Tyler, 1998) and though “relatively realistic...they can never be *real*” (McNamara, 2000: 8). Similarly, some question i) whether the social impacts of candidates and interlocutors negatively influence the co-construction of the verbal encounter, and ii) whether the nature of exams, providing “deliberate samples...of an individual’s language knowledge or language behaviour” (McNamara, 2004: 767) are capable of reflecting successful learner speech in general or instead, are only capable of simply reflecting an ‘ad-hoc’ samples of speech.

Despite these views, it was felt that the UCLanESB speaking exams were still a suitable research tool for collecting learner speech. Consultation with the CEFR's descriptions of communicative, productive and interactive activities (see CoE, 2001: 57, 58, 73) confirmed this view by demonstrating that UCLanESB's OPI design did, in fact, satisfy many of the activities' criteria. Not only are candidates expected to act both as producers and receivers in a series of turns, but they are also required to use various communication strategies that may provide examiners with a wider impression of their spoken abilities (see Saville & Hargreaves, 1999; Taylor, 2011; Weir, 2005a). Furthermore, exams were used only as a tool for eliciting speech. When priority is placed on the interaction, rather than the adequacy of the testing model, importance shifts towards "the people and what passes between them" (Underhill, 1987: 3). As this was not a study into testing, and since the collection of data was of more importance than the potentially negative characteristics of exam design, its flaws as an instrument for testing language were of less significance. Finally, the UCLanESB speaking exams presented a uniform, accessible source of learner language which had been pre-assessed. Collecting samples of genuine learner-to-learner conversation would have been an ethically and logistically problematic: the content could have been private or personal, ensuring learner level prior to conversation would have been difficult, defining which conversations were successful would have been challenging, and comparison of differing conversations would have been very complex. The OPIs, regardless of their imperfections, thus offered a homogeneous, comparable structure – a requirement of solid corpus design (see Sinclair's (2004) tenth principle) – in which learner level could be ensured, success could be defined, informed consent could be obtained and in which interlocutor speech was (for the most part) controlled. As a "systematic gathering of language related behavior" (Chapelle & Plakans, 2013: 241 cited by Callies & Götz, 2015: 2), the exams meant that language could be collected, but more importantly, the scores assigned to them acted as a validatory tool.

To recap this section on the design, limitations and strengths of UCLanESB speaking exams in this study of success, it is clear that the capture of perfectly natural conversation at B1, B2 and C1, of course, would have been optimal. However, due to the difficulties in obtaining such data, the exams were chosen as they satisfied the criteria specified by the CEFR for communicative, interactive and productive spoken communication. Not only would the rigorous UCLanESB exam procedures ensure consistency in the collection and marking of usable exams, but the varied tasks and

interaction patterns would allow a range of speaking abilities to be evidenced. Their limitations can be counteracted using test and communication research, but from a corpus linguistic stance, it is also accepted that learner corpora will inevitably involve language with “some degree of ‘unnaturalness’” (Gilquin, De Cock & Granger, 2010: 5). Sections 5.2.2-5.2.9 of the methodology chapter will now outline the decisions taken in ensuring that the UCLanESB corpus design maximised the data’s relation to successful language.

5.2.2 Corpus type

Type, or “orientation” (Sinclair, 2004: 3), has been chosen as the second point for discussion as it is integral to subsequent corpus design considerations such as size and content. General corpora aim “to be representative of a particular language” (Baker, 2010: 99) via the inclusion of multiple text types to achieve a broader coverage (Adolphs, 2006; Aston, 1997; Hunston, 2002; Kennedy 1998). They can consist of hundreds of millions of words: the British National Corpus containing 100 million words and the World English Corpus containing 220 million words being cases in point. Monitor corpora also tend to be very large as they are “open-ended” (McEnery & Wilson, 2001: 30) and are continuously extended and updated, for example the Corpus of Contemporary American English (COCA) containing 450 million words and the Collins Corpus (COBUILD) containing 4.5 billion words. Specialised corpora, on the other hand, tend to be much smaller (Baker, 2010; Hunston, 2002). Although some specialised corpora, such as the 5 million word Cambridge and Nottingham Corpus of Discourse in English (CANCODE), do extend into the multi-million word range (see also Koester, 2010), most are smaller due to restrictions in text type and due to their carefully tailored construction with particular research questions in mind (Adolphs, 2006; Baker, 2010; Hunston, 2002). With specialised collections being as small as 20,000 to 200,000 words (Aston, 1997) or up to 250,000 words (Flowerdew, 2004 cited by Koester, 2010), it is clear that corpus *type* would have a stark effect on this study’s corpus composition.

It was evident that the corpus to be constructed was of a specialised nature. Focussing specifically on language from UCLanESB speaking exams, the corpus would be unable to be representative of an entire language variety but it would help to support conclusions about the language that could be evidenced by B1, B2 and C1 learners. The corpus would also satisfy other corpus type definitions. Despite containing some scripted language on the part of the exam interlocutor, the learners’ language was spoken and

assumedly spontaneous. The corpus, therefore, was an example of a spoken corpus: “a unique resource for the exploration of naturally occurring discourse” (Adolphs & Knight, 2010: 1) which, again, tend to be reduced in size due to the challenges and practicalities of obtaining and transcribing data (Adolphs & Knight, 2010; Baker, 2010; Leech, 1995; McEnery & Hardie, 2012; O’Keeffe, McCarthy & Carter, 2007; Thompson, 2004). It also satisfied definitions of a learner corpus. Learner corpora, as the name suggests, contain language, either spoken or written, produced by learners of a language (Baker, 2010; Hunston, 2002; Meyer, 2002). Aiming to supply enhanced descriptions of learner language” (Granger, 2002), they range in purpose but are often a source of comparison (Hunston, 2002) resulting in increased knowledge of the errors which may occur in learners from particular L1 groups or of errors which arise following native speaker comparison (Tono, 2003). Despite belonging to this corpus type, it was decided that the intended corpus was to have a different aim. Rather than identifying the errors made by learners, it would adopt a more positive stance. It would still correspond to definitions such as that provided by Hunston (2002: 15) as it would “identify in what respects learners differ from each other” at B1, B2, and C1, but it would do so while acknowledging the success the language afforded test candidates, rather than the errors contained within it. To summarise, the learner corpus to be compiled was specialised in that it would contain samples of spoken learner language at B1, B2 and C1.

5.2.3 Corpus compiler

In terms of the people compiling the corpus, Sinclair (2004) favours an individual who has expert knowledge of the communicative function of the language to be studied. This not only assists in the selection of suitable texts, but it ensures that corpus contents are selected for their function rather than solely for the language they contain. The corpus compiler in this study had acquired several years’ experience working with the written and spoken UCLanESB exams before embarking on this PhD. Experience of acting in roles as interlocutor, assessor and standardiser across a range of levels gave a vital level of understanding not only of the individuals involved, but also of the structure of the exam and the criteria required in achieving various grades. Standardisation and knowledge of the marking criteria (see Appendix 2) also provided additional insight into how global scores were representative of varying degrees of success in respect to grammatical accuracy, lexical range, interaction, discourse management and pronunciation.

5.2.4 Corpus representativeness

Sinclair (2004: 2) states that individuals should “strive” to maximise corpus representativeness, employing intuition and language awareness to make the corpus as demonstrative as possible of the language focus. In order to satisfy the study’s purpose and to ensure data is well-matched to research aims, it is essential that corpus compilers fully understand “what it [the corpus] is meant to represent” (Biber et al., 1998: 246). To achieve this, external criteria should be established prior to data collection to facilitate text selection and to reduce error. Sinclair (2004: 4) thus advises that the following considerations are made so that external criteria can be determined:

- a) the mode of the text; whether the language originates in speech or writing, or perhaps nowadays in electronic mode
- b) the type of text; for example if written, whether a book, a journal, a notice or a letter
- c) the domain of the text; for example whether academic or popular
- d) the language or languages or language varieties of the corpus;
- e) the location of the texts; for example (the English of) UK or Australia
- f) the date of the texts.

Applying this to the intended corpus was rather straightforward (see Table 22). However, it still left questions to be answered since other specific criteria templates, such as Tono’s (2003: 800) design considerations for building learner corpora, identified other potentially influential criteria. Table 23 demonstrates how Tono’s considerations related to the intended corpus. Its learner-related criteria, in particular, highlighted how individual factors such as age, background, L1 and language proficiency were potentially significant. It was also vital that factors such as learner background and proficiency were considered prior to data collection as they could have had implications for the corpus’ ability to be generalisable (representative) to the population of B1, B2 and C1 learners (Biber, Conrad & Reppen, 1998; Gass & Selinker, 2001; Gilquin, De Cock & Granger, 2010; Linquist, 2009; McEnery & Wilson, 2001; Meyer, 2002; Sinclair, 2005; Tognini-Bonelli, 2001).

Table 22: Applying Sinclair's (2004) criteria to UCLanESB corpus

Sinclair's criteria	Application to UCLanESB
a) the mode of the text	Spoken
b) the type of text	Language test
c) the domain of the text	Academic but general. Tests held in academic setting but test elicited general English
d) the language or languages or language varieties of the corpus	Learner English at B1, B2 and C1
e) the location of the texts	UK higher education institutions
f) the date of the texts	2013-2016

Table 23: Applying Tono's (2003) learner corpus criteria to UCLanESB corpus

Tono's considerations			Application to UCLanESB corpus		
Language-related	Task-related	Learner-related	Language-related	Task-related	Learner-related
Mode [written /spoken]	Data collection [cross-sectional/ longitudinal]	Internal-cognitive [age/cognitive style]	Mode: Spoken	Data collection: Cross-sectional	Internal-cognitive: Over 18
Genre [letter/diary/ fiction/essay]	Elicitation [spontaneous/ prepared]	Internal-affective [motivation/ attitude]	Genre: Language test	Elicitation: Spontaneous	Internal-affective: Unknown
Style [narration/ argumentation]	Use of references [dictionary/ source text]	L1 background	Style: Question and answer/ discussion	Use of references: Written question for Part B only	L1 background: Varying
Topic [general/ leisure/etc]	Time limitation [fixed/free/ homework]	L2 environment [ESL/EFL]/ [level of school]	Topic: Varying/ general	Time limitation: Fixed; 10- 15minutes depending on level	L2 environment: EFL, foundation level
		L2 proficiency [standard test score]			L2 proficiency: B1, B2, C1

The decisions made regarding learner background, specifically learner L1 and age, will be explained first. The UCLanESB speaking exams are taken by a range of students within the UK, Greece, China and Spain. Despite large amounts of exams being conducted in Greece, the decision was taken that the corpus should be representative of B1, B2 and C1 speaking exam candidates at the University of Central Lancashire's Preston campus. This meant that i) L1 would not be a controllable factor as the Preston campus has a range of student nationalities (see Appendix 5), and ii) data from the Greek exam intakes would not be used. Capturing a range of student nationalities, and therefore

L1s, was thought to increase generalisability. Although Greek candidates do comprise a vast portion of total exam candidates, including only Greek data may have made the corpus more specialised or inadvertently skewed; instead of being a learner corpus for general B1, B2 and C1 learner language use, it may have been representative only of Greek learners. Rationale for this decision was also attributed to ethical issues, data reliability and the practicalities of collecting data. Firstly, as highlighted in the pilot study (see Section 4.1), age is a factor which varies considerably amongst Greek exam candidates. Not only would this have complicated the process of obtaining informed consent for candidates under the age of 18 (see British Association for Applied Linguistics, 2006), but it would also have affected the reliability of conclusions drawn from the data. As was seen in Section 4.1, a learner's age can influence their linguistic development. More importantly, controlling age "does not necessarily ensure that the subjects selected are comparable in terms of language proficiency" (Tono, 2003: 801): a child learner may not have developed full competence in their L1 nor in their knowledge of the world which could impact on their ability to discuss the UCLanESB exam topics in the L2 and the corpus' ability to generalise successful language use across all learners. Also, the practicalities of research similarly needed contemplation since obtaining accurate and audible recordings from a range of institutions may also have resulted in the need for additional time to conduct the study and could have led to further complications had recordings been of poor quality. In summary, the Preston campus allowed for age to be regulated since enrolled students had to be at least 18 years of age but nationality was not a controlled factor. However, although learner L1 was not used as an external criterion for selecting usable exams, it is important to acknowledge that it can still be a key variable in analyses of successful learner speech. As was noted by Norton (2005) in a small-scale study into the effects of learner L1 and gender on paired speaking test performance, linguistic performance is intrinsically linked to learners' sociocultural and pragmatic competences which may be both elevated or hindered by monolingual or multilingual pairings. Capturing information about learner L1 and displaying it in the UCLanESB metadata (see Section 5.1.2) was therefore essential: in spite of it not affecting the data collection process, it would undoubtedly be an influential factor in examining learner success.

The second focus, that of language proficiency, could have had strong implications for the corpus' ability to exemplify successful language use at the three chosen levels both in relation to ensuring students were of the specified level and that

their spoken exam data reflected *successful* language use only. The B1, B2 and C1 levels were the sole focus of this study; no other exam levels were incorporated into the corpus. To ensure students were of the correct levels, chosen exam intakes followed courses in which the students had completed a placement exam on entry. Despite not being a failsafe guarantee of language level, proficiency tests did provide some evidence that students were not taking exams above or below their level. Secondly, and more importantly for documenting language success, careful consideration had to be given to the exam grades allocated to candidates. As was discussed in 5.2.1, candidates in UCLanESB speaking exams are given a grade from 0-5 on a range of aspects (see Appendix 2). A global mark of 2.5 represents a borderline pass. The marking scheme had strong implications for which exams could be included in the study since candidates with a global mark of 5 could have possessed an ability beyond that of B1, B2 or C1 level, whereas candidates obtaining a borderline global mark of 2.5 may not have exhibited a solid performance. Furthermore, decisions had to be made as to whether a score of 2.5 for each language aspect needed to be obtained or whether a global mark would be sufficient. It was decided that all candidates whose data was to be incorporated had to achieve a global mark of 3.5 or 4 to ensure a solid performance and to minimise the possibility of students receiving lower marks in the different spoken sub-categories; nevertheless, regardless of individual criteria, the overall performance had still been deemed successful. It also meant that each candidate in the exam dyad or triad had to obtain a pass of this grade. Despite reducing usable data heavily, an outcome that was forewarned by Tono (2003) and a consequence identified in the pilot study, such an approach was thought to increase validity.

To further ensure exams were of the correct grade, procedures were followed to check inter-rater reliability. Inter-rater reliability relates to independent judges assigning grades to the same performance; the raters do not confer or collaborate but the item under inspection remains the same (Bachman, 2004). Cohen's Kappa can be used to perform such a calculation. A selection comprising over 20% of all 45 exams collected were therefore sent for second-marking to see if they could be accepted as examples of successful learner language. However, exams for the study had already been pre-selected according to whether a grade of 3.5 or 4 had been awarded. Since all selected exams from the original raters showed a constant score, computing Cohen's Kappa against the second-marker's scores did not work; a constant was created resulting in a score of 0 despite there evidently being agreement. Although not ideal, the decision was taken to calculate inter-rater agreement as a percentage. In terms of students being graded at either 3.5 or 4, there

was an agreement rate of 50%. This immediately seems low, but additional rationale explained why the exams were still used. Firstly, all exams sent for review came back with a score of 2.5 or more. As indicated in Section 5.2.1, all learners were judged as displaying successful language as they all passed the exam. Secondly, UCLanESB has strict procedures in place for the standardisation and monitoring of examiners' rating. Individuals would have been working from the same agreed marking standards but with second-markers often working from audio recordings, some of the context of the speech would have been lost. Finally, the second-marker acted, in fact, as a third marker. UCLanEB guidelines (n.d.-b) state that an assessor completes the marking grid for each student but that the interlocutor must also provide a global score. The argument presented here is that two people, working independently of each other had already agreed on a score. The fact a third person assessed all speaking exams as successful provided further evidence that corpus representativeness for spoken language was assured.

To sum up this section on representativeness, here are the final external criteria used for the B1, B2 and C1 UCLanESB corpus:

Table 24: Final external criteria for UCLanESB corpus

Language-related	Task-related	Learner-related
<i>Mode:</i> Spoken	<i>Data collection:</i> Cross-sectional. Exam intakes from 2013-2016	<i>Internal-cognitive:</i> Aged over 18
<i>Genre:</i> Language test	<i>Elicitation:</i> Spontaneous	<i>L1 background:</i> Varying (not controlled)
<i>Style:</i> Question and answer/ discussion	<i>Use of references:</i> Written question for Part B only	<i>L2 environment:</i> UCLan Preston Campus: EFL, foundation level
<i>Topic:</i> Varying/general	<i>Time limitation:</i> Fixed; 10-15minutes depending on test level	<i>L2 proficiency:</i> B1, B2, C1 Global exam score of 3.5 or 4

5.2.5 Corpus balance

After ensuring that the design criteria of the corpus would allow results to be as generalisable as possible to the population of UCLanESB speaking exam candidates, the concept of 'balance' needed to be addressed. Balance refers to the weightings different sections carry in *general corpora* (Kennedy, 1998); it can also account for the proportions

of varying text types and topics incorporated. As well as being difficult to define in concrete terms (Hunston, 2002; Kennedy, 1998; McEnery & Hardie, 2012; Sinclair, 2004), balance has been declared “largely heuristic” (McEnery & Hardie, 2012: 10), a feature only discovered and determined by the compiler during the design stage process. It is significant as even for more specialised corpora, it can have considerable influence on design (Kennedy, 1998) and on the interpretation of data (Hunston, 2002). Any texts selected for inclusion therefore needed to be balanced carefully to ensure that they did not “skew the corpus as a whole” (Baker, 2010: 96). For this study, balance had to be contemplated in order to avoid disproportions in three aspects: cross-level comparisons, the use of partial or complete texts and topic choice.

Firstly, for general corpus compilation, Hunston (2002: 29) advises listing possible text types and including “roughly equal amounts of data” from these types. For the UCLanESB corpus, text type was uniform; only speaking exams would be used. The use of texts from different *levels* thus became more relevant. Admittedly, it was easy to make sure that each level consisted of an equal number of texts (exams). However, with corpus size attributable to either the number of texts per type, the number of words per sample *or* the total number of words (Biber et al., 1998), following such a technique may have introduced skew into the corpus which would have been difficult to control: keeping the number of B1, B2 and C1 texts equal would be simple, but keeping their relative and unpredictable wordcounts comparable would have been difficult to regulate. This brings the discussion to the second aspect: that of using complete or partial text samples to maintain balance. In this respect, the UCLanESB corpus was subject to contradictory advice. To reduce skew in general corpora, text fragments or partial text samples are often integrated (Baker, 2010; Meyer, 2002). Although perhaps not fully applicable to small, specialised corpora, this practice helps regulate size across texts. Despite this advantage, it is not a fully endorsed technique. Adolphs (2006), Meyer (2002) and Sinclair (2004) instead indicate that it is advisable to include texts in their entirety to maintain context. Using intact texts inevitably leads to varying text sizes (Sinclair, 2004) but also, it contradicts Hunston’s previous advice and that of Biber et al. (1998: 249) who remark “If too few texts are included, a single text can have an undue influence on the result of an analysis.” It seemed that including entire speaking exams may have increased the potential for individual exams, larger or smaller in size, to skew analysis given the specialised, reduced size of the UCLanESB corpus. Finally, opposing views about subject matter will be examined. Topic is of particular relevance here since UCLanESB speaking

exams cover a range of general, non-specialised themes (see 5.1.4). In literature for designing general corpora, some writers advocate a conscious attempt by the compiler to achieve balance via controlling the text topics (Biber et al., 1998). Sinclair (2004: 9) himself acknowledges that it is indeed a feature which creates “a strong urge to control” but since subject matter is realised through the lexis of a text, it cannot be used as an indicator of whether a text should or should not be included as only *external*, not *internal* criteria, should be used in making such decisions.

For the UCLanESB corpus, the following decisions were made. The number of exams to be captured at each level was to be equal; following the pilot study, the total number of exams per level raised from 10 to 15. Although this would inevitably lead to wordcount differences across individual exams, it would not necessarily be a drawback: for instance, a higher combined word total at C1 level would have the potential to reveal fluency differences in comparison with B1 and B2. Next, no exams were to be reduced in size; they would be included in their entirety, again to help with additional insights into success, but more importantly to uphold context for subsequent analyses of lexis. Finally, exam topic would not be used as an external design criterion. It was hoped that a lack of control in this aspect would naturally capture a random variety of subject matter at B1, B2 and C1, thus reducing subjectivity. As the pilot study also demonstrated, usable exams were already restricted by consent and global score; introducing the criterion of topic would have had a drastic effect on the ability to collect sufficient exams.

5.2.6 Corpus size

Similar to balance, size can also be a contentious issue when designing a corpus. With no maximum size (Hunston, 2002; Sinclair, 2004) and no concrete rule as to magnitude (Baker, 2010), it appears easy to accept the notion that “In general, the lengthier the corpus, the better” (Meyer, 2002: 33). Sinclair (2004 cited by Koester, 2010: 66) himself declares that in corpus design “small is not beautiful; it is simply a limitation.” Notwithstanding, creating a large corpus simply because it is possible provides no additional gains when a smaller corpus would suffice (Kennedy, 1998), so it is perhaps understandable that corpus size can be seen as “controversial” (Tognini-Bonelli, 2001: 62). Discussion here will thus focus on three areas: the type of corpus, the linguistic aspects to be analysed and the practicalities of spoken corpus construction.

Corpus type can have a dramatic effect on size. General corpora can range to billions of words in size whereas specialised corpora can contain anywhere between tens of thousands of words and millions of words: as a more precise figure, Aston (1997) states that small corpora can contain between 20,000-200,000 words. One key feature which dictates a corpus' volume is the *variation* of language which is required to achieve representativeness (Kennedy, 1998): "a rule of thumb is that the more varied the language, the larger the corpus required" (Baker, 2010: 96). Specialised corpora, especially those focussing on a particular type of language within a particular context, tend to be more homogeneous (Sinclair, 2004): they therefore tend to be much smaller in size. With this study being extremely specialised, the size would always be greatly reduced in comparison with other corpora.

Related to the previous point is the effect of research purpose. With corpus research mostly centred on quantitative patterns of language (Baker, 2010; Biber et al., 1998; Sinclair, 2004; Timmis, 2015), size cannot be established without careful consideration of the language aspects under examination and the analysis required (Baker, 2010; Biber et al., 1998; Meyer, 2002; Sinclair, 2004). Lexicographers, for instance, require vast amounts of data due to the high frequencies of grammatical words and the high numbers of infrequent lexical items which appear in corpora only once or twice (Adolphs, 2006; Biber et al., 1998; Koester, 2010; Meyer, 2002): as a guideline for this research purpose, a million word corpus would be considered insufficient (Baker, 2010). Studies focussing on grammatical patterns, on the other hand, can be conducted using corpora of various sizes (Meyer, 2002) since these patterns are more repetitive than lexical items (Adolphs, 2006). Despite this, specialised corpora, whilst smaller in size, can still benefit from their more homogeneous design. It is thought that lexis in such corpora occurs in higher concentrations since specialised lexis will appear with more regularity and prominence (O'Keeffe et al., 2007; Sinclair, 2004). Also, frequent word types analysis can be conducted using smaller corpora (Adolphs, 2006) and much more can be said about the use of language in the original context unlike large, less homogeneous corpora (Flowerdew, 2004 cited by Koester, 2010). The lexicogrammatical nature of this study may have therefore been susceptible to criticism for its very small size, but it was not wholly unjustified.

Small sizes in corpora can also be a product of the practicalities of compiling a spoken corpus which can conflict directly with a desire for larger corpus size (Adolphs, 2006). McEnery and Hardie (2012) remark that a surprising development in corpus

linguistics was that corpora became much smaller in size. Whilst justified in some cases, in others, such an outcome could be due to the practicalities of corpus construction: there is often “a compromise between the desirable and the feasible” (Stubbs, 2004: 113). The pragmatics of corpus design therefore acts as a limitation; text availability, funding and time restrictions can all inhibit the proposed corpora size (Baker, 2010). In the case of spoken corpora, this is especially true. Compiling spoken corpora can be problematic and time-consuming as often, speech needs to be recorded, appropriately and accurately transcribed, and then analysed (Baker, 2007; O’Keeffe et al., 2007; Thompson, 2004). For the UCLanESB corpus, the compiler thus had to bear research aims in mind and be realistic with the available resources. Sufficient data for basing valid and reliable conclusions had to be obtained, but not at the cost of obtaining an overwhelming amount of data, impossible to transcribe and analyse according to the time and funding available.

The pilot study corpus of C1 language consisted of 23,239 words. This figure was utilised to inform calculations of final corpus size. Whilst general corpus compilation can base size predictions on scientific calculations (see Meyer, 2002), they do not translate well to specialised corpora (Sinclair, 2004). The strategy followed in the pilot study was thus adopted. In this strategy, advised by Biber et al. (1998), Kennedy (1998), Meyer (2002) and Stubbs (2004), the size of individual corpus components, and not the intended overall corpus size, are focussed upon. By using the estimates offered by Adolphs and Knight (2010) and O’Keeffe et al. (2007) the approximate sizes of individual exams were used to calculate the size of the final UCLanESB corpus. The pilot study, however, found the calculations to be slightly optimistic when applied to learner language. Ten 15-minute exams at C1 level, the highest proficiency level, had been expected to yield approximately 25,000 words. In reality, 12 exams only captured 23,359 tokens. The decision was taken to increase the number of exams at each of the three levels to 15. This number was believed to surpass the 25,000 word target for C1. Though B1 and B2 learners would presumably produce fewer words, differences in total word count per level would in themselves allow additional conclusions to be drawn.

The final number of 45 exams (15 at B1, B2 and C1) was deemed appropriate for two reasons: practicality and purpose. As discussed above, spoken corpus compilation can be time-consuming. To build the final UCLanESB corpus, research aims had to be balanced against the resources available and what was realistically achievable. The pilot study ascertained that each exam took a maximum of 2 hours’ transcription; total transcription time would thus equate to approximately 90 hours. As the sole researcher

and corpus compiler in this study, an overwhelming amount of data would have clearly extended the amount of time needed for creating the corpus of B1, B2 and C1 language resulting in a delay of the analysis stage. In terms of purpose, the gains in collecting a larger amount of data had to be contemplated. As a spoken, specialised corpus, its size would inevitably be very small in comparison with multi-million or multi-billion word corpora. However, with strings and patterns repeated in corpora of a more homogenous nature, it was expected that the UCLanESB's corpus' target of 60-75,000 words would still allow the research questions and aims to be achieved. In total, the UCLanESB corpus contained 70, 578 tokens. Nevertheless, its size could still be considered a limitation as the ability to generalise findings does diminish.

5.2.7 Corpus sampling

In a statistical sense, samples refer to groupings of individual language examples which represent the population being investigated (McEnery & Hardie, 2012). If done correctly, sampling allows generalisations to be made “reliably and validly about the population as a whole” in order to satisfy research goals (Kennedy, 1998: 74). Since population refers to “the group of people whom the study is about” (Dornyei, 2007: 96) it is easy to surmise that this study's potential research population could have referred to *all* successful B1, B2 and C1 learners; any findings regarding success would ultimately be generalised to all learners at this level partaking in similar spoken tasks. Clearly, such a population would extend beyond the realistic scope of this study's aims given the expansive linguistic variation to be encapsulated, the differing definitions of success available, the numerous contexts surrounding the learners' language use, and the changing make-up of individuals producing it. Hence for data collection purposes, the population was taken to be B1, B2 and C1 learners at the University of Central Lancashire's Preston campus. This population, according to the research aims, was refined to *successful* B1, B2 and C1 learners taking UCLanESB speaking exams using the criteria exemplified in Section 5.2.4.

As “spoken texts are harder to sample” (Kennedy, 1998: 74), consideration of the research population was vital in this study as it helped to identify a suitable sampling method from the many that exist. Publications dealing with corpus design, such as McEnery and Hardie (2012), Meyer (2002), and Tognini-Bonelli, (2001), often refer to

the norms followed in producing general corpora which rely upon a *sampling frame*. A sampling frame contains:

“A definition, or set of instructions, for the samples to be included in a corpus...[it] specifies how samples are to be chosen from the population of text, what types of texts are to be chosen, the time they come from and other such features. The number and length of the samples may also be specified.”

(McEnery & Hardie, 2012: 250)

As such, references are often made to techniques and calculations that can be used by general corpus compilers to determine the genres to be encapsulated, the proportions they comprise and the size or number of words to be captured in each text. It was clear, however, that the specialised nature of the UCLanESB corpus would mean that individual texts would be homogeneous and would be comparable (dependent upon level and fluency) in size. Some of the advice given was not entirely applicable. Despite this, a sampling frame was still constructed. The sampling summaries offered by Kennedy (1998) and Tognini-Bonelli (2001) suggest that external criteria can be used to identify texts for possible inclusion. In a way, Sections 5.2.4 to 5.2.5 outline the external criteria decisions which had already been taken. The sampling frame thus consisted of the following decisions: language had to be from B1, B2 and C1 UCLanESB speaking exams conducted at Preston Campus; all candidates had to secure a score of 3.5 or 4 to be considered successful; exams would be included in their entirety; and 15 exams at each of the three levels would shape the final UCLanESB corpus. In a similar vein, candidate characteristics such as gender, age, nationality and time spent learning English would be borne in mind but not used as external criteria unless there was a potential for skew in nationality (see Section 5.2.4). These factors can sometimes be used when individual characteristics may be a focus of research (see McEnery, Xiao & Tono, 2006; Meyer, 2002) but in this study, the language itself was more important than the individuals who created it. For example, had calculations been employed to identify the number of males or females to be included, there would be no guarantee that they would be representative of *successful* speakers. In the end, the resulting sampling frame worked well to identify a number of usable exams, but sampling method still required contemplation.

In scientific research, a vast range of sampling techniques exist (see Cohen et al., 2011; Dornyei, 2007). These are divided between probability and non-probability techniques. According to Meyer (2002), probability sampling, the most reliable, involves

the use of pre-selected demographic information and mathematical formulas to maintain text representativeness; non-probability sampling, often common in corpus linguistics, is typified by its unprincipled, convenient character. Additionally, a method stated by Kalton (1983 cited by Meyer, 2002: 43) refers to the use of expert choice sampling where “one decides before-hand who would be best qualified to be sampled”. In the UCLanESB corpus, a combination of expert choice and non-probability sampling was utilised. Using my knowledge of the UCLanESB speaking exams, exams that would not be representative of the population of successful learners were discounted. However, as the pilot study discovered, since external criteria had a dramatic effect on the number of usable exams, convenience sampling was then employed for selecting exams to be incorporated into the corpus. For instance, in one exam intake, a total of 60 exams took place. Of these, only three were useable. Of course, the convenience sampling technique is open to criticism. The resulting “*opportunistic*” corpora are said to operate without a rigorous sampling frame; they represent “nothing more nor less” than data that was collectable (McEnery & Hardie, 2012: 11). However, as noted earlier, a sampling frame had been established so it would not be entirely accurate to deem this corpus wholly opportunistic. Criteria allowing for expert choice sampling were adhered to, but due to the small numbers of usable exams collected, an element of convenience sampling had to be integrated.

5.2.8 Transcription

Once decisions regarding the composition of the corpus had been made, the capture of learner spoken performance in the UCLanESB exams had to be planned. It is important to note, however, that whichever transcription system was used, resulting data would inevitably offer an incomplete account of the original speech. As a multimodal channel of language, various layers would be lost, the original mode of speech would be altered in its written embodiment, and ultimately, the product would offer a selective, subjective portrayal of the original interaction (see Adolphs, 2006; Adolphs & Knight, 2010; De Cock, 2010; Edwards, 1992; Gilquin & De Cock, 2013; Kennedy, 1998; Ochs, 1979; Thompson, 2004). Though writers warn of the risks of documenting speech using written conventions, very few alternatives remain for researchers wishing to keep transcriptions. For this reason, the decisions taken in the transcription of data need to be documented so that ‘spokenness’, the extent to which the transcriptions depicted the

original discourse (De Cock, 2010), can be determined. Transcription, quite simply, is the “process of representing spoken language in written form” (Leech, et al., 1995: 15). Despite requiring great skill, time and consistent application, transcriptions will form selective representations based on their relevance to the researcher’s aims and uses for the data (Chafe, 1995; Edwards, 1992).

The options available to corpus linguistics in transcribing accurate accounts of speech create somewhat of a “black hole” (McCarthy, 1998: 13). Not only do decisions need to be made as to whether transcriptions will be orthographic, prosodic or phonetic, but multiple layouts are also available (Thompson, 2004). Ultimately, the chosen system should match the research purpose, the time and budget available and the readability of the resulting transcripts; it should also be transparent and replicable to aid other enquiries made of the corpus (Adolphs & Carter, 2013; Adolphs & Knight, 2010; Crowdy, 1994; Kennedy, 1998; McCarthy, 1998). The researcher opted for a *broad* transcription system rather than *narrow*. Narrow transcription systems can contain considerable detail as to the intonation, stress, gesture, voice quality and pronunciation of speech but they can be difficult to read and interpret; broad transcriptions whilst lacking in detail can offer gains in transcription accuracy and do not suffer from an overload of irrelevant, distracting details (De Cock, 2010; Dressler & Kreuz, 2000). This study of success focussed heavily on the lexis produced by learners and what was said so readability, transcription rates and transcription accuracy were of more significance:

“For a linguist whose interest is in the patterning of language and in lexical frequency...there will be little need for sophisticated transcription...the main consideration will be the quantity and speed of transcription work.”

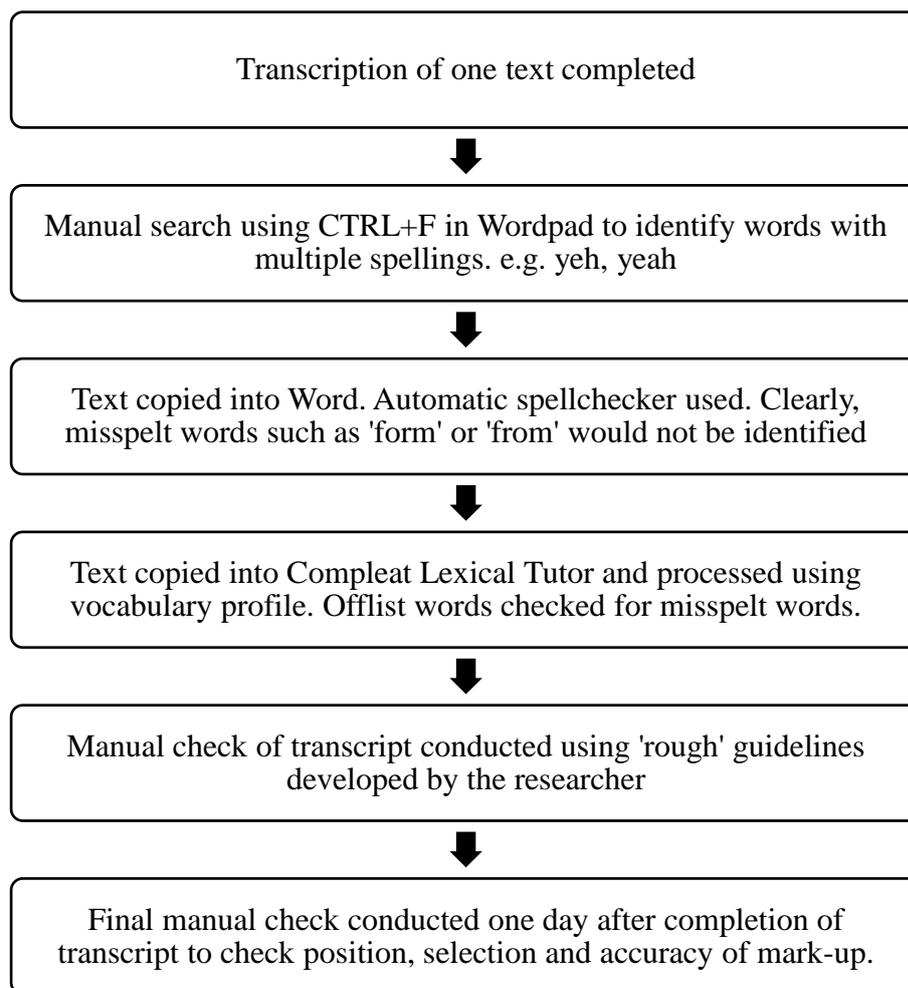
(Thompson, 2004: 59-60)

CANCODE conventions (Adolphs & Carter, 2013) were thus chosen as the researcher felt that they satisfied the readability criteria established by Edwards (1992): codes were placed in-line with text, thus strengthening the temporal representation of items; a minimum number of symbols helped to reduce the demands placed on readers’ short term memories; and a vertical layout, as opposed to columnal, seemed more appropriate for adults whose conversations are mostly balanced equally between speakers. Similarly, with potential users of the UCLanESB corpus searching information for pedagogic purposes, the broad description system would suffice as it is easy to learn and decipher and it satisfies the obligation that transcriptions should be useful to a wide-

ranging group of users (see Chafe, 1995; Timmis, 2015). Whilst CANCODE allows for simple orthographic transcriptions of speech to be combined with the inclusion of simple mark-up, no annotations were added. Mark-up refers to the application of codes which act as metalanguage for signifying typography and layout (Adolphs, 2006). In speech, features such as overlaps, changes in speaker, pauses and other extralinguistic information can be signalled using mark-up. Annotation, conversely, refers to the process of inserting analytical information such as part-of-speech or error tags to transcripts (Adolphs, 2006). Clearly of use for subsequent analysis of texts, the researcher felt that such annotation would i) obscure the clarity of transcriptions thus reducing readability, and ii) would not offer significant gains considering the amount of time that would have been spent checking the accuracy of automatic annotation programs (e.g. CLAWS, UCREL, n.d.). Only CANCODE codes were thus included.

As warned by McEnery and Hardie (2012), the transcription process was incredibly time-consuming. The wordcount estimates identified in Sections 4.1 and 5.2.5 suggested that each 15-minute exam would yield approximately 2500 words (see Adolphs & Knight, 2010; O’Keeffe et al., 2007) so with one hour of conversation taking up to 20 hours to transcribe (McCarthy, 1998), it was apparent that one sole transcriber would require a great deal of time. Fortunately, the researcher discovered that transcription and mark-up took approximately 2 hours for each 15-minute exam. In spite of the amount of time taken in transcribing more than 40 exams, it was felt that greater benefits in mark-up consistency were gained. This therefore left the topic of transcription consistency to be considered. Humans by definition are fallible (Goedertier, Goddijn & Martens, 2000). Choosing the CANCODE system and stipulating decisions prior to transcription did help in terms of consistency but careful proofreading was still required if texts were to be “of high quality and as free as possible of error” (Kennedy, 1998: 82). To achieve this, a procedure for checking each transcription, as recommended by Crowdy (1994) and Thompsen (2004) was devised. Incorporating both manual and automatic procedures, Figure 16 displays the steps that were taken in ensuring that transcriptions were of the highest accuracy and consistency.

Figure 16: Procedure followed for checking transcription accuracy and consistency



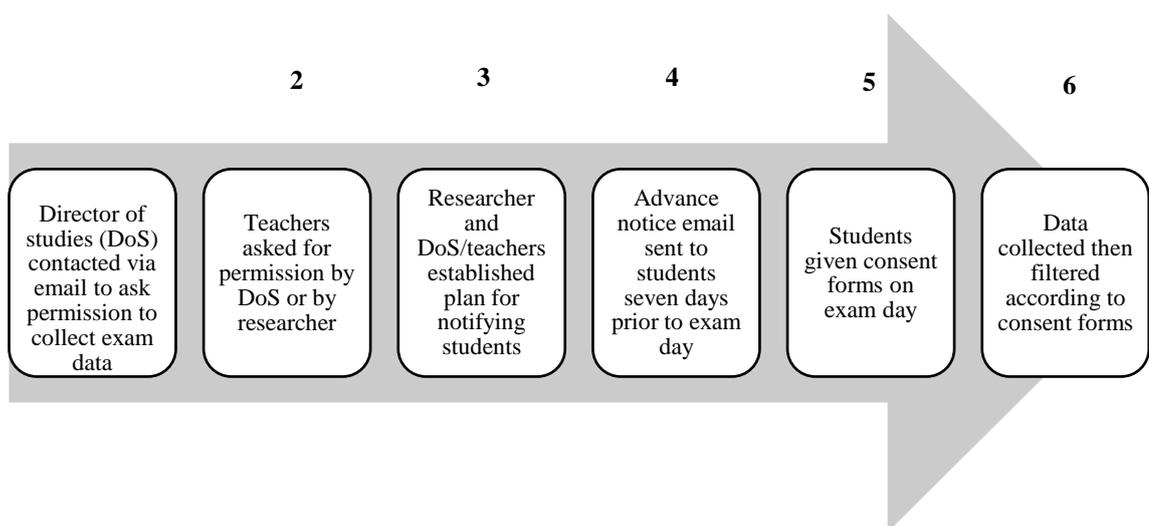
5.2.9 Ethics

This study adhered to the following ethical guidelines and principles: British Association for Applied Linguistics [BAAL] (2006), the Data Protection Act [DPA] (1998), UCLan’s Ethical Principles (2012a) and UCLan’s (n.d.-a) ‘e-Ethics System’. Whilst some corpus linguistic studies have remained “silent” on the matter (McEnery & Hardie, 2012: 67), ethical considerations were made throughout the research process with particular respect paid to informed consent, anonymization and confidentiality and data storage.

Informed consent is identified as the effort made by the researcher “to provide sufficient information about the purpose and methods of the study...that might reasonably be expected to influence willingness to participate” (UCLan, 2012a: 3). A researcher should inform potential respondents about the nature of research, its duration, the communication of results and the length of time that data may be kept. BAAL (2006)

concurr but also stipulates that details relating to objectives, potential consequences and data security also need to be made clear. This study opted for overt consent via the use of consent forms in Appendix 8 rather than assumed consent (see Dornyei, 2007 for active and indirect approaches to consent) since “explicit consent” was called for (DPA, 1998: 53; UCLan, 2012a: 3). The process below in Figure 17 was thus followed to ensure consent was explicit and that the purpose of the research was fully communicated. Whilst this process was largely unchanged, amendments were made when necessary to ensure that participants understood the information given, that full information was given and that participants voluntarily gave their consent (see Cohen et al., 2011). Also as some exams came from real exam sessions, efforts were made to minimise disruption and avoid any undue stress for staff and students. Therefore, to forewarn learners of the research taking place, an advanced-notice email (see Appendix 6) was distributed seven days prior to the research or, by teacher request, a short presentation was given (see Appendix 7). At all times “power relations” were considered and attempts made to reduce them (BAAL, 2006: 4). This not only concerned learners but teachers who may have been asked to participate by Directors of Studies. For some students, the exams acted as “inducements” or incentives to participate, not in a monetary sense, but in an ‘opportunity to practise’ sense (Cohen et al., 2011: 80). Even still opportunities for questions were given and learners not wishing to take part were not denied the extra exam practice. Following collection of all recordings and consent forms, data was stored securely as per the DPA (1998) and UCLan (2012b)’s principles. Had anyone other than the researcher gained access, confidentiality and anonymity would have been jeopardised.

Figure 17: Overview of informed consent process



As a corpus linguistics study in which personal data and exam data were collected, data storage was not the only threat to anonymity and confidentiality. Transcriptions also had to meet the same standards. At the root of anonymisation lies the tenet that “information provided by participants should in no way reveal their identity” (Cohen et al., 2011: 91). In short, any subjects involved in research should not be identifiable to others including the researcher (Cohen et al., 2011). However, the researcher possessed personal data about the test candidates, thus making it difficult for them to be fully incapable of identifying individuals. With complete anonymity being difficult to accomplish, perhaps definitions of confidentiality bore more significance. This notion instead concerns the researcher’s undisclosed knowledge of connections between data and the individuals to which it belongs. Often befitting of more sensitive data content, confidentiality ensures that the identification of participants is protected (Cohen et al., 2011). Furthermore, the very act of turning audiovisual recordings of conversation into written transcriptions omitting identifiers could constitute one way to make candidates anonymous (McEnery & Hardie, 2012) but this cannot be guaranteed. Whilst not always satisfied by the sole omission of participant names or proper nouns (Adolphs & Knight, 2010; McEnery & Hardie, 2012; Rock, 2011), the following decisions were taken. Firstly, CANCODE transcription conventions (Adolphs & Carter, 2013), discussed in Section 5.2.8, allowed for candidate and teacher names to be taken out. As a “sex-based code” (Rock, 2001: 11), candidates were differentiated first numerically and then by gender e.g. <\$1F> <\$2M> for speaker one female, speaker two male; examiners received the neutral code <\$0> with no reference to gender being made. Secondly, any third parties mentioned in the data, such as friends of test candidates, were replaced using the <\$E> </\$E> code. Thirdly, although many writers advocate the omission of details such as place names (see Rock, 2001), country names were included. Town names were felt to have been too specific and may have potentially helped transcript readers to identify a particular speakers. Countries, on the other hand, were often discussed due to exam topics such as ‘travel’; taking these names out may have been problematic for analysis or achieving a full understanding of candidate answers. Finally, metadata and additional details obtained in the consent forms were not stored in the same place as the transcripts themselves in accordance with ethical and legal guidelines (BAAL, 2006; DPA, 1998; UCLan, 2012b).

5.3 Data analysis

Sections 5.1 and 5.2 introduced the UCLanESB corpus demographics and explained the decisions taken in its design. This final segment will deal with the methods used in analysing the corpus' data and the rationale underpinning them. It is important to point out here that this study took a corpus-driven (see Section 3.1.2) approach to language analysis; items for analysis arose from the text and were not determined prior.

To aid discussion, analytical techniques will be organised according to research question although some crossover will be evident. For instance, RQ2a and RQ2b utilised very similar qualitative analyses of data using concordance lines, a topic discussed in depth in 5.3.2 but one which was omitted in 5.3.3 so as to avoid repetition. It is important to highlight also, that there will evidently be some overlap with the processes outlined in the pilot study chapter, though admittedly, a much greater amount of detail will be offered here.

5.3.1 RQ1: What percentage of the words used by successful B1, B2 and C1 learners come from the first 1000 and second 1000 most frequent words in English?

To answer the first research question, the pilot study procedure (see Section 4.5.1) was once again followed. Lexical frequency profiles (Laufer & Nation, 1995) were computed using uncoded text files of the B1, B2 and C1 data via the Vocabprofile function on *Compleat Lexical Tutor* [LexTutor] (Cobb, n.d.-b). As a freely available program, this software contains a “unique feature” not offered to users of programs such as *AntConc* and *WordSmith Tools 6.0*: that of allowing users to create vocabulary profiles from their corpora from pre-embedded word lists such as the BNC, BNC-COCA, and GSL-AWL (Diniz, 2005: 25). In answering this RQ, few alternative, well-known options remained for producing vocabulary profiles so LexTutor was chosen as a result.

In preparation for this process, coded text files containing *only* learner turns for each of the 45 exams had to be filtered to remove mark-up from the transcriptions. Any transcriptions containing coding would not have worked well as coding could have been counted as an ‘offlist word’ thus distorting family, type and token figures. Removing mark-up was, at first, done manually. Then, to ensure no codes had been missed, text was copied into a Word document so that the ‘Find’ function could be used to identify any unnoticed symbols. Speaker codes, extralinguistic information, pauses, overlaps,

interruptions and truncated words were thus removed; only the utterances produced by learners in the exams remained. To improve ease of use, due to the fact that LexTutor only allows text files to be uploaded separately (Diniz, 2005), individual text files were combined for each of the three levels. This process was repeated for all Part A, Part B and Part C exam sections at each the three levels. This meant that instead of copying and pasting exams and exam sections on multiple occasions, one text file for each analysis could be used. To produce relevant vocabulary profiles, the BNC 1-20k option was selected as a basis for comparison. BNC 1-20k generates vocabulary profiles based on the first 20,000 word families in English. Although a vocabulary profiler based on Laufer and Nation's classic "4-way sorter" is also available for use (Cobb, n.d.-c: para: 1), words are only sorted into K-1, K-2 and AWL lists; results beyond K-2 would therefore have a more academic than general English emphasis. Despite embedded lists being proportionally founded upon written English more than spoken English, the BNC was considered appropriate for the purposes of this study. Any lists based on the General Service List, Academic Word List or Corpus of Contemporary American English could have skewed findings for spoken, non-academic English produced by learners in a British English education context. The BNC 1-20k lists may not have been the optimum choice, but it was more suitable than the alternatives.

Vocabulary profiles were created for Parts A, B and C as well as for entire exam scripts at B1, B2 and C1. The resulting outputs (see Appendix 9 for an example) at each level were then compared according to the word families, word types, and word tokens used. This was done only for K-1 and K-2 words for two reasons. Firstly, general consensus (see Section 3.3) states that the 2000 most frequent words in English comprise major portions of receptive or productive language skills; it is essential that learners learn these words if they are to at least express or comprehend ideas in either written or spoken form (see Section 3.3). Secondly, the pilot study based on C1 learner language discovered that very few words, less than one in twenty, originated from a different word family band. The gains in examining such vocabulary or basing findings upon it may ultimately not have had a similar level of significance as those based on K-1 and K-2 vocabulary.

When differences in K-1 or K-2 percentages and numbers were identified, findings had to be strengthened via significance testing. Significance testing is essential as it examines the likelihood that a particular event or result is attributed solely to chance, i.e. that it is a coincidence (McEnery & Hardie, 2012). Significance testing has much more weighting than saying that a particular aspect simply occurs more or less often than

another. To compare means, therefore, independent t-tests were conducted. T-tests were chosen as they are suitable for research with small samples (Oakes, 1998); independent, not dependent tests were performed as the two groups did not consist of the same individuals (Salkind, 2014). The tests would not indicate whether the same individuals had improved or worsened, but whether different groups of learners at two distinct levels produced different lexis. In preparation for the tests, family, type and token occurrences at the K-1 and K-2 levels were inputted into *SPSS* (IBM Corporation, 2015). Once the figures had been entered, two-tailed t-tests were conducted using the selected data for instance, the number of B1 K-1 types compared against the number of C1 K-1 types. Though a critical value of 0.05 is usually used to ascertain whether the null hypothesis can be rejected or not, in many calculations, significance was found to be lower than 0.01.

The last stage of analysing data to answer RQ1 involved calculating type-token ratios. Type-token ratios [TTR] are useful as they can provide insights into lexical variation and the amount of repetition evidenced in speech. Although TTRs are clearly affected by the overall size of corpora, thus making their comparison more problematic (Barnbrook, 1996), they were only used across UCLanESB data. Token totals did change across B1, B2 and C1 (see Section 5.1) but they were comparable in that a general corpus, for example the BNC, was not directly compared to the small UCLanESB corpus.

5.3.2 RQ2a: What were the 20 most frequent words at B1, B2 and C1 and their notable collocations, colligations and functions?

In response to question 2a, frequency lists were generated using the Wordlist function in *WordSmith Tools 6.0* (Scott, 2015). This software was chosen partly due to the range of functions permitted within it (see McEnery & Hardie, 2012) and partly due to the level of detail available in frequency, concordance and collocational analyses of large data sets (Kennedy, 1999). Its ease of use was also advantageous and allowed for successful data analysis in the pilot study, so its application was extended into the main study. With regards to the use of frequency lists, it is acknowledged that the first stage of basic corpus analysis involves observing the occurrence of individual words; the lists identify word types within a corpus, the raw frequencies and their percentage coverages (Baker, 2006; Barnbrook, 1996; Hunston, 2002). To conduct analysis for RQ2a, frequency lists were thus compiled from the coded text files produced in the transcription

stage. These were produced for B1, B2 and C1, as well as the individual exam sections at each level.

Once the lists had been generated, a procedure similar to that used in the pilot study (see Section 4.5.2) was followed. Firstly, introductory analysis involved comparing the B1, B2 and C1 frequency lists i) with reference corpora, and ii) against each other. When compared in such a way, frequency lists can become “much more meaningful” (Barnbrook, 1996: 46), especially in the case of specialised corpora (Hunston, 2002). Since speakers have a large option of choices from which to choose when interacting, the high frequencies of some items will inevitably reveal speaker preferences and intentions “whether conscious or not” (Baker, 2006: 48) (see Section 3.1.2). The choice of reference corpus does, of course, affect the reliability of comparisons. Defined as a large corpus, “representative of a particular language variety”, they act as ‘benchmarks’ demonstrating typical language usage within that variety (Baker, 2007: 30). If chosen imprecisely, findings could suffer due to the lack of similarity or ‘appropriateness’ with the language under inspection (Scott & Tribble, 2006). In the pilot study, frequency lists from the Spoken BNC and CANCODE were utilised. Though this approach was applied in the main study, criticisms do exist of findings based on NS-learner comparison. Admittedly, the Spoken BNC and CANCODE data were based on a variety of native speakers in numerous interaction contexts; the extent to which they are representative of speech similar to that contained in the UCLanESB learner corpus clearly diminishes. However, with definitions of success needing to counter the comparative fallacy, such comparison may have been able to distinguish similarities between learners and NSs for those assuming that the NS represents the ultimate target of second language acquisition (see Chapter 1). The decision was subsequently taken (in Section 4.8) to supplement comparison of UCLanESB data with a reference corpus containing *learner* speech. The Louvain International Database of Spoken English Interlanguage (Gilquin, De Cock & Granger, 2010) was selected. Containing over a million words of learner speech from individuals varying in nationality, age and proficiency, it was considered ideal for comparison; based on learner interviews consisting of three tasks (set topic, free discussion and picture description), it was thought to be particularly comparable to the data in the UCLanESB corpus (see Gilquin, De Cock & Granger, 2010). In preparation for subsequent analyses, a wordlist composed of learner turns only, was constructed from LINDSEI using the Wordlist function in WordSmith.

It is now necessary to explicate the manner in which comparison took place so that the identification of word types for further investigation can be understood. At first, raw observations of data were carried out to identify words which differed greatly in their list positions or frequencies. Any words appearing at one UCLanESB level but not others, or any words appearing in the UCLanESB data but not the NS data were immediately earmarked for further investigation. Changes in word type position on the lists, were also taken into consideration for additional analysis. In the pilot study, however, using word rankings and frequencies on their own left a gap in the analysis. Since corpus size was not accounted for, it was difficult to adequately understand how word frequency changed between the UCLanESB and NS data: word frequency *within* a corpus was established, but word frequency differences *across* corpora were not distinguished. Normalised frequencies were hence incorporated into the analysis for RQ2a. Normalised frequencies essentially reveal how often a word can be expected within a set number of tokens (the base of normalisation), for instance per every 10,000 words or 1,000,000 words. They are calculated by dividing a word type's total occurrence by the size of the corpus; this figure is then multiplied by the base of normalisation to allow comparisons to be made across corpora of differing sizes (McEnery & Hardie, 2012). Since the base of normalisation adapts to corpus size, normalised frequencies in this study utilised a base of 10,000 for LINDSEI, Spoken BNC and CANCODE comparisons, and cross-level comparisons of the UCLanESB data.

Discussion of RQ2a's data analysis has so far examined the creation of wordlists, the choice of reference corpora, and comparison methods. With corpus linguistics merging quantitative and qualitative research methods, it is clear that thus far, only quantitative approaches have been considered. The qualitative aspect of data analysis now needs to be explored to understand the way in which a clear frequency list flaw was minimised. One striking limitation of frequency lists is that items within them are isolated and removed from their original contexts; word types with the capacity of expressing multiple meanings are combined and presented as one single item (Barnbrook, 1996). For this reason, once a word was selected for further investigation, the use of a concordancer was integrated into the process. A concordancer is:

“a program that searches a corpus for a selected word or phrase and presents every instance of that word or phrase in the centre of the computer screen, with the words that come before and after it to the left

and right. The selected word, appearing in the centre of the screen, is known as the **node word**”

(Hunston, 2002: 39)

The concordancer in *WordSmith Tools 6.0* was extremely helpful in this respect as once the lines were presented, collocational, colligational and functional insights could be gained using the collocates, patterns and clusters features. Together they helped to paint a picture as to a word’s function in learners’ utterances; not only was information obtained as to what a particular word occurred with, but information was obtained as to what structure it usually occurred in and the purpose it fulfilled.

When performing this qualitative analysis via the use of concordance lines, one approach in particular was utilised; that of ascertaining salient collocations within the corpus via the use of significance testing. Word co-occurrence is a concept closely related to concordance lines “since the idea of two words occurring in a common context is similar to that of two words occurring in the same concordance window” (Oakes, 1998: 159). However, the ability to identify the co-occurrence of two words via frequency alone is insufficient in queries of collocational importance or strength. Though the pilot study was able to identify patterns of words stemming from the frequency of collocations, it was unable to assess their worth in learner speech in the UCLanESB exams: raw frequencies are incapable of suggesting whether any degree of importance is attached to the word pairing (Hunston, 2002) and using frequency alone can often result in function words occupying the highest positions (Baker, 2006).

For this reason, quantitative measures of collocational strength were adopted. Essentially, these measures would compare the number of appearances of a collocate within a four- or five-word window (or span) to the left and right of the node (Baker, 2006; McEnery & Hardie, 2012): in WordSmith, calculations were left at the default window of 5 words either side of the node. However, the range of statistical tests can seem somewhat baffling to a novice corpus linguist, so for assessing collocational strength, three of the “most commonly used measures” (Hunston, 2002: 70) were considered: z-scores, t-scores and Mutual Information (see Barnbrook, 1996; Hunston, 2002; Oakes, 1998). A z-score compares the frequency of a co-occurrence within a specific window with the frequency expected if they were to co-occur within that window by chance alone; a high score indicates a stronger “degree of collocability” (McEnery & Wilson, 2001: 87). A t-score, on the other hand, though similar to the z-score in some ways, is considered more accurate in the analysis of words with relatively low frequencies

due to a slight change in the calculation (Baker, 2006; Barnbrook, 1996): any t-score greater than 2 would signify a potentially “interesting” focus (Barnbrook, 1996: 98). The small size of the UCLanESB corpus suggested that this might have been the most suitable score. Despite this, at first, Mutual Information [MI] was used as it seemed to receive the most attention in discussions of collocational statistics. Mutual Information is concisely defined by Oakes (1998: 253) as the “probability of two things happening together compared with the probability of their occurring independently; it is thus a statistical measure of the degree of relatedness of two elements.” It essentially compares the co-occurrence found in data with expected co-occurrences if words appeared randomly. To make sense of scores obtained, an MI score of 3 or above, as stated by Hunston (2002), was considered significant.

Upon calculating MI scores for UCLanESB data, a problem arose. Collocations of a particular node were difficult to analyse due to low-frequency collocates appearing high in the list (see Figure 18 below containing MI scores for collocates of *so* in C1 data). As is remarked by Baker (2006) and Hunston (2002), MI scores can suggest that a collocate is significant when in fact, it occurs rarely in the corpus; the high score is thus a product of the low-frequency collocate appearing on most occasions in close proximity to the node. Though intuition can be used to such filter results (Baker, 2006), it was felt that a greater degree of subjectivity would have been introduced into the analysis than was necessary. T-scores were therefore incorporated into the analysis. When using WordSmith’s collocates tab to compute t-score relationships, the resulting lists appeared much more relevant; t-scores were often much higher than the threshold of 2 and the frequency of collocates coincided with vocabulary that had been expected to appear. Nevertheless, though t-scores were found to be more useful, MI scores were still computed. Since there is no single, definitive significance score providing “the best possible assessment” of collocational strength, a combination of scores is sometimes needed in order to gather a broader impression of the data (Barnbrook, 1996: 101). For instance, though the MI score can reveal the collocational relationship between a node and collocate, it is the t-score that determines the amount of evidence for that collocation within the corpus (Hunston, 2002); the combination of these measures can reveal the meaningfulness of the collocation.

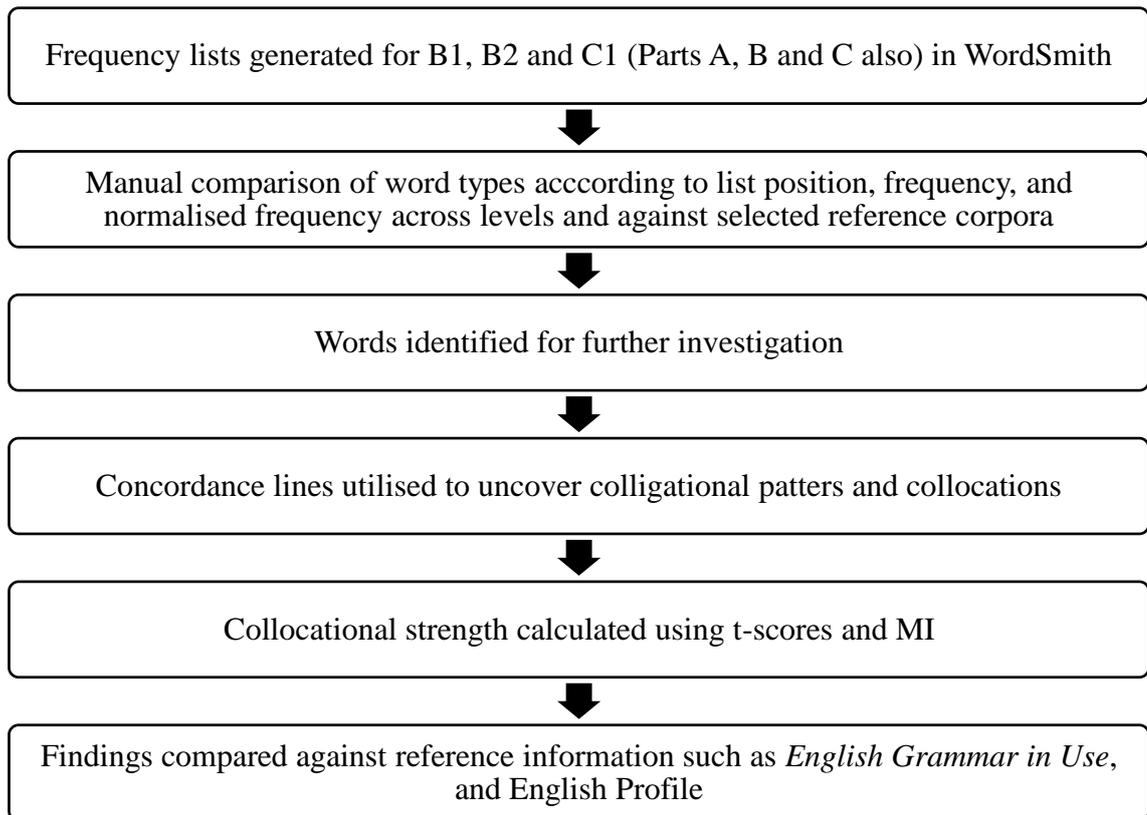
Figure 18: Collocates of 'so' with high MI scores in C1 data

N	Word	With	Relation	Set	Texts	Total	Total Left	Total Right
1	TASTE	so	7.339		1	2	1	1
2	BLACK	so	7.339		1	2	1	1
3	BEACHES	so	7.339		1	2	2	0
4	TOW	so	7.339		1	2	1	1
5	FINISHED	so	7.339		1	2	0	2
6	WOMENS	so	7.339		1	2	0	2
7	ARABS	so	7.339		1	2	2	0
8	TICKETS	so	7.339		1	2	2	0
9	PACKED	so	7.339		1	2	1	1
10	CURIOUS	so	6.924		1	3	1	2
11	SO	so	6.469		15	279	12	12
12	ITALIAN	so	6.339		1	3	2	1
13	PEOPLES	so	6.339		2	2	1	1
14	AMERICAN	so	6.339		1	2	1	1
15	AGO	so	6.339		2	2	1	1

Finally, to consolidate potential findings for RQ2a, information external to the corpora was exploited. This took different forms. The former involved the use of *English Grammar in Use* (Carter & McCarthy, 2006), a publication offering comprehensive insights into the usage of individual words in speech. Based on the CANCODE corpus, it contains NS examples of spoken and written English which were used to explain the patterns discovered in concordance analysis. To develop findings further, comparisons of individual words, or one of its colligational patterns, were carried out using information from EVP (CUP, 2015a) and EGP (CUP, 2015b). Though based on *written* learner language, and therefore questionable in their applicability to spoken conversation, these tools were found to be very useful. Not only were they constructed using examples of learner language, unlike *English Grammar in Use*, but they identified the Common Reference Level at which a word or a particular usage function is acquired together with the CEFR can-do statement to which it could correspond. Whilst aiming to document what learners “DO know, not what they SHOULD know” (CUP, 2015c: paragraph 2), such information was clearly invaluable for explaining or identifying potential differences in success between B1, B2 and C1 learners. Finally, concordance lines from LINDSEI’s learner turns wordlist and the BNC (using Brigham Young University British National Corpus BNC, n.d.) were then compared against the learner language in UCLanESB’s corpus to identify similarities or differences.

To summarise this section, the flowchart below in Figure 19 gives a visual representation of the steps followed in RQ2a’s analytical process:

Figure 19: Summary of RQ2a's analytical procedure



5.3.3 RQ2b) What were the 20 most frequent keywords at B1, B2 and C1 and their notable collocations, colligations and functions?

As noted previously, keywords are words that appear significantly, or unusually, more or less often than in a chosen reference corpus. This elevated occurrence is said to indicate the ‘aboutness’ of a text since function words typically occupying high positions in frequency lists may be replaced with lexis of more interest for the researcher. Typically, the words elicited occupy three categories. In addition to a text’s proper nouns, words revealing the topics or content included in a text are presented, and finally, words associated with the style, or nature, of the language are captured (Scott, 2015). For researchers interested only in words giving the ‘aboutness’ of a text, the use of a stoplist (see Baker, 2006: 127-8) can be used but as the features making B1, B2 and C1 candidates successful had not yet been determined, no such list was utilised. Therefore, to discover which words were key in the UCLanESB corpus, keyword lists were generated using WordSmith. This software was useful as it supplied keyword information relating to the frequency and coverage of a keyword both in the corpus under inspection and the

reference corpus; more importantly, it automatically computed a keyness figure using log-likelihood accompanied by a p score confirming that the keyword indeed occurred significantly more frequently (see Baker, 2006). The lists were compiled using coded text files (as in RQ2a) for B1, B2 and C1, as well as Parts A, B and C. Before doing so, however, consideration had to be given to the process underpinning keyword list construction.

Scott and Tribble (2006) identify four crucial stages in producing and interpreting keyword lists. The fundamental principle states that any words repeated throughout a text or texts are likely to be contained within a keyword list. In corpus terminology, this relates to repeated *types* rather than *families*: as an example, inflections of a verb, for instance *walk*, *walks*, *walked*, are treated separately meaning that only one of these items may appear in the keyword list. The pilot study had already detected lexis found to be key in 12 of the C1 texts. This indicated that some of the words associated with exam language, for instance *think* and *because*, were presumably likely to be identified once again in the main study. The second step, again similar to RQ2a, involved the selection of a large but appropriate reference corpus wordlist that would act as a suitable comparator against the UCLanESB corpus. Careful consideration was therefore given to the reference corpora used previously. Since data for CANCODE and the Spoken BNC is not readily available, only two options were left: use of the LINDSEI wordlist and use of the BNC World list offered in WordSmith. With LINDSEI data being of a similar structure and nature to UCLanESB data, it was predicted that resulting keyword lists may only be able to reveal lexis typically associated with topic or ‘aboutness’. As LINDSEI centres on an interview in which candidates offered their thoughts on a set topic, free discussion and picture description, the words considered key may have only related to the subjects students discussed and not the abilities evidenced. It would therefore not reveal what made students successful as similar structures would appear throughout LINDSEI and UCLanESB. The decision was thus taken to combine analysis of keyword lists produced using LINDSEI learner turns, and keyword lists produced using BNC World. Though BNC World contains NS, written language as opposed to learner spoken language, the researcher anticipated that it would reveal more about the structures used by learners to preface discussions about particular exam topics; as Scott and Tribble (2006: 65) remark, “even the use of a clearly inappropriate reference corpus...may well suggest useful items to chase up using the concordancer”. Since this study sought to examine what makes learners successful in relation to the CEFR, a keyword list produced using general English

was assumed to be able to present keywords that would i) associate well with the abilities B1, B2 and C1 learners can demonstrate, and ii) combine with the findings from comparison with LINDSEI.

The penultimate principle of Scott and Tribble's (2006) process related to the statistical rigour that would help extract meaningful keywords. Before processing keyword lists, a threshold level for minimum occurrence within the corpus must be set. In WordSmith, this figure is automatically set to 3; it eliminates words which may be high in keyness, but low in frequency (Scott, 2015). This condition is also reinforced with a default 5% figure for occurrence across texts; any keyword identified has to appear in a minimum of 5% of texts. For the UCLanESB data, these thresholds were perhaps a little low. As an illustration, each level consisted of 15 exams. If a keyword had to appear in 5% of texts, this would equate to occurrence in 0.75 of texts. In hindsight, both this and the minimum occurrence figure could have been raised so as to identify significant words that appeared in a greater number of exams. It transpired, however, that increasing the text threshold from 5-20% (0.75 to 3 exams), for instance, would only have affected keyword data in a total of 2 cases for the entire exam lists at B1, B2 and C1. Hence, the small thresholds did not have as big an impact as expected.

The final stage concerned using statistical significance scores to elicit both the order of keyword lists, but more importantly the meaningfulness of the words it comprises. In WordSmith, two options, chi-square or log-likelihood are automatically calculated during comparison with reference corpora (Scott, 2015). Used for nominal data dealing with frequencies within categories, or used to ascertain whether categories are linked or independent of each other (see Oakes, 1998: 24-25), chi-square calculations were not used. Log-likelihood was instead chosen as although it can assist analyses of longer texts, it is said to provide "a better estimate of keyness" (Scott, 2015: para 4). In conjunction with a p value set at <0.000001 expressing the certainty with which a word can be considered truly key (Scott, 2015), every word in the B1, B2 and C1 lists had a p value of 0.0000000000 (a likely outcome in such calculations, see Baker, 2006). This meant that it was extremely unlikely that the words in the lists were established as key by chance alone. Comparisons across proficiency level keyness lists was then conducted using manual observations of data, concordance lines and exploration of collocations, colligations and functions, as per the process detailed in RQ2a.

A final point to be explored in the analysis of RQ2b relates to an earlier potential flaw. If thresholds for total occurrence and text occurrence are set too low, there is a

possibility that individual words or texts could skew the keyword lists compiled. As an example, though not discussed in Section 4.6.3, the word *hotel* appeared in the C1 level pilot data as 7th in the keyword list for the entire exam (i.e. not Parts A, B or C). Despite appearing 82 times within the pilot corpus, it only occurred in 4 out of 12 exams. Although its appearance in the list was significantly more frequent than its appearance in the BNC World frequency list, it could be argued that it did not represent a typical feature of successful language use for all C1 candidates examined. For this reason, another statistical measure was introduced, a measure that had not been used in the pilot study. For corpus linguistics, the topic of dispersion can be of equal importance to frequency. Dispersion refers to the occurrence of a particular item within a corpus. Whereas frequency identifies how often it occurred, dispersion can give more indication of its distribution and whether it was typical of all texts in a corpus (see Oakes, 1998). To assist in calculating this measure, Brezina's (2014) online toolbox for corpus linguistics was utilised.

5.3.4 RQ2c) What were the most frequent 3- and 4-word chunks at B1, B2 and C1 and their notable collocations and functions?

Section 3.3.2 in the literature review introduced the concept of lexical chunks. The definition offered from Wray (2002) clearly related to their role in language but methodologically speaking, chunks do not always fall within the typical two to four word parameter described by McCarthy (2010). Lexical chunks are "recurring sequences of *n* words" (McEnery & Hardie, 2012: 110); their length can be determined by the analyses the researcher wishes to make. This study chose Ngrams of three and four words in length. Three-word Ngrams were included as they are extremely common in speech (Biber et al., 1999). Although two-word chunks would be easier to extract, due to their commonality, four-word chunks were added as their extended length may have resulted in greater fluency gains. Whilst less common than three-word chunks (Biber et al., 1999), the longer stretches of lexis were thought to be of more interest for comparisons of B1, B2 and C1. Clearly, four-word chunks would encapsulate some of the three-word Ngrams (see Carter & McCarthy, 2006), for example *what do you*, *do you think*, *you think about*, but the longer stretches would hopefully provide more context as to whether they always combined or whether they joined with other words. Regardless of whether they were complete chunks or not, their appearance in the lists would still suggest that they are

typically retrieved and produced as ‘one large word’. As a final note to definitions offered, for Ngrams, a sequence qualifies as a chunk if it has a frequency of 10 occurrences per million words (Biber et al., 1999). Although this number can be adjusted and can be made of more relevance to larger, general corpus studies, it would have very little impact for this study.

To produce the chunk lists, uncoded text files were uploaded onto *AntConc*. Though CANCODE mark-up was ignored by the programme, terms such as *candidates* and *laughter* did interrupt some chunks. Uncoded textfiles were chosen as a result and were computed for each of the three UCLanESB levels. Manual comparison of the spoken BNC (using the Brigham Young University British National Corpus [BYU-BNC] (n.d.) online resource) and LINDSEI chunks was undertaken once again but comparison here instead prioritised chunk use across learner levels as their resulting frequencies meant that the variety of chunks seen in the larger corpora would not be replicated. In taking this approach, the previously discussed approaches to extracting collocational information and analysing concordance lines were still followed. Given their functional uses in speech (see Section 3.3.2), they would elaborate more on how they were actually used by learners. Though *AntConc* does automatically provide statistical measures for collocation, comparisons with LINDSEI or the spoken BNC were calculated using the online tools created by Brezina (2014). Once most observational and quantitative analyses had been completed, findings were then compared to Biber et al. (1999) and Carter & McCarthy’s (2006) grammars of English to determine whether any similarities or differences could be identified.

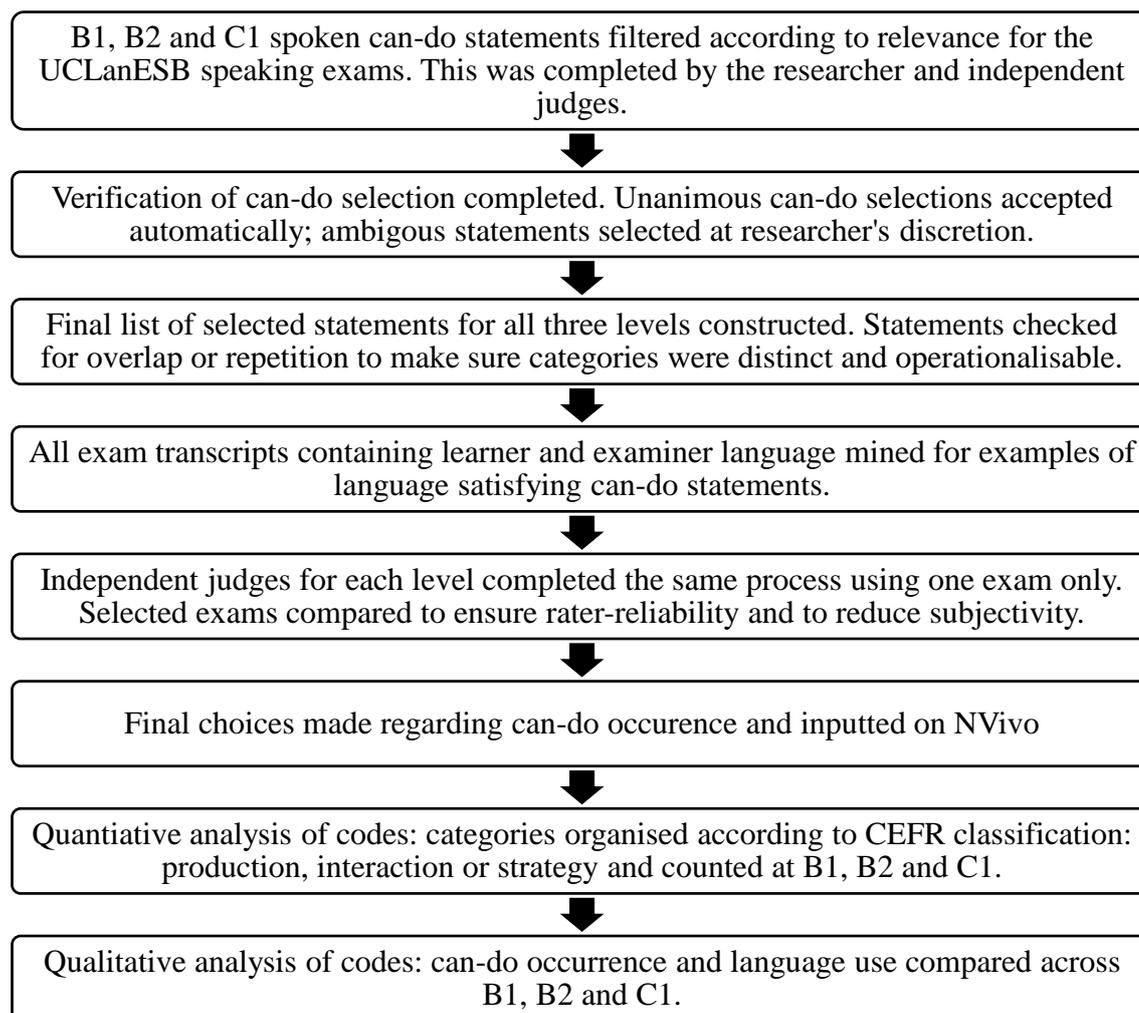
5.3.5 RQ3) What Common European Framework of Reference for Languages (CEFR) indicators are present in terms of spoken interaction, spoken production and strategies at B1, B2 and C1 and how are they realised?

As Section 4.5.3 of the pilot study remarked, data analysis for the third research question required a change in approach. Though the compilation of a corpus resulted in the collection of learner language, corpus linguistic methods were not employed in ascertaining how utterances in learner language did or did not satisfy spoken can-do descriptors in the CEFR. This section therefore documents how language from the corpus was exploited further.

It has been demonstrated that corpus linguistics methods can reveal frequent and significant lexicogrammatical patterns in a particular body of language. However, corpus techniques do have limitations with respect to the rationale underpinning this particular research question. Firstly, qualitative concordance lines can add greater context to colligational and collocational patterns of singular search terms, but they are unable to show how larger structures combine to produce utterances, and how these utterances then combine to fulfil distinct functions during spontaneous interaction. Another problem created by the use of corpus techniques is that delving deeper into the context surrounding particular lexis or structures requires prior knowledge of the items being searched. As Section 3.1.2 demonstrated, if data are to provide a true reflection of real language use, predetermined biases could cloud or overlook aspects of genuine interest to researchers' aims. Of additional pertinence here, is the fact that corpus linguistics methods also predominantly employ quantitative methods of analysis. When such an approach is adopted, there is a risk that "simplified interpretations...distort the bigger picture" (Dornyei, 2007: 39). Though referring to qualitative data in the traditional sense, e.g. responses obtained during interviews, Dornyei's comments clearly resonate with analyses of spoken language use by learners. By investigating specific details such as individual words and chunks, the 'bigger picture' can be missed in terms of how language is used by B1, B2 and C1 learners, and how and when it relates to the can-do statements offered in the CEFR.

Analysis of data for RQ3 required a systematic and robust procedure so that objectivity and reliability could be upheld as much as possible. Since there is no definitive process suitable for all qualitative research (Cohen et al., 2011), analysis ultimately rests on the processes followed and the researcher's own interpretations of data (Dornyei, 2007). Ensuring that illustrations of can-do occurrence across B1, B2 and C1 were not *solely* a product of personal account was therefore essential. Before clarification is given as to how this was achieved, an overview of the process implemented is displayed in Figure 20.

Figure 20: Overview of analytical process for RQ3



Although discussion will elaborate on each of the steps taken during data analysis for RQ3, it is necessary to outline two of the key differences which distinguished traditional qualitative research methodology to that used in this study. A key facet of such research is that making sense of a participant's utterances can occur organically. The use of codes, for instance, emerges from increased familiarity with one or more of the collected texts and thus does not occur prior to the reading of data so as to maintain researcher objectivity (Cohen et al., 2011; Dornyei, 2007; Miles et al., 2014). However, it was clear that this study would not conform to this convention. As learner language was being examined in accordance with CEFR statements of ability, it was clear that codes for analysing data at B1, B2 and C1 had already been predetermined. This study sought not to identify new codes for describing learner speech, but instead it aimed to elaborate on those already defined in an attempt to illustrate them more fully. Secondly, a notable limitation of qualitative research is that it usually focuses on small sample sizes and as

such, produces findings which are difficult to generalise to wider populations (Dornyei, 2007). The sample of 45 exams in this study, 15 at each level, could be deemed small in comparison to the population affected and to other studies such as the EVP and EGP. However, though it is impossible to document or predict all the conceivable structures in a language, it could be the case that particular structures are repeated for certain language functions. The much quoted statement that “language is never ever ever random” (Kilgarriff, 2005: 263) implied that it was possible that certain functions could be realised by B1, B2 and C1 learners using the same, if not similar, lexicogrammatical structures. Though the sample in this study was small, its findings could potentially be considered typical of other learners if this was the case.

Responses to RQ3 were essentially based on the use of codes during analysis. Coding refers to “the ascription of a category label to a piece of data, that is either decided in advance or in response to the data that have been collected” (Cohen et al., 2011: 559). As previously stated, these codes had already been established by the CEFR and would remain largely unchanged during analysis. The first step in Figure 20 therefore centred on the identification of relevant can-do statements that could be used as codes. With the CEFR aiming to be comprehensive and applicable to a range of language learning situations, not all abilities documented within it would be evidenced in the speaking exams. As qualitative research is typically time consuming (Dornyei, 2007), efficiency had to be maximised by searching for only the abilities that were expected to arise. To do this, a list of all spoken CEFR can-do statements at B1, B2 and C1 was made. The researcher read each statement and decided upon its relevance in the exams, categorising each as ‘yes’, ‘no’ or ‘maybe’. Though this process was followed in the pilot study, no independent verification of the selected statements took place; the degree of researcher bias or subjectivity could have been considered high even before analysis of exams occurred. For this reason, independent raters were asked to complete the same procedure (see Cohen et al., 2011) marking statements as relevant, irrelevant or potentially relevant. Unanimous decisions were immediately accepted and the researcher decided upon statements displaying a lack of consensus. In total, five independent raters assisted at this stage. Each had several years’ experience using the exams as either an interlocutor, assessor, senior examiner or examination standards official; they had all completed the required standardisation procedures and each completed a refresher activity via the *CEFRTrain* website (University of Helsinki, n.d.).

The use of can-do statements as codes for categorising chunks of language from the exams revealed an immediate challenge. Upon looking at the statements selected, it was clear that at some levels there was overlap and repetition. For instance, in the CEFR, overall descriptors for spoken production were often broken down into two or three individual descriptors; on other occasions, statements from spoken interaction appeared extremely similar to statements belonging to the strategy group (see Figure 21 below). In qualitative research, attempts should be made to make codes discrete; this can be completed via a systematic process of refining preliminary codes or by a process of changing coding labels into fuller sentences describing actions or patterns (see Cohen et al., 2011; Miles et al., 2014). However, since this study aimed to investigate ways in which can-do statements were realised in learner speech, modification of CEFR terminology was considered unfavourable. The following decisions were taken. Where several descriptors combined to formulate an overall descriptor, the individual categories were chosen; where overall descriptors were not itemised, the researcher broke them down into discrete labels; and when statements were repeated, only one statement was chosen. As a result, 22, 14 and 9 CEFR can-do statements (see Appendix 12) were selected at B1, B2 and C1, respectively, as a basis for coding examinations.

Figure 21: Can-do statement similarity

B1 can-do statements	
Overall oral production (production)	Can reasonably fluently sustain a straightforward description of one of a variety of subjects within his/her field of interest, presenting it as a linear sequence of points.
Sustained monologue (production)	Can give straightforward descriptions on a variety of familiar subjects within his/her field of interest.
	Can reasonably fluently relate a straightforward narrative or description as a linear sequence of points.
B2 can-do statements	
Goal-oriented co-operation (interaction)	Can help along the progress of the work by inviting others to join in, say what they think, etc.
Interaction strategies: Co-operating (Strategy)	Can help the discussion along on familiar ground, confirming comprehension, inviting others in, etc.
C1 can-do statements	
Overall oral production (production)	Can give clear, detailed descriptions and presentations on complex subjects, integrating sub-themes, developing particular points and rounding off with an appropriate conclusion.
Sustained monologue (production)	Can give clear, detailed descriptions of complex subjects.
	Can give elaborate descriptions and narratives, integrating sub-themes, developing particular points and rounding off with an appropriate conclusion.

The next stage involved analysis of exam transcripts according to the finalised set of codes. Researchers should not initiate this step immediately. Instead, they should make themselves familiar with the data by reading transcripts or listening to audio files so as to achieve “a sense of the whole” before assigning codes (Cohen et al., 2011: 556). Such advice presented obstacles in terms of practicalities of the research and validity. Firstly, written exam transcriptions, rather than audio files were chosen for the basis of the analysis. As each exam was 15 minutes long, listening to audio files would not be feasible: exams would have to be listened to for each of the 45 statements and possibly rewound when language had to be deliberated upon. Written transcripts of learner and examiner speech were deemed to be most appropriate. Conversely, some may argue that using written scripts to analyse spoken language denies researchers the opportunity to capture a true ‘sense’ of data. Features such as the rate of speech, pronunciation and intelligibility of candidates are undeniably omitted. However, since the researcher transcribed nearly all of the exams using CANCODE conventions, it was felt that familiarity was already of a high standard and key features of spoken interaction such as pauses, overlaps and false starts had been kept. Analysis of data, level-by-level, by the researcher was conducted following a preliminary read-through of exams to refresh familiarity and via read-throughs of exams for each statement identified at the proficiency level. Although language deemed to match codes was initially highlighted manually, findings were ultimately recorded using *NVivo*, as in the pilot study, to facilitate the recording, comparison and counting of data according to the assigned codes (Cohen et al., 2011; Dornyei, 2007).

The final step in the qualitative analysis of data involved reducing the effects of the researcher upon RQ3 findings. The largest part of analysis was conducted by the researcher; though code selection was reliant upon verification from independent judges, the choice of language deemed to relate to can-do statements was the researcher’s alone. The degree of subjectivity here was therefore high. The optimum approach would have been to have two or more individuals repeating the same process with the same set of exams. However, realistically this would not have been an option due to the vast amount of time required to analyse 15 exams at three levels. An alternative, convenient method for confirming language analysis had to be found. The chosen option was to create a reduced process that would require judges to repeat the procedure followed for a condensed set of codes on one exam; their findings could then be compared to the researcher’s to ensure agreement or distinctions in analysis. A total of 8 statements were

selected at each of the three levels. Though this represented less than a third of B1 statements and nearly all of C1 statements, a higher number of statements would significantly have increased time demands for completing the task. After contacting several individuals to help with this process, three independent judges were identified: one per level. They were instructed to complete the task in Appendix 13 and assigned an exam to read at random. This achieved mixed results. Many of the can-do statement language choices were confirmed; new utterances were highlighted and in some cases, the selected exams showed no evidence of a statement at all. Discussions with the raters, as is advised by Miles et al. (2014: 84), was able to identify “initial difficulties” in interpreting the statements and the language deemed to relate to them. Though codes could not be amended, discussion highlighted once again the struggle experienced in relating the abilities documented in the CEFR to the language demonstrated in real life contexts. To complete analysis, all can-do occurrences was inputted to NVIVO (QSR, 2012). Once statements had been counted, most frequent statements at B1, B2 and C1 for production, interaction, production strategies and interaction strategies were selected for subsequent qualitative comparison.

5.4 Chapter conclusion

This chapter has documented the composition, design rationale and analysis of the UCLanESB learner corpus. As an example of a spoken, learner corpus with a specialised nature, its size is much reduced in comparison with other corpora. However, using external criteria and careful consideration on the part of the corpus compiler, efforts were made to make it as representative as possible of successful B1, B2 and C1 learners at the University of Central Lancashire’s Preston campus. With each level consisting of 15 OPI exams achieving a solid pass score, all texts were included in their entirety and ethical guidelines were contemplated and adhered to at all times. The next chapter thus presents and discusses the findings extracted from the data using the corpus and qualitative analyses documented in the latter stages of the methodology discussion.

6. RESULTS AND DISCUSSION

6.0 Introduction

This chapter presents the results and discussion generated from the data gathered. Since the research questions examined learner success in differing ways, i.e. using vocabulary profiles, frequent and key words and chunks, and analyses of can-do statement occurrence, this chapter will be split into three main sections. To aid discussion, each section will contain results and findings for each main research question; a series of bullet points will also be used to summarise findings before another question is introduced. Section 6.1 will shed light on the results from RQ1's vocabulary profile data, Section 6.2 will reveal the findings from RQ2a, RQ2b and RQ2c, and Section 6.3 will present RQ3 can-do statement occurrence and the language used.

6.1 RQ1: What percentage of the words used by successful B1, B2 and C1 learners come from the first 1000 and second 1000 most frequent words in English?

6.1.1 RQ1 Results

Table 25 displays the vocabulary profile results for K-1 and K-2 bands in the UCLanESB data.

Table 25: Percentage of K1 and K2 words at B1, B2 and C1

Exam Level	Freq. Band	Families (%)	Types (%)	Tokens (%)	Cumul. Token %	Total types across all bands	Total tokens across all bands	Tokens per type	Standardised TTR (<i>WordSmith Tools</i>)
B1	K-1	515 (63.50)	715 (66.88)	13751 (93.33)	93.33	1069	14733	13.78	25.60
	K-2	161 (19.85)	189 (17.68)	483 (3.28)	96.61				
B2	K-1	595 (62.57)	844 (67.09)	17710 (93.51)	93.51	1258	18939	15.05	25.69
	K-2	205 (21.56)	235 (18.68)	748 (3.95)	97.46				
C1	K-1	607 (58.48)	927 (63.71)	19307 (92.25)	92.25	1455	20929	14.38	27.66
	K-2	236 (22.74)	285 (19.59)	976 (4.66)	96.91				

In response to RQ1, a percentage of approximately 97% of words at B1, B2 and C1 originated from the 2000 most frequent words in English. When *tokens* were examined, very little difference was evidenced across the learner levels in this study. The 1000 most frequent words afforded learners at each level a token coverage between 92%

and 93%; the second 1000 most frequent words offered an additional token coverage of 3% to nearly 5%. When combined, Table 25 reveals that these two bands provided a token coverage of approximately 97% meaning that only 1 in every 33 words at B1, B2 and C1 came from a different frequency band. An examination of coverage according to word *families* also supported the finding that there was little difference across the levels. Combined K-1 and K-2 family coverages remained rather stable at 81-84%: B1 had a family coverage of 83.35%, B2's family coverage stood at 84.13% and C1's family coverage decreased slightly to 81.22%. It can be inferred from this data that for learners to be successful and present a solid performance in their speech at the Threshold, Vantage and Effective Operational stages, the two most frequent vocabulary bands in English are essential. Extracts 20-22 below, taken from the B1, B2 and C1 data, demonstrate that successful language at these levels is indeed comprised of vocabulary mostly from the K-1 and K-2 bands (words in blue = K-1, words in red = K-2, words in **bold** denote a different frequency band or an off-list word):

Extract 20: B1, Exam 10

<\$19M> *Er actually the most popular sport in my country is football er I I like football and er I've found national team. Er actually national team played yesterday last night and er er losed the cup silver cup. I'm sad today but er the the you know the football it's a game I think it's help the politics to keep the people of the country happy to keep the people in the country you know er fans watch the TV <\$/=> it's </\$/=> and also the people er happy when they when they watch er the the foot= the football match.*

Extract 21: B2, Exam 14

<\$29F> *Well they look for higher qualifications and it depends actually on the job if you're gonna apply like if it's for sport they will look for sport section <\$/=> if it's for the </\$/=> it depends which major you're gonna you're gonna apply for not major meaning you're gonna apply for the if you're studying business like I'm studying business now I might t= er go to a company and manage the company so it depends on the subject you studied and it depends on the company you're applying to.*

Extract 22: C1, Exam 14

<\$29M> *To learn to drive I think it must be sixteen or seventeen but to start driving it should be eighteen because if people just learn to drive in one month or in one session they won't learn everything so <\$/=> they start </\$/=> two years ago they start having classes some safety some safety classes about the classes about the cars <\$/=> can </\$/=> what they are capable of what they can do what they can't do they must tell them everything before they can use the car.*

Though similarities in combined token and family coverage scores permitted a preliminary response to RQ1 to be gauged, they overlooked the subtle differences

evidenced in the occurrence of K-1 and K-2 words across B1, B2 and C1. It was necessary to analyse the changes in the number of *types* and *tokens* used by learners to see whether learners exhibited usage of a ‘broader’ range of distinct words as proficiency grew. In a sense, this would provide an insight into the number of words ‘known’ or used by learners at each level. The steps taken in identifying potential differences examined i) raw type and token frequencies, and ii) individual family proportions.

Observations of raw type and token frequencies revealed an extended range and usage of K-1 and K-2 words as proficiency levels rose. The third column in Table 26 below shows an incremental use of K-1 and K-2 *types* from B1 to C1. K-1 types increased by 129 ($\approx 18\%$) from B1 to B2, and by 83 ($\approx 10\%$) from B2 to C1. K-2 types increased by 46 (24%) from B1 to B2, and by 50 ($\approx 21\%$) from B2 to C1. When independent t-tests were performed on data from the three exam sets, it was confirmed that B2 learners produced significantly more K-1 and K-2 types than B1 learners ($p < 0.01$ for K-1; $p < 0.05$ for K-2) and that C1 learners produced significantly more K-1 and K-2 types than B1 ($p < 0.01$ for K-1 and K-2) and B2 learners ($p < 0.05$ K-1, and $p < 0.05$ for K-2). This data shows that to be successful at higher levels, a wider range of types within the 2000 most frequent words in English must be evidenced. Similarly, raw *token* frequencies (column 4) displayed a change in the number of words used across levels. Whereas B1 learners produced a total of 13,751 K-1 tokens, B2 generated 17,710 (an increase of $\approx 29\%$) and C1 learners reached 19,307 (an increase of $\approx 40\%$ from B1 and $\approx 9\%$ from B2). K-2 token usage also more than doubled from B1 to C1. When tested for significance, a distinction arose. B2 and C1 students clearly used significantly more K-1 and K-2 tokens than B1 learners ($p < 0.01$ in both instances). C1 learners, on the other hand, did not produce significantly more K-1 or K-2 tokens than B2 candidates. The data implies that learners need to use more individual words at B2 than B1 to be considered successful, but from B2 to C1, the difference rests on alternative factors.

Table 26: Types, tokens, means and SDs for K-1 and K-2

Exam Level	Freq. Level	Types			Tokens		
		Frequency (%)	Mean	SD	Frequency (%)	Mean	SD
B1	K-1 Words:	715 (66.88)	193.13	27.38	13751 (93.33)	833.00	169.90
	K-2 Words	189 (17.68)	19.00	7.15	483 (3.28)	32.20	15.27
B2	K-1 Words:	844 (67.09)	226.93	30.59	17710 (93.51)	1105.47	262.15
	K-2 Words	235 (18.68)	25.53	6.33	748 (3.95)	49.87	13.61
C1	K-1 Words:	927 (63.71)	254.87	42.92	19307 (92.25)	1212.67	333.97
	K-2 Words	285 (19.59)	34.30	10.95	976 (4.66)	65.07	31.31

Taken together, type and token frequencies allowed conclusions to be made regarding success in RQ1. As was presented earlier and has been confirmed here, learners at any of the three levels cannot, quite obviously, be successful without knowledge of the 2000 most frequent words in English. Furthermore, it is apparent that progression from B1 to B2 rests on the utilisation of a broader and more frequent use of lexis from the K-1 and K-2 words categories. However, the stability in token and family coverage, as well as increased type and token numbers directly contradicts Laufer and Nation's (1999) finding that higher-level students use fewer high frequency words. In fact, successful UCLanESB students used more lexis from the 2000 most frequent words as proficiency rose. This finding was also corroborated when the number of K-1 and K-2 word *families* used by B1, B2 and C1 learners were examined. B2 learners produced significantly more K-1 and K-2 families than B1 learners ($p < 0.01$ for K-1, $p < 0.05$ for K-2) and C1 learners produced more K-2 families than B2 learners ($p < 0.05$). To see if there were any underlying differences, therefore, individual, rather than combined, word families were scrutinised more closely to see if percentage gains or losses could be identified across the K-1 and K-2 bands from B1 to C1. It was found that K-1 families decreased by 5.08% from B1 to C1, K-2 families increased by 2.89% and the remaining deficit of 2.19% was recouped in the K-3 to K-6 bands. Though the previous data suggested otherwise, success in learner speech is still subject to an overall fall in word families of highest frequency but unlike Laufer and Nation's observations, which refer to high frequency lexis more generally, this outcome occurred only at the K-1 band in the UCLanESB data.

Following these findings from the vocabulary profiles, one final analysis was performed. Rather than ascertaining the bands which words came from, UCLanESB data was inputted into EVP's *Text Inspector* tool (WebLingua, 2014) to supplement and enhance RQ1's findings. Though undeniably *Text Inspector* is used to analyse written language, it would provide a deeper understanding of the words categorised by the vocabulary profiles and it would inform assumptions by some that lexis increases in 'complexity' as proficiency rises. Table 27 shows the average percentage coverage of words at B1, B2 and C1 according to the different CEFR bands (for a more complete picture of how this fluctuated across the exam parts, see Appendix 10).

Table 27: Text Inspector analysis of B1, B2 and C1 learner speech

UCLanESB data set						
CEFR level	B1	Cum.%	B2	Cum.%	C1	Cum.%
A1	48.59	48.59	45.61	45.61	41.19	41.19
A2	20.98	69.57	22.52	68.13	21.92	63.11
B1	14.79	84.36	15.31	83.44	17.45	80.56
B2	4.06	88.42	5.64	89.08	6.86	87.42
C1	0.99	89.41	0.97	90.05	1.67	89.09
C2	0.65	90.06	0.52	90.57	0.63	89.72
Unlisted	9.95	100.01	9.33	99.9	10.29	100.01
Total	100.01	--	99.9	--	100.01	--

*Figures rounded to 2 decimal places resulted in variance in total figures.

Approximately two thirds of speech at B1, B2 and C1 were composed of lexis from A1 and A2. As can be seen, another noticeable difference across levels is the cumulative percentages for A1, A2 and B2 lexis. Figures for each of these three levels falls across the B1, B2 and C1 UCLanESB data whilst percentages stabilise at B2 and display little variance in the C1, C2 and unlisted categories. In conjunction with UCLanESB data demonstrating that the percentage of K-1 word families falls, the data also indicates that some more basic vocabulary (A1, A2 and B1) does make way for vocabulary from the intermediate levels but that the vast majority would probably be considered more ‘simplistic’.

6.1.2 RQ1 Discussion

The CEFR states that there should be development in vocabulary range from B1 to C1. It does not suggest what this development is in quantitative terms, but instead uses vague expressions such as “enough language to get by, with sufficient vocabulary [B1]”, “has a sufficient range of language [B2]” and “a broad range of language [C1]” (CoE, 2001: 28) to infer that the amount of vocabulary known by learners should grow in some way. RQ1 therefore set out to establish the extent to which B1, B2 and C1 learners used the 2000 most frequent words in English. Whilst vocabulary research can focus on the number of word families, derivations and inflections known to learners (see Section 3.3.1), the literature review demonstrated that such enquiry has pinpointed a common target for all learners: that they learn the “heavy duty vocabulary” (McCarthy, 1999: 4) comprising the 2000 most frequent word families in English. To distance this study from

the NS-NNS dichotomy, and subsequently the comparative fallacy, comparison across B1, B2 and C1 learner speech was conducted. This also sought to determine whether these learners differed in their use of K-1 and K-2 lexis, as had been indicated in previous literature.

According to Laufer and Nation (1999), learners at more advanced levels make less use of high frequency vocabulary than learners at lower levels. Conversely, the UCLanESB learners at the B1, B2 and C1 levels displayed comparable token and family coverages at the K-1 and K-2 bands: the predicted changes at the higher levels did not materialise. When K-1 and K-2 coverages were combined, progression in learner proficiency failed to demonstrate a meaningful fall in the use of high frequency words as combined token coverages at B2 and C1 actually showed an increase from B1. However, such a finding is not unique. In a study conducted by Galaczi and French (2011), it was discovered that frequency profiles remained relatively fixed as learner proficiency increased. Learners equivalent to B1, B2 and C1 displayed coverages of 97.05%, 97.61% and 97.75%, respectively, which demonstrates that the learners in this study were comparable in their fixed coverages across levels. Despite the intuitive appeal of these results, only looking at frequency counts provides a limited picture of language use. Support is thus lent to Galaczi and French's (2011: 160) claim that such quantitative lexical variables alone are unable to "consistently show the lexical improvement in candidate speech". Put simply, whilst K-1 and K-2 words enabled learners at the three levels to be successful, vocabulary profiles alone cannot distinguish how that success varied. Furthermore, a crucial point that needs to be acknowledged is that lexical frequency profiles alone fail to illustrate the full extent of word usage. Though it was possible to create lists of all word types used by UCLanESB learners, no indication could be offered as to how they were employed. The notion of polysemy (see McCarthy, 1990; Schmitt, 2010) is hence neglected in the analysis of this research question as words varying in meaning can be assimilated into the same frequency count. Polysemy relates to the way in which a particular word form can occur in differing contexts and can convey differing meanings, for example *a plant* could refer to a herb or shrub, a nuclear of manufacturing facility, a trap to fool criminals or unsuspecting individuals, or an action of growing a crop, tree or flower from a seed. Though comparisons across levels are highlighted in this discussion, it is clear that words from the most frequent 2000 word bands could convey a much more numerous amount of meanings than is first suggested.

That being said, analysis into the number of types and tokens used did reveal distinctions between the proficiency levels. Since vocabulary, can be quantified according to the “number of words” of which the learners have some knowledge (Qian, 2002: 515), this study has been able to identify that there was a statistically significant rise in the number of K-1 and K-2 types used not only from B1 to B2, but from B2 to C1 also. The study also established that token numbers used increased significantly between B1 and B2, but not from B2 to C1. Relating the successful speech in this research to the CEFR descriptors adds some clarification as to how the transitions from B1’s ‘enough language to get by, with sufficient vocabulary’ to B2’s ‘sufficient range of language’ to C1’s ‘broad range of language’ exemplify themselves. After B1, it is through the number of different words (types) ‘known’ or used that learner proficiency increases, rather than the number of words (tokens) employed. Pertinent to mention here, in discussion of the number of words used, is the matter of lexical diversity. Table 25 in Section 6.1.1 presented a standardised type-token ratio [TTR] at each of the three levels. Standardised TTRs reflect the lexical diversity in speech according to the number of type and tokens and the size of the data sample (Götz, 2013); the higher the score, the more varied the lexis (Biber et al, 1999). Since TTRs in conversation are said to approximate to 30% (in samples of 1000s of words, Biber et al., 1999) and since the standardised TTR for LINDSEI stands at 29.42%, it is clear that UCLanESB data at all three levels reflected the typical nature of speech and features of spoken grammar. Due to the real-time, spontaneous nature of speech, TTRs are expected to be lower as the repetition and recycling of vocabulary can aid planning, emphasis, and the communication of messages (Biber et al, 1999; Carter & McCarthy, 2006; Nation, 2001; Nation & Waring, 1997). The appearance of repetition and a relatively low variety of words should thus not be penalised. Though learners are “characterized” by lower levels of lexical diversity (Götz, 2013: 12), repetition of word types can actually be a sign of success since it can help alleviate some of the demands placed on learners during spontaneous, unplanned speech.

Finally, this last section will focus on the shifting characteristics of learner vocabulary both with respect to word family percentages across levels and their relation to CEFR levels. Previous discussion highlighted that in contrast to literature stating that low frequency words reduce as proficiency grows, the UCLanESB learners remained relatively similar in terms of combined token and family coverages; their use of types from the K-1 and K-2 bands significantly enlarged also. With regards to individual coverages for K-1 and K-2 bands across B1, B2 and C1, a difference emerged: the most

frequent 1000 words fell in percentage by 5% from B1 to C1, a percentage which was regained at K-2 and above. Laufer and Nation's (1999) finding was earlier disputed by the UCLanESB data but it in fact has some weighting where K-1 lexis is concerned. The learners did reduce their use of words from this band from B1 to C1 but to be successful, learners still employed very high numbers of high-frequency words. Though still lower than the NS threshold of 96% for K-1 and K-2 families in speech (Adolphs & Schmitt, 2004; Schonell et al., 1956 cited by Adolphs & Schmitt, 2003) the 2000 most frequent words are evidently crucial to success. Put simply, just because a learner reaches higher levels of proficiency should not mean that they have to or should be expected to produce less frequent lexis. This conclusion, explains the analysis performed using *Text Inspector*. Personal experience as a teacher and assessor has resulted in many encounters in which teachers and assessors, native or not, give 'extra credit' for 'lexical gems' in learner output or, more generally, place higher expectations upon their learners in terms of the vocabulary they should produce. However, two-thirds of learner speech at B1, B2 and C1 comprised lexis from the A1 and A2 bands according to the EVP. Such lexis will be of a more frequent nature but it is the researcher's opinion that judgements of 'success' at higher levels should not become confused with a need to produce more complex vocabulary. To be successful in speech, learners are said to avoid 'difficult' words due to their effect on processing effort (Götz, 2013) but similarly, difficult words may not be evidenced simply because they are not always required by learners. In unplanned speech, research has demonstrated that NSs simplify their speech so as to reduce processing demands. It would seem that in some circles, a type of 'double-standard' exists as this strategy is not extended to learners. In fact, a less target-like performance is sometimes expected. Therefore, it should be remembered that "a fundamental measure of improving lexical proficiency is *how well* the words are used, rather than if they are of lower frequency" (Galaczi & French, 2011: 160).

6.1.3 RQ1 Summary

The results for RQ1 can be summarised as follows:

- B1, B2 and C1 learners were comparable in their combined K-1 and K-2 token and family coverages. K-1 and K-2 tokens stood at 97% whereas family coverage fell between 81-84%. Less than 1 in 33 tokens came from another band.

- Learners did not use less frequent vocabulary, according to K-1 and K-2 bands, as proficiency developed so the 2000 most frequent words are therefore fundamental to success.
- A rise in proficiency was characterised by higher numbers of word types. Corpora cannot be representative of a learner's full language knowledge but the number of words used did increase.
- The number of words used, i.e. tokens, increased significantly from B1 to B2. It was not, however, a distinguishing feature of the changes from B2 to C1.
- The percentage of K-1 words did fall from B1 to C1; it was not felt that this fully corroborated claims that frequent vocabulary amounts in learner speech would fall more generally.
- A high amount of less difficult vocabulary was exhibited by successful learners. Higher levels of proficiency do not necessarily equate to the use of more complex vocabulary in learner speech so expectations placed on learners should be realistic and acknowledging of previous corpus findings for speech.
- It must be remembered that the vocabulary profiles were not demonstrative of the polysemic nature of individual words in this data. They allowed for comparisons of types and token usage, but they nevertheless revealed little about the way words were used.

6.2 RQ2a: What were the 20 most frequent words at B1, B2 and C1 and their notable collocations, colligations and functions?

The 20 most frequent words occupied a large portion of the total corpus at each of the three levels. When percentages for these words were combined, it was discovered that at B1, the words comprised 44.57%; at B2, the words comprised 41.09% and at C1, the words comprised 39.36% of their respective corpora. Compared to the cumulative percentage of LINDSEI's top 20 words (see Appendix 11), which stood at 38.01%, it is noticeable that UCLanESB students relied a little more heavily on the words in the tables below than the learners in LINDSEI. However, with more than two thirds of the top 20 words at B1, B2 and C1 (70%, 80%, 85%) corresponding to LINDSEI's top 20 list, there is evidence that being successful in spoken interaction still necessitates knowledge of the words identified. This result could also be interpreted differently. The combined UCLanESB sub-corpus proportions clearly fell as proficiency grew. Learners relied on

these frequent words to form around two fifths of their speech, but the decreasing percentages showed that dependency on these words lessened: transitions from B1 to B2 and from B2 to C1 resulted in increasing usage of word types beyond the 20 most frequent threshold.

Table 28: 20 most frequent words at B1, B2 and C1

N	B1					B2					C1				
	Word	Freq.	% of B1 corpus	Texts	% Range	Word	Freq.	% of B2 corpus	Texts	% Range	Word	Freq.	% of C1 corpus	Texts	% Range
1	ER	1,220	8.4	15	100	ER	1,220	6.48	15	100	THE	1,071	5.19	15	100
2	I	943	6.49	15	100	I	808	4.29	15	100	ER	733	3.55	15	100
3	THE	592	4.08	15	100	THE	775	4.12	15	100	I	718	3.48	15	100
4	AND	454	3.13	15	100	AND	517	2.75	15	100	AND	590	2.86	15	100
5	TO	387	2.66	15	100	TO	501	2.66	15	100	TO	581	2.81	15	100
6	ERM	300	2.07	15	100	YOU	454	2.41	15	100	ERM	450	2.18	15	100
7	IN	249	1.71	15	100	IN	334	1.77	15	100	IS	377	1.83	15	100
8	A	244	1.68	15	100	YEAH	307	1.63	15	100	IN	369	1.79	15	100
9	IS	236	1.62	15	100	A	289	1.53	15	100	YOU	360	1.74	15	100
10	YOU	227	1.56	15	100	ERM	280	1.49	15	100	YEAH	354	1.72	15	100
11	MY	222	1.53	15	100	IS	278	1.48	14	93.33	THINK	313	1.52	15	100
12	THINK	203	1.4	15	100	THINK	263	1.4	15	100	A	309	1.5	15	100
13	LIKE	190	1.31	15	100	LIKE	259	1.38	15	100	LIKE	302	1.46	15	100
14	SO	171	1.18	14	93.33	SO	231	1.23	15	100	OF	281	1.36	15	100
15	BECAUSE	168	1.16	15	100	MY	222	1.18	15	100	SO	255	1.24	15	100
16	YEAH	140	0.96	14	93.33	BUT	203	1.08	15	100	THEY	250	1.21	14	93.33
17	OF	139	0.96	15	100	THEY	201	1.07	14	93.33	IT'S	234	1.13	15	100
18	CAN	132	0.91	13	86.67	OF	199	1.06	15	100	IT	230	1.11	15	100
19	IT'S	132	0.91	14	93.33	IT	197	1.05	15	100	THAT	174	0.84	15	100
20	WE	124	0.85	14	93.33	HAVE	194	1.03	15	100	BECAUSE	173	0.84	15	100

Attention, following this preliminary overview, will now turn to the individual words which provoked further investigation. Results have been summarised using bullet points but will be unified and linked to success in each word’s discussion. The words explored more deeply are: *we*, *er* and *erm*.

6.2.1. *We*

- Initially appeared to be a feature more typical of B1; it was not in the top 20 most frequent words at B2 or C1
- Normalised frequencies (see Table 29) were however similar in the UCLanESB and LINDSEI data, though there was a dramatic fall in the Spoken BNC data
- 19 common and significant collates were found across B1, B2 and C1 (see Table 30)
- Differences in function identified; sometimes differences correlated with collocates not shared across levels

Table 29: Normalised frequencies for *we*

Word	UCLanESB									LINDSEI			BNC Spoken		
	B1			B2			C1			List Pn	Freq.	NF	List Pn	Freq.	NF
WE	20	124	85.358	23	171	90.856	23	165	79.969	26	5773	72.801	13	10448	10.448

*red = beyond 20 most frequent word boundary

Table 30: Collocates of *we* at B1, B2 and C1

B1				B2				C1			
	Freq. with node	T-score	MI score		Freq. with node	T-score	MI score		Freq. with node	T-score	MI score
Can	39	6.065	5.113	Have	39	5.963	4.469	Have	44	6.43	5.025
Have	25	4.816	4.761	So	33	5.379	3.976	So	27	4.804	3.727
Should	18	4.188	6.287	Can	31	5.263	4.19	Should	27	5.113	5.967
So	17	3.769	3.542	Think	21	4.061	3.136	Know	25	4.858	5.135
If	16	3.878	5.039	Like	21	4.069	3.159	Can	24	4.657	4.342
Is	15	3.353	2.897	Know	19	4.186	4.656	Because	22	4.396	3.992
Think	12	2.964	2.792	Are	17	3.998	5.038	If	19	4.164	4.487
Because	11	2.884	2.939	Need	17	4.042	5.661	Think	14	3.073	2.484
Are	11	3.206	4.905	Will	16	3.719	3.829	Is	13	2.77	2.109
Need	10	3.116	6.107	What	16	3.798	4.307	Don't	13	3.415	4.241
Different	9	2.92	5.235	If	15	3.568	3.668	Like	11	2.589	2.188
Know	8	2.584	3.532	Because	15	3.491	3.341	Are	10	2.826	3.233
Don't	8	2.569	3.446	Do	15	3.575	3.701	Need	10	3.097	5.588
Go	6	2.213	3.37	Go	15	3.65	4.12	Do	9	2.731	3.479
Like	3	0.796	0.887	Should	14	3.691	6.198	Can't	8	2.721	4.719
Do	3	1.387	2.328	Don't	14	3.531	4.147	Will	7	2.205	2.584
Will	3	1.427	2.503	Can't	10	3.105	5.783	Go	5	1.932	2.879
Can't	3	1.619	3.934	Is	9	2.159	1.834	Different	3	1.543	3.194
What	0	0	0	Different	0	0	0	What	0	0	0

*Words in red below threshold for significance

**Words in bold were common and had significant t- and MI scores across all three levels

Elicited functions of *we* as a subject pronoun across B1, B2 and C1 matched three distinct purposes identified by Carter & McCarthy (2006: 379). Though they are learnt at B1 level according to EVP (CUP, 2015a), they were still representative of speech at all three levels. The functions included:

- *Specific* – Inclusive of speakers in immediate context
- *Third party* – Exclusive as it refers to speakers and persons absent from discourse
- *General* – Reference to larger groups of people e.g. society or people in general

Part A of the exam revealed that:

- B1 (71%) and B2 (77%) saw the majority of usage in the third party category (see Extracts 23 and 24). Usage was attributed to familiarity of questions such as *can you tell me who you get on best with in your family?* and *what do you do when you go out with your friends?* Prevalence of *go*, *don't*, and *like* in collocates list for B1 and B2 linked to such usage
- C1 saw *third party* usage occupy 41% of all occurrences of *we* but this was superseded by general usage (59%). Again questions triggered such a result.

Though familiar questions were posed, questions also incorporated the more complex topics alluded to in the CEFR, e.g. *what are the major transport problems in your country?* (See Extract 25).

Extract 23: B1, Exam 9

<\$0> *Who do you most like to spend time with?*

<\$17F> *Okay I like to spend my time or all my free time with my friends.*

<\$0> *Mhm.*

<\$17F> *Er we go to shopping and er we sometimes go out to eat out restaurant. Yeah we have er nice time <\$O26> when </\$O26> we together.*

Extract 24: B2, Exam 13

<\$0> *+ how long have you been studying English?*

<\$25M> *I have been er studying English er eight months. Er I started er in August er no sorry in October last year 2012. Er er you can say in er when I was in school I sta= I start from the business English because we don't er use it English in our country. We use Arabic language. So that's it.*

Extract 25: C1, Exam 8

<\$0> *The most environmentally friendly way to travel? Do you understand that?*

<\$16F> *The most er invently +*

<\$0> *Environmentally friendly.*

<\$16F> <\$O17> *Oh </\$O17> Yeah I got the meaning.*

<\$15F> <\$O17> *Environment </\$O17>*

<\$0> *Mm.*

<\$16F> *Mm I answer the question?*

<\$0> *Yeah.*

<\$16F> *In the travel I think the environment friendly it means we go to travel not by a car and it not er mm we can protect er environment and that is er mm er green green travel yes erm green travel when we go to travel I think if we use bicycle I think it's not harmful the er fresh air and we can't throw the rubbish anywhere I think it's er mm friendly environment.*

Part B revealed:

- A range of differences across the levels
- B1 learners began using *we* in a more general sense (53%) whilst third party usage fell to 42%. Learners applied statements of personal interest or relevance to everyday life, an ability learnt at B1 according to EGP (CUP, 2015b)
- B2 learners did not demonstrate a clear majority in either of the three functions. We distributed between the specific (44%) and general (49%) categories. Learners related discussion to those involved in the interaction and to wider contexts

- C1 learners used generic *we* in 74% of their utterances. Less familiar and more diverse topics were considered responsible for the spike generated when this figure was compared to B1 and B2
- *We* for specific meanings uncovered another use at the B2 and C1 levels. Moving on from the content of statements to the ability to confirm details and manage discourse, there were small numbers of strategies in the data (see Figure 22) (7 at B2 and 6 at C1)
- *We* enabled learners to express thoughts, manage discourse and clarify meaning

Figure 22: Usage of *we* for confirmation and discourse management at B2 and C1

B2	C1
• Shall we begin? (Exam 12)	• And why should we choose two? (Exam 14)
• So do we discuss the two topics or the first and then the second? (Exam 14)	• Should we discuss it with you? (Exam 13)
• So what do you think about it if we just refer to these two words? (Exam 4)	• Yes yeah and er shall we focus on the other element? (Exam 6)
• We should start the topic at first (Exam 12)	• So we have to speak to each other? (Exam 14)
• Can we start? (Exam 6)	• How many minu= minutes can we? (Exam 3)
• Er should we start? (Exam 15)	• For me can we start? (Exam 6)
• Er shall we start? (Exam 7)	

Part C revealed:

- A uniform use of *we* dominated at all three levels: the use of *we* to create general meaning
- Part C represented the ‘probe’ stage of the exam so unsurprisingly, numbers did increase here to occupy two thirds of learner uses of *we*: B1 (67%), B2 (74%) and C1 (68%)
- Collocation data (Figures 23-25) displayed the greatest differences in this exam part with respect to modal and auxiliary verbs in R1 position:
 - At B1, general usage was expressed largely through combinations with *can* (evidenced on 17 occasions). The frequency of *can* and MI and T-scores were highest at this level. Other forms utilised included *couldn't*, *don't*, *need to*, and *should* but these were not used as frequently. When used in such a way, *we* often meant people, language learners or students in general:

Figure 23: Modal and auxiliary verbs in R1 position in Part C at B1

N	Concordance
1	we can erm like improve and they language or we can er know how how they er m= we can
2	Oh so in English + Yeah I know er the summary. We can have a different er + We have a thought
3	with you because it's the most popular and we can like er you know er er see a new food
4	er the same with you er we can remember we can remember what happened at that time
5	er if if they keep a diary eve= every day er we can er we can memorise some significant
6	if they keep a diary eve= every day er we can er we can memorise some significant er er things
7	or we can er know how how they er m= we can make friendly with them. Yes. What
8	because er er if we start a business maybe we can get a lot of money but we have we are
9	er is very good for us and erm I mean we can if we want to change the new work er
10	also can Oh yes because I think er the together we can feel happy always all the time I think so
11	meet er a good new people? Yes I I think er we can erm like improve and they language or
12	to us because er after many years er we can er we can memorise er the the days the days
13	because er we don't need the cooker because we can buy the food fast food or er food from
14	very important because er the same with you er we can remember we can remember what
15	if we ha= if we forget some something we also we can also erm remember through it. The
16	could help me to do solve the problem and we can help each other mm but others are also
17	to our to us because er after many years er we can er we can memorise er the the days
18	in Youtube yes because in Korea we can c= we couldn't get music for free so I go to
19	many strangers around us erm er some people we don't know but we have seen each other
20	Yes. It's er the most useful because er we don't need the cooker because we can buy
21	Erm er we have to begin? I start or you? You you.
22	us erm er some people we don't know but we have seen each other mm when I came to a
23	I think I think is very important so we we need learn some some skill we er another
24	friend er I think er have many challenge and er we should to er do something we never done
25	but it's not the er the area for the work they we should when the organisation applic= recruit
26	like you have lots of experience so I think er we should er get a lot of new experience. Yeah

- At B2 and C1, there was a slightly broader range of modal and auxiliary verbs in R1 position despite a fall in their frequency. These allowed learners to refer to societies, particular nationalities, students and people in general according to the topics discussions. Modal and auxiliary verbs such as *can't*, *didn't*, *don't*, *have*, *have to*, *need to*, *should*, *shouldn't* and *will* were employed

Figure 24: Modal and auxiliary verbs in R1 position in Part C at B2

N	Concordance
1	winter again. Mm because in in that reason erm we can I like snow because in province that
2	Mhm. Yes. Yes. Er it's er our local er traditional. We can say er we have erm special food Arabic
3	weather tomorrow + + we will know what things we can plan + + mm and er if er what if we will
4	if we will have party er if tomorrow is er Sunday we can have the party outside and er if mm
5	these things they they don't have the history so we cannot do this. Erm one thing I think is erm
6	which is er last you can say er forty years ago. We didn't use the like this building say this
7	because you know er er next er hundred years we don't know actually we don't know what's
8	er next er hundred years we don't know actually we don't know what's what will happen. But I
9	music. Mm to me I I think that mm in the shop we don't use the music because we just want
10	know. By media by news by news technology so we have to learn from er to l= to learn er what
11	know what's what will happen. But I think er we have to to be ready for anything you know
12	things items and er you know or sometimes the we need to stay alone you know maybe er to
13	erm you know we are we are young yes so erm we need do a lot of experience of my erm
14	see the snow. Mm? Er mm I like that because if we need er know the weather tomorrow + +
15	. I think during this er during this game we will make new friends and er know about
16	good. Yes. Er fresh air fresh air erm also erm we will say here it's good for our children they
17	if we need er know the weather tomorrow + + we will know what things we can plan + + mm
18	things we can plan + + mm and er if er what if we will have party er if tomorrow is er Sunday

Figure 25: Modal and auxiliary verbs in R1 position in Part C at C1

N	Concordance
1	deserts and beaches and yeah so on so we are really working on improving the tourism
2	students if we know the foreign language we can communication with the others er who
3	. Yeah it's very important in Qatar because now we can before it wasn't but now it's very
4	learning English is also useful for the future job we can work job in global companies I think
5	we ca= we don't have to repeat them and we can learn from them. Er I agree with you on
6	should erm we should do our best such as we can go to work or go school by public
7	to keep the environment because we can erm we usually see the weather
8	others er who are come from other countries we can know their culture and can make
9	people will think yeah it's a higher price maybe we can go next time and also they it's very
10	water and er air is the is the the w= any life we can't live without the water and the air if
11	pollution b= er if they're big po= pollution + + we can't erm it's it's can't imagine it's can't
12	Erm and I think if we know the mistakes we ca= we don't have to repeat them and we can learn
13	options about like good food or may= maybe we don't know all the options. Yeah promote
14	in it. The train there must be one to manage it we have to stop in emergencies people to
15	about this question I think mm we have the erm we need to do this the government should do
16	imagine it's can't imagine erm so if er so I think we should er the governments and individuals
17	many animals in the world no people and now we should take the responsible to protect them
18	people so when we communicate with others we should listen what er what people said so if
19	difficult because when we learn a new language we should mm learn the how to writing and er
20	to pollution the air and the individual should erm we should do our best such as we can go to
21	Oh. Oh. So y= so you mean erm you mean we we shouldn't to change the countryside. Can we
22	I think it will get better. Erm by we teenagers we will improve things that we're seeing going

- Collocational differences established one change between B1 and B2-C1 learners but B1 learners were similarly distinguished once again according to the use of *we* for third parties.
- Carter and McCarthy's (2006) third party category was found to be sub-divided in the UCLanESB data: third party either referred to friends or people specific to learners' own lives, or a broader group of people to which the listener(s) did not belong
- 3 out of 52 (6%) of B1 instances of *we* referred to a third party: frequency was small but all three uses referred to friends or people specific to their lives (see Extract 26)
- At B2 (6, 13%) and C1 (20, 29%), *we* was used in the second instance to emphasise exclusivity on a wider scale (see Extracts 27 and 28). In these extracts, *we* could be considered to share parallels with general usage. However, the speakers identified a distinction: that they were part of the group whilst the listener and other members of society were not

Extract 26: B1, Exam 14

<\$27F> I <\$E> laughs </\$E> I I never record some video but I always download some video er for for for example such as and **we** sings songs and **we** er sing +

Extract 27: B2, Exam 13

<\$27F> I don't know <\$E sighs /\$E> I agree with er these guys but I need to add some little bit from my culture or my traditional. I think in Saudi Arabia <\$=> **we** </\$=> our culture rely on the religion first of all. Everything. All all our life start with the religion everything **we** follow the religion what is the holy Qur'an said and **we** follow it. So **we** have the similar the similar culture with all the Arab countries because **we** have all the same language and the same traditional and the same erm colouring you can say + <\$26M> Skin.

<\$27F> + so we're we're very very similar to them yeah.

Extract 28: C1, Exam 13

<\$0> Okay <\$E> laughs </\$E> Erm is national history or international history more interesting for you <\$26F>?

<\$26F> For me? Erm I'd have to say national history because from where I come from **we** have the pyramids of course because I'm from Egypt. Erm **we** **we** learnt about pharaohs and hieroglyphics and how they used to read and write how they used to live in the desert <\$O41> yeah + </\$O41>

6.2.2 Discussion of the frequent word *we*

Analysing the functions of the personal pronoun *we* uncovered the subtle influences it had on success at B1, B2 and C1. The CEFR states that the nature of the topics discussed evolves across the levels from those of “familiar, or personal interest” at B1, to “a wide range of subjects related to [a learner’s] field of interest” at B2, and finally to those deemed to be “complex” at C1 (CoE, 2001: 26-27). *We* demonstrated how such changes occurred across levels, but also how they occasionally altered within levels. At B1 and B2, third party usage was most prevalent; at C1, general usage was most common. Changes at B1 from third party to general usage also showed that being successful at this level necessitated an ability to expand topics of familiarity to broader contexts, an ability documented in the CEFR (CoE, 2001: 26) when it states that B1 students can “enter unprepared into conversation on topics that are familiar, of personal interest or pertinent to everyday life.” The capacity to extend statements of personal preference or habits, which did not involve the listener, to those relating to people in a general sense or ‘everyday life’ were therefore exhibited through a simple change in the usage of *we*. At B2, specific usage of *we* was nearly joint-highest in Part B. Whereas B1 learners amended topics to refer to more general issues, B2 learners demonstrated an ability to take topics of a more general nature and turn them into topics of direct relevance for the speakers involved. A can-do statement for oral production at B2 (CoE, 2001: 28) describes the capability of giving “clear, detailed descriptions...on a wide range of subjects related to *his/her field of interest*” [emphasis added]. In a literal sense, the use of *we* to signify the speakers and listeners involved evidenced an ability to take a generic subject and highlight its specific relevance for the speakers. In a sense, the influence of individuals on the jointly-constructed skill of speaking became ever apparent at this level as learners were more aware of their partner and tried to produce statements of relevance to them both. Being successful at B2, and progressing from B1, called for the ability to relate topics not only to personal or general situations, but to situations shared by speakers partaking in the discourse.

We also enabled another function relevant to success to be performed. Specific meanings expressed using *we* shed light on the strategic importance the pronoun held for managing interaction. The CEFR highlights the ability at B2 and C1 to understand NSs, although a distinction is made between conversing with a NS in “noisy” environments at B2, and confirming details given by an NS “especially if the accent is unfamiliar” at C1

(CoE, 2001: 75). The questions in Figure 22 show that at B2, emphasis was placed on ‘kick-starting’ discussion by involving others whereas at C1, clarifying task instructions and details from the NS interlocutor were prioritised. The use of *we* clearly afforded learners some flexibility not only in expressing their opinions, but in managing discourse and confirming comprehension. Since other can-do descriptors highlight the ability of B2 and C1 learners to “initiate” discourse and to “get clarification” (CoE, 2001: 86-87), it is clear that the seemingly unremarkable pronoun *we* performed multiple roles in successful learner speech at all levels.

In terms of success, the analysis of *we* on its own merits may not initially have seemed worthwhile. As a personal pronoun, its prevalence in conversation would already have been expected to be high (Biber et al., 1999), but what it has revealed is of importance to learner speech. As an example of a deictic word, orienting the speaker and listener to the discourse and context surrounding them (see Carter & McCarthy, 2006), it provided learners a means to refer to people involved in the discourse, to people exclusive to their own lives and interests, and to people or society in general. While *we*’s deictic function is often specific to the immediate contexts of the speakers and listeners, the data also showed that it allowed B2 and C1 learners to differentiate themselves from others or accentuate an exclusive group to which the listener did not belong; it created exclusion as much as it created inclusivity. For this reason, *we* fulfilled an important pragmatic function which could be adapted as required. In addition to offering some elaboration as to how subject matter can change and the strategic role it realised across proficiency levels, collocational data was also able to demonstrate the potential range of auxiliary verbs it combined with. With B1 learners said to have sufficient vocabulary, B2 learners said to have a sufficient *range* of vocabulary and C1 learners said to have a *broad* range of vocabulary, personal pronouns such as *we* may not immediately convey how such change is realised. It has offered some evidence to suggest that B1 learners may ‘cling’ to lexical or collocational “teddybears” (see Götz, 2013: 30; Hasselgren, 1994: 237) such as *can* whilst B2 and C1 learners ‘branch out’ a little more albeit though collocations are still of high frequency. However, closer examination of *we*’s collocates has still been able to show the general flexibility it affords learners as proficiency increases.

6.2.3 *Er/Erm*

- *Er* occupied first and second position in the B1, B2 and C1 reference lists
- *Erm* was 6th most frequent word at B1 and C1, and 10th most frequent at B2
- The percentage coverage and frequencies of *er* gradually fell as proficiency increased
- *Erm* showed no such pattern. It fluctuated between 1.5-2.2% of the B1, B2 and C1 corpora
- Normalised frequencies of *er* and *erm* were comparable, albeit higher, than in LINDSEI data but they exhibited much greater differences with the BNC spoken data (see Table 31)

Table 31: Descriptive statistics for *er* and *erm*

ER							
Level	Freq. list position	Freq.	%	Normalised freq.	Mean	SD	Juillard's D
B1	1	1220	8.4	839.82	81.33	25.74	0.92
B2	1	1220	6.48	648.21	81.33	43.57	0.86
C1	2	733	3.55	355.26	48.87	30.80	0.83
LINDSEI	4	23925	3.02	301.71	--	--	--
BNC SPOKEN	17	88188	0.89	88.51	--	--	--
ERM							
Level	Freq. list position	Freq.	%	Normalised freq.	Mean	SD	Juillard's D
B1	6	300	2.07	206.51	20.00	11.50	0.95
B2	10	280	1.49	148.77	18.67	16.16	0.77
C1	6	450	2.18	218.10	30.00	30.73	0.73
LINDSEI	11	10354	1.31	130.57	--	--	--
BNC SPOKEN	27	62086	0.62	62.31	--	--	--

- T-tests revealed that frequency changes for *erm* across the levels and exam parts (e.g. Part A of B1 compared with Part A of B2) showed no statistically significant differences. However, with higher standard deviation scores at C1 and a drop in dispersion between B1 and C1, individual exams may have skewed data slightly. For example, one exam at C1 accounted for 28% of all uses of *erm* alone.

- T-tests revealed that frequency changes for *er*, on the other hand, showed one key difference. Overall test figures uncovered that B1 ($p < 0.01$) and B2 ($p < 0.05$) learners used *er* significantly more than C1 learners
- This change was attributed to Part B of the exam which by far saw the biggest frequency changes: mean uses of *er* for B1, B2 and C1 decreased from 34, to 33 and finally to 18 per exam.
- Significance was nearly reached ($p = 0.57$) in Part C of the exam between B2 and C1 with means decreasing from 30 uses per exam to 17. This cannot entirely account for the difference seen overall across B1, B2 and C1

Examination of collocational information, particularly in L1 position, also revealed an interesting pattern among the data for *er* and *erm*. After calculating t-scores and MI scores, words presenting higher in the lists for B1, B2 and C1 related to the same linguistic function: that of creating cohesion within utterances via coordinating and subordinating conjunctions (see Table 32).

Table 32: Collocates of *er* and *erm*

ER											
Collocate	B1			B2			C1			BNC	LINDSEI
	L1 freq.	T score	MI score	L1 freq.	T score	MI score	L1 freq.	T score	MI score	MI score	MI score
AND	169	18.976	3.505	165	18.912	3.655	78	14.348	3.553	1.806	3.306
BECAUSE	33	10.213	3.215	27	9.990	3.506	13	7.232	3.381	1.193	3.273
BUT	13	6.256	2.917	37	9.855	3.214	10	5.589	3.104	2.098	3.170
ERM	30	13.283	3.152	18	10.058	2.896	15	10.575	3.151	0.423	3.067
SO	12	7.404	2.496	10	8.504	2.740	16	7.615	3.050	0.403	3.108
THINK	46	12.547	3.478	39	11.604	3.293	16	9.481	3.320	-0.300	3.024
ERM											
Collocate	B1			B2			C1			BNC	LINDSEI
	L1 freq.	T score	MI score	L1 freq.	T score	MI score	L1 freq.	T score	MI score	MI score	MI score
AND	29	8.020	3.129	9	6.969	3.035	21	8.251	2.838	1.016	3.246
BECAUSE	13	5.602	3.453	6	4.290	3.246	20	6.853	3.840	1.301	3.217
BUT	2	3.577	3.240	9	5.587	3.616	6	2.813	2.186	1.973	3.180
ER	32	13.283	3.152	12	10.058	2.896	15	10.575	3.151	-0.426	3.067
SO	2	2.798	1.987	7	4.850	3.127	7	4.976	2.654	0.681	3.110
THINK	13	6.630	3.633	13	7.240	3.940	18	8.992	3.784	-0.403	3.131

*Words in red fall beneath threshold for significance

- UCLanESB collocational data differed greatly in this analysis from the NS BNC corpus but it was comparable to MI scores computed using LINDSEI. Such collocations were typical of this type of interaction

- Frequency and evidence of *and er* fell despite a relatively stable MI. *And erm* was not a strong collocation at C1 but at B1 and B2, the words commonly co-occurred
- *Because er* or *because erm* often co-occurred. Though evidence of *because er* fell from B1 to C1, *erm because* increased
- *But er* was not a noticeable feature of B1 speech however, there was a lot of evidence of this at B2. *But erm* was not a significant collocation at C1
- B1 learners showed highest evidence and frequency for *er erm* or *erm er*. Such a collocation could have emphasised hesitation in learner speech. Though this was not a significant collocation at B2, it did reappear at C1
- *So er* was only significant at C1. *So erm* did co-occur at B2.
- Though not a conjunction, the verb *think* was included in the collocational analysis. There was ample evidence to show that *think* appeared with *er* and *erm*. As Extract 29 shows, learners might have used it as a less obvious hesitation device

Extract 29: B2, Exam 15

<\$0> *Okay is that your opinion as well <\$32M>?*

<\$32M> *Yeah I think er if er if er the older people er find help from er his sons or something like that or for young people that er use er smartphone I think er it's er become because older people have experience to to deal with the er any device I think nowadays*
+

6.2.4 Discussion of the frequent words *er* and *erm*

The pilot study established that *er* and *erm* were very much a feature of C1 learner speech. Though sometimes viewed as a performance error or deficiency in speech, they were shown to act as a brief planning device, often preceding answers to questions for learners to quickly prepare their utterances. However, whilst intrinsically linked to spoken grammar as features of spontaneous speech (see Section 3.2), their implications for better or worse are varied.

The CEFR remarks that fluency increases across the levels. Though this discussion by no means equates fluency solely to the use of filled pauses like *er* and *erm* (see Gilquin & De Cock, 2013; Götz, 2013 for more in-depth treatments of fluency), the CEFR does explicitly stress that noticeable hesitations or pauses are evident at B1 and B2. At B1, “pausing for grammatical and lexical planning...is very evident”, at B2 few long pauses should occur though learners “can be hesitant as he/she searches for patterns

and expressions” but at C1, speech should flow “almost effortlessly” for most subject areas (CoE, 2001: 28-29). In the UCLanESB data, both *er* and *erm* were high in all the word frequency lists: normalised frequencies of *er* were more than double those of LINDSEI for B1 and B2 whilst *erm* was still much higher at B1 and C1. However, for usage of *er* at least, there was a statistically significant fall from B1 to C1, and from B2 to C1. Successful speech therefore displayed a reduction in filled pauses using *er*. Though no difference was established for the use of *erm* across the proficiency levels, this study has determined that CEFR descriptors of fluency are to an extent upheld.

As stated earlier in Section 2.2, filled pauses such as these can be discouraged in teaching materials as, rather bluntly, they can make speakers “sound stupid” (Viney & Viney, 1996 cited by Hughes, 2011: 37), but this data compounds questions as to whether such a concerted effort is required. For instance, Biber et al. (1999) group fillers under ‘dysfluency’ features but for Götz (2013: 36), they act as a “fluency enhancement strategy” which can alleviate some of the pressures of spontaneous speech, especially in the case of a second language. *Er* and *erm* also form part of Bygate’s (1987) facilitative communication skills so actively avoiding them can be detrimental to success. They permit learners to pause before making important lexical choices, plan what they will say and retain turns when they have not fully completed their utterances. Furthermore, though *er* did reduce in the B1, B2 and C1 data, *erm* did not. This did not deny learners success in speech; in fact, as Canoz’s (1998) study of second language speech discovered, filled pauses are actually more typical of high proficiency learners who require time for planning. Ultimately, what this data shows is that whilst fillers such as *er* and *erm* are much more frequent than in the LINDSEI or BNC Spoken data, they should not immediately be dismissed as indicators of unsuccessful or disfluent speech. They can instead be an indicator of the application of strategic competence.

With regards to the collocational analysis of the UCLanESB data, another function of *er* and *erm* was identified. Supported by comparison with LINDSEI, though different to BNC Spoken data, speech at B1, B2 and C1 was able to demonstrate that filled pauses performed a role in discourse cohesion. MI scores shows that *er* and *erm* collocated significantly with *think* at all levels; they also collocated with several conjunctions albeit at varying rates across the three levels. This indicates that they facilitated cohesive links across utterances, a feature of Bachman and Palmer’s (1996) textual competence, and enabled learners to convey and link ideas. Since real-time communication exhibits simplified ‘chains’ of utterances (Carter & McCarthy, 2006), this

data illustrated how hesitation between messages often displayed some degree of hesitation in successful learner speech. Not only does this relate to fillers to bridge utterances and maintain turns, as mentioned, it also exemplifies how coherence and cohesion as per the CEFR (see Figure 26) can be seen as a combination of ‘shorter, discrete elements’ which can occasionally be ‘jumpy’:

Figure 26: CEFR descriptors for cohesion and coherence (CoE, 2001: 125)

COHERENCE AND COHESION	
C2	<i>Can create coherent and cohesive text making full and appropriate use of a variety of organisational patterns and a wide range of cohesive devices.</i>
C1	<i>Can produce clear, smoothly flowing, well-structured speech, showing controlled use of organisational patterns, connectors and cohesive devices.</i>
B2	<i>Can use a variety of linking words efficiently to mark clearly the relationships between ideas.</i>
	<i>Can use a limited number of cohesive devices to link his/her utterances into clear, coherent discourse, though there may be some ‘jumpiness’ in a long contribution.</i>
B1	<i>Can link a series of shorter, discrete simple elements into a connected, linear sequence of points.</i>

6.2.5 RQ2a Summary

The findings from RQ2a data can be summarised as follows:

- The 20 most frequent words at B1, B2 and C1 comprised approximately 40% of all speech. Knowledge of these words is vital for success.
- Pronouns such as *we* perform a range of functions and express a number of meanings which change across and within levels.
- B1 learners related topics to third parties but were also able to widen discussion to more general contexts.
- B2 learners are able to take a topic and relate it to their own lives or contexts
- C1 learners mostly used generic *we* but this was related to the nature of the topics discussed.
- Change in uses across levels reflects topic changes in the CEFR from those of familiarity to those of a more complex nature.
- Strategic competence exhibited using *we* for clarification and initiation of discourse.
- Collocates of *we* at B1 showed less variety. Evidence provided to demonstrate how vocabulary repertoire can ‘grow’.
- *Er* and *erm* very frequent at all levels.

- Hesitation via *er* did reduce as proficiency grew which reflected changes documented in CEFR descriptors of fluency.
- Hesitation was found to be a feature of strategic competence and success in general.
- Hesitation was also concentrated around conjunctions which linked utterances. Successful speech will show hesitation at these points but this also characterises strategic competence for keeping turns whilst speaking.

6.3 RQ2b: What were the 20 most frequent keywords at B1, B2 and C1 and their notable collocations, colligations and functions?

Analysis using the BNC World and LINDSEI wordlists produced two very different sets of data. Whilst BNC World identified the lexis that was more typical of this discourse type, e.g. *er, erm, I, think, so, because*, LINDSEI instead extracted lexis more characteristic of the topics discussed (see Section 5.1.4 for overview of topics). Topic related words do, of course, feature in the BNC World lists but clarifying this distinction here is useful for explaining why two distinct lists are presented for each level. Following Tables 33-38, once again, words selected for further exploration will be presented individually.

Table 33: B1 keywords - BNC World

N	Key word	Freq.	%	Texts	RC. Freq.	RC. %	Keyness	P
1	ER	1,220	8.4	15	90,254	0.09	8,721.31	0
2	I	943	6.49	15	732,523	0.74	2,481.08	0
3	ERM	300	2.07	15	63,095	0.06	1,512.72	0
4	THINK	203	1.4	15	88,700	0.09	739.21	0
5	MY	222	1.53	15	146,775	0.15	639.27	0
6	BECAUSE	168	1.16	15	100,659	0.1	513.3	0
7	LIKE	190	1.31	15	147,936	0.15	491.12	0
8	YEAH	140	0.96	14	83,012	0.08	430.19	0
9	MOVIE	46	0.32	4	1,753		385.06	0
10	OKAY	67	0.46	13	12,098	0.01	356.63	0
11	FILM	60	0.41	5	9,948	0.01	329.29	0
12	MAYBE	57	0.39	14	10,023	0.01	306.3	0
13	IT'S	132	0.91	14	126,792	0.13	292.33	0
14	CINEMA	37	0.25	6	1,873		289.06	0
15	WATCH	52	0.36	11	9,134		279.52	0
16	SO	171	1.18	14	239,549	0.24	271.83	0
17	SPORT	41	0.28	5	4,407		259.67	0
18	MM	74	0.51	11	34,736	0.03	258.99	0
19	VERY	106	0.73	14	119,611	0.12	205.65	0
20	CAN	132	0.91	13	211,093	0.21	182.26	0

Table 34: B1 keywords – LINDSEI

N	Key word	Freq.	%	Texts	RC. Freq.	RC. %	Keyness	P
1	ER	1,220	8.4	15	23,925	3.02	953.07	0
2	SPORT	41	0.28	5	30		233.96	0
3	WATCH	52	0.36	11	160	0.02	187.63	0
4	MY	222	1.53	15	4,526	0.57	156.28	0
5	CAN	132	0.91	13	2,224	0.28	124.76	0
6	FOOTBALL	27	0.19	4	45		123.38	0
7	CINEMA	37	0.25	6	168	0.02	109.93	0
8	I	943	6.49	15	37,060	4.67	94.65	0
9	MOVIE	46	0.32	4	371	0.05	93.69	0
10	THINK	203	1.4	15	5,131	0.65	93.14	0
11	TELEVISION	22	0.15	8	53		87.98	0
12	FILM	60	0.41	5	811	0.1	74.93	0
13	HARRY	11	0.08	3	4		71.15	0
14	SOME	108	0.74	13	2,289	0.29	70.66	0
15	GAME	19	0.13	3	59		68.24	0
16	FAVOURITE	21	0.14	7	85	0.01	66.34	0
17	POTTER	10	0.07	3	4		63.76	0
18	MEMORY	13	0.09	4	21		60.01	0
19	PERSONALITY	14	0.1	1	30		58.56	0
20	TV	7	0.05	5	0		56.25	0

Table 35: B2 keywords - BNC World

N	Key word	Freq.	%	Texts	RC. Freq.	RC. %	Keyness	P
1	ER	1,220	6.48	15	90,254	0.09	8,071.28	0
2	I	808	4.29	15	732,523	0.74	1,532.93	0
3	YEAH	307	1.63	15	83,012	0.08	1,245.82	0
4	ERM	280	1.49	15	63,095	0.06	1,233.05	0
5	THINK	263	1.4	15	88,700	0.09	957.24	0
6	LIKE	259	1.38	15	147,936	0.15	692.69	0
7	OKAY	108	0.57	15	12,098	0.01	620.58	0
8	YOU	454	2.41	15	588,503	0.59	596.48	0
9	MAYBE	97	0.52	14	10,023	0.01	572.59	0
10	MY	222	1.18	15	146,775	0.15	535.98	0
11	MM	125	0.66	15	34,736	0.03	499.69	0
12	YES	134	0.71	14	58,669	0.06	422.13	0
13	BECAUSE	163	0.87	15	100,659	0.1	412.7	0
14	SO	231	1.23	15	239,549	0.24	382.5	0
15	CAN	187	0.99	15	211,093	0.21	284.09	0
16	AGREE	54	0.29	14	8,060		280.02	0
17	LOT	78	0.41	13	27,912	0.03	274.6	0
18	CULTURE	50	0.27	5	8,481		246.9	0
19	IT'S	134	0.71	14	126,792	0.13	241.4	0
20	SMARTPHONE	13	0.07	2	0		222.89	0

Table 36: B2 keywords – LINDSEI

N	Key word	Freq.	%	Texts	RC. Freq.	RC. %	Keyness	P
1	ER	1,220	6.48	15	23,925	3.02	566.75	0
2	AGREE	54	0.29	14	61		250.49	0
3	CAN	187	0.99	15	2,224	0.28	197.66	0
4	TECHNOLOGY	29	0.15	3	8		180.08	0
5	PRESTON	23	0.12	9	0		173.16	0
6	WILL	124	0.66	15	1,165	0.15	172.25	0
7	YOU	454	2.41	15	9,842	1.24	160.09	0
8	CULTURE	50	0.27	5	174	0.02	146.79	0
9	USE	48	0.25	6	200	0.03	127.1	0
10	SPEND	37	0.2	10	100	0.01	123.44	0
11	UK	16	0.08	8	0		120.45	0
12	THINK	263	1.4	15	5,131	0.65	119.66	0
13	SUCCESSFUL	28	0.15	2	42		118.56	0
14	SPORTS	28	0.15	5	61		102.82	0
15	COMPUTER	25	0.13	6	44		99.93	0
16	SMARTPHONE	13	0.07	2	0		97.87	0
17	MY	222	1.18	15	4,526	0.57	90.9	0
18	SOME	136	0.72	14	2,289	0.29	83.69	0
19	NEED	37	0.2	10	197	0.02	83.52	0
20	TV	11	0.06	4	0		82.81	0

Table 37: C1 keywords - BNC World

N	Key word	Freq.	%	Texts	RC. Freq.	RC. %	Keyness	P
1	ER	733	3.55	15	90,254	0.09	3,966.81	0
2	ERM	450	2.18	15	63,095	0.06	2,315.95	0
3	YEAH	354	1.72	15	83,012	0.08	1,470.82	0
4	THINK	313	1.52	15	88,700	0.09	1,187.85	0
5	I	718	3.48	15	732,523	0.74	1,112.82	0
6	LIKE	302	1.46	15	147,936	0.15	841.37	0
7	IT'S	234	1.13	15	126,792	0.13	609.12	0
8	MAYBE	87	0.42	14	10,023	0.01	479.44	0
9	TOURISM	55	0.27	7	1,461		460.82	0
10	HOTEL	82	0.4	4	10,911	0.01	428.94	0
11	BECAUSE	173	0.84	15	100,659	0.1	428.16	0
12	SO	255	1.24	15	239,549	0.24	425.19	0
13	COUNTRY	99	0.48	15	27,959	0.03	375.43	0
14	DUBAI	31	0.15	2	141		363.62	0
15	PEOPLE	162	0.78	15	116,196	0.12	342.17	0
16	MM	100	0.48	10	34,736	0.03	340.58	0
17	YOU	360	1.74	15	588,503	0.59	305.3	0
18	UM	32	0.16	7	651		284.71	0
19	LAUGHS	31	0.15	9	588		280.03	0
20	REALLY	98	0.47	10	46,477	0.05	277.94	0

Table 38: C1 keywords – LINDSEI

N	Key word	Freq.	%	Texts	RC. Freq.	RC. %	Keyness	P
1	HOTEL	82	0.4	4	108	0.01	348.6	0
2	TOURISM	55	0.27	7	13		338.62	0
3	UM	32	0.16	7	0		235.2	0
4	IMPORTANT	89	0.43	15	338	0.04	234.56	0
5	WILL	146	0.71	15	1,165	0.15	217.37	0
6	ENVIRONMENT	46	0.22	9	46		212.95	0
7	DUBAI	31	0.15	2	2		212.86	0
8	LAUGHS	31	0.15	9	6		195.36	0
9	THINK	313	1.52	15	5,131	0.65	169.77	0
10	COUNTRY	99	0.48	15	708	0.09	163.43	0
11	AGREE	37	0.18	13	61		145.17	0
12	LOCATION	20	0.1	4	3		129.33	0
13	TRAVEL	43	0.21	10	136	0.02	125.67	0
14	GOVERNMENT	23	0.11	9	16		117.06	0
15	FOOD	47	0.23	12	232	0.03	104.37	0
16	YEAH	354	1.72	15	7,498	0.95	101.87	0
17	AIR	26	0.13	6	46		99.27	0
18	ERM	450	2.18	15	10,354	1.31	98.93	0
19	CANDIDATES	15	0.07	13	2		98.03	0
20	CARS	28	0.14	7	62		97.38	0

6.3.1 Think

- *Think* was chosen for analysis as it appeared in the BNC and LINDSEI lists at all levels
- *Think* was of fundamental importance to learners participating in exam-like discourse. Juilland's D of approximately 0.9 across all levels demonstrates its significance in the UCLanESB data
- Expected to arise as a keyword but its appearance in both BNC World and LINDSEI lists showed that it still occurred significantly more frequently in the UCLanESB data
- As Table 39 shows, *think* increased in frequency as proficiency developed
- It occurred on average 7 times more at C1 than it did at B1 and grew in keyness by 230.61 for BNC World and 50.11 for LINDSEI

Table 39: *Think* at B1, B2 and C1

Level	Raw freq.	% of corpus	Range	Mean freq.	SD	Juilland's D	BNC World		LINDSEI	
							Position in KW list	Keyness	Position in KW list	Keyness
B1	203	1.4	15	13.53	7.45	0.85	4	739.21	10	93.14
B2	263	1.4	15	17.53	5.67	0.91	5	957.24	12	119.66
C1	313	1.52	15	20.87	8.92	0.89	4	1187.85	9	169.77

- Clearly, *think* was most often used when learners wished to give their opinions
- *I think* appeared on 178, 206 and 261 occasions at B1, B2 and C1, respectively
- Unsurprisingly, *I* topped the collocates lists for t-scores and MI scores at the three exam parts across all levels, a finding reflected in the LINDSEI data
- Colligational patterns in the data were analysed to see how *I think* was used

Table 40: Colligational patterns of *think*

Usage	B1					B2					C1				
	A	B	C	Total	%	A	B	C	Total	%	A	B	C	Total	%
I [er/erm] think + object + is/'s + adjective phrase. <i>Example: I think er sport is very important</i> [B1]	8	14	9	31	64.58	2	11	4	17	40.48	8	13	18	39	52.70
I [er/erm] think + object + is/'s + noun phrase. <i>Example: I think it's er a matter of culture</i> [C1]	2	6	7	15	31.25	6	4	9	19	45.24	8	12	8	28	37.84
I [er/erm] think + object + is/'s + comparative or superlative. <i>Example: I think outside is better</i> [B2]	0	1	1	2	4.17	1	3	1	5	11.90	1	4	2	7	9.46
I [er/erm] think + object + is/'s + relative clause. <i>Example: I think the most important thing in our life is that there's no things should be happened in the future</i> [B2]	0	0	0	0	0.00	0	1	0	1	2.38	0	0	0	0	0.00
	Total			48	100	Total			42	100	Total			74	100

- Most common colligation contained a complement with an adjective phrase (this was set apart from comparative and superlative phrases to enhance analysis)
- One structure in particular *I [er/erm] think + object + is very important* was prominent in learner utterances:
 - B1 = 10 occurrences
 - B2 = 10 occurrences
 - C1 = 28 occurrences
- Learners used this to express their thoughts: it acted as a precursor to additional explanation. *Very important because* appeared on 6 out of 7 occasions at C1 for example
- Phrases such as *I think it is very important* also acted as a potential stalling device for giving learners valuable thinking time as they formulated ideas and speech. When it was used in such a way, it often followed or preceding other hesitations or stalls (see Extract 30). *Important* was found to be a collocate of *think* in Part B at B1 (t-score: 2.151, MI 4.716) and C1 (t-score: 3.754, MI 4.025) and in Part C at C1 (t-score:3.332, MI 4.712).

Extract 30: B1, Exam 1

<\$0> Thank you. <\$1M>. What do you do to keep healthy?

<\$1M> Keep healthy ***I think er sport is er very important*** you know er sometimes <\$=> I every </\$=> I like basketball and er sometimes I really with my roommate or my classmate go to basketball.

- The second most common colligational pattern combined *I think* and a noun phrase

- Across levels, this was achieved using lexis such as *way, place, city, thing, habit, subject, man, sport, or problem*
- Such utterances were simply used to provide an answer to a question briefly or to round off an utterance (see Extracts 31-33)

Extract 31: B1, Exam 5

<\$0> *Tell me about an interesting place you have visited recently.*

<\$9F> *Er erm in recent? <\$O15> Recently </\$O15>*

<\$0> <\$O15> *Recently yes </\$O15>*

<\$9F> *I went to York. <\$=> York is </\$=> I think York is good pla= good place.*

Extract 32: B2, Exam 3

<\$0> *Okay. Would you prefer TV without advertising? Would you like +*

<\$5M> *No. No I I don't agree it because I I have studies two years media and er in the future I want er go to a TV station. The er the television is erm is a very important part of earn money er in the of the to the TV er station. So I think it is not a good thing for the TV station.*

Extract 33: C1, Exam 8

<\$0> *Er okay <\$15F> what are the major transport problems in your country? Major transport problems in your country.*

<\$15F> *Mm major transport problems in my country? Okay I think it's the traffic jams*

Since the occurrence of comparative and superlative and relative clause colligations were low, *I think* was analysed for the collocations which surrounded it in learner discourse using a window of 5 words left and right to see if differences arose (see Table 41).

Table 41: Collocations of *I think*

Collocates of <i>I THINK</i>	B1 (I think = 178)			B2 (I think = 204)			C1 (I think = 261)			LINDSEI (I think = 4482)		
	Freq.	T-score	MI	Freq.	T-score	MI	Freq.	T-score	MI	Freq.	T-score	MI
Actually	0	--	--	3	1.482	2.792	0	--	--	52	6.225	2.871
Also	2	1.189	2.650	8	2.568	3.441	14	3.461	3.738	81	80.490	3.242
Basically	0	--	--	1	--	--	0	--	--	3	1.337	2.133
First	6	2.314	4.181	8	2.671	4.171	4	1.855	3.782	69	7.379	3.163
However	0	--	--	0	--	--	2	1.361	4.720	3	1.510	2.965
Maybe	8	2.581	3.518	20	4.237	4.250	11	2.985	3.322	88	8.348	3.182
Okay	7	2.335	3.092	3	1.057	1.359	7	2.426	3.589	229	14.488	4.554
In my opinion	5	2.198	5.865	8	2.771	5.622	2	1.289	3.498	7	2.494	4.125
Well	7	2.562	4.988	6	2.273	3.792	8	2.547	3.328	341	16.793	3.464
Yeah	13	3.130	2.922	28	4.663	3.074	32	4.866	2.838	744	25.723	4.134
Yes	10	2.926	3.742	11	2.879	2.922	5	2.021	3.379	441	19.287	3.616
And	28	4.240	2.332	32	4.667	2.514	47	5.767	2.655	1085	27.766	2.671
But	12	3.174	3.578	23	4.337	3.387	20	4.102	3.594	552	21.316	3.431
Because	22	4.252	3.418	23	4.428	3.703	20	3.983	3.193	341	16.732	3.413
cos	0	--	--	0	--	--	0	--	--	31	4.984	3.254
So	28	4.896	3.740	34	5.402	3.764	39	5.729	3.596	683	23.955	3.584

*words in red fall below the threshold for significance

**words in bold denote conjunctions referred to in Section 6.2.3 with *er* and *erm*

I think was found to be very multifunctional acting as a discourse marker, stance marker and hedging device (see Carter & McCarthy, 2006):

- It displayed a sequencing function due to combinations with *and*, *first* and *so*; the latter often used to conclude or round off remarks (see Section 6.3.3). Conjunctions such as those identified in the section on *er* and *erm* again were found to be significant here
- It helped to focus, divert or shift attention via the use of *well* e.g. *well I think*. Though this is generally used to change focus, in the UCLanESB data it allowed students to pause and sometimes introduce an idea different to the one expected by the listener
- The use of *actually*, *basically*, and *to be frank* (evidenced once at B2) was extremely low. The use of *I think* as a stance marker appeared to fulfil learners' needs at all levels
- *In my opinion* is omitted as a stance marker in Carter and McCarthy (2006). It was not overly frequent in the UCLanESB data but at B1 and B2, it strongly correlated with *I think*:
 - Learners either combined the two phrases because they did not think *in my opinion* worked in isolation, or:

- *I think* was added to emphasise that the subsequent utterance was simply an opinion. It added a degree of politeness or helped to maintain relationships
- In the BNC Spoken data, *in my opinion* only combined with *I think* on two occasions in the way the B1, B2 and C1 learners used it
- In the UCLanESB data, *in my opinion I think* was often placed at the beginning of an utterances whereas in BNC Spoken data, *in my opinion* was often used at the end of an utterance
- *Think* was also used as a hedging device either on its own or in combination with *maybe*. This collocation was found to be significant in the UCLanESB and LINDSEI data. Though it sometimes sounds rather repetitive, it is also a feature of NS speech as its appears 120 times in the BNC Spoken data with an MI score of 3.897 (see Figure 27)

Figure 27: BNC Spoken sample of *I think maybe*

BNC:KB8 S_conv	a heavy cold that was making her feel miserable. (SP:PS14L) Yes (SP:PS14B) And I think maybe it was with all of them. They were just slightly under the weather)%>
BNC:KC7 S_conv	What in here? In this room? (SP:PS0BL) Yeah. Yeah. (pause) I think maybe he's just come to realize (unclear) that (pause) (unclear) (pause) (SP:PS0BK) Yeah.)%>
BNC:KC9 S_conv	to (unclear) it was, it's such a waste that washing machine, I think maybe I've done a few shirts (unclear) hand washing and all the (unclear) water)%>
BNC:KD2 S_conv	. (pause) So (pause) (sighing) Yeah. So you got (SP:PS0J7) I think, I think maybe I would refuse to go, if I go in front of a T)%>
BNC:KPU S_conv	is married. Er erm (pause) and then later on I, I, I think maybe she did realize, you know, cos I said (pause) I wasn't)%>
BNC:KPV S_conv	(SP:KPVPSUNK) (unclear) (SP:PS6RW) It is difficult to guess how much transcription is needed. I think maybe (cough) what I'm going to do with transcriptions again. Cover the variations)%>
BNC:FY8 S_interview	n't think there was any way they were going to get anything different. I think erm maybe some of them felt it would be a waste of time fighting for)%>
BNC:GYK S_interview	Yeah. (pause) And also having got into the f-- into the complex, I think maybe having said that the spaces, you know the, the space between the)%>
BNC:HV5 S_interview	whole vocabulary as well. Which is the vocabulary is disappearing first and then I think maybe the actual intonations and phrases will disappear then too. But it'll take)%>
BNC:JP0 S_meeting	there for you to perhaps say all you wanted to say. (SP:PS4GM) Erm I think maybe we need to think about it, we need to give a bit more)%>

6.3.2 Discussion of the keyword *think*

Think, in particular, *I think*, was found to fulfil a range of functions in learner speech. Firstly, as an integral feature of B1, B2 and C1 learner speech, it was most associated with presenting and then expanding upon personal opinion. In Bygate's (1987)

treatment of production and interaction skills (see Section 2.2), he declares that during communication via spontaneous, unplanned speech, the use of routines is essential. With respect to *I think*, it was clear that learners at all levels employed one common routine: *I [er/erm] think + object + adjective phrase*. Such a routine allowed learners not only to express their initial thoughts, but it would also have reduced the demands on decision making during the interaction. The saliency of *I think it's very important* also fulfilled a strategic purpose as it occasionally acted as a stalling device, thus satisfying CEFR descriptions regarding 'buying time' (see CoE, 2001: 86).

Following this, a deeper layer of multifunctionality arose during collocational analysis. Often combined with *so* and *first*, *think* was found to be closely associated with sequencing language as was found in *er/erm*'s analysis. It allowed for chains of ideas to be linked and often signified the starting clause which would then be expanded upon using the "add-on" strategy (Biber et al., 1999: 1078). Routines containing *I think* such as that discussed in the previous paragraph would therefore supply learners with a 'way into' the discourse. In terms of learner success, *I think* often satisfied B1 calls for learners to express ideas as a "linear sequence of thoughts", B2 calls for more "detailed descriptions" and for C1 calls for learners to "develop particular points" as per the productive sustained monologue descriptors (CoE, 2001: 59). *I think* also played a pragmatic role in learner speech as a stance marker (see Carter & McCarthy, 2006). Stance, specifically epistemic stance, expresses a speaker's comments, evaluations and attitudes towards a particular topic as well as their origin (Biber et al., 1999; Carter & McCarthy, 2006). Though *actually*, *basically* and *in my opinion* relate to such meanings, it was clear that the sheer dominance of *I think* at all levels signified that the chunk was sufficient for learners to be successful. For instance, when *in my opinion* was used, it nearly always combined with *I think*, e.g. *in my opinion, I think....* Personal experience has demonstrated that stance and opinion-based language are often taught so as to add variety to learner speech and reduce repetition. However, this creates two contradictions. If *I think* suffices NSs in the majority of their regular conversations, doubt is raised as to whether learners should be discouraged from using it so frequently. Similarly, speech by its very nature is repetitive so instructing learners to deviate from this norm could have rather big effects on the 'naturalness' of their speech. Also, the fact that B1, B2 and C1 learners merged *in my opinion* with *I think* demonstrates that learners did not see the former's value in isolation and that tuition of this form created less target-like speech (they only collocate twice in the Spoken BNC).

Finally, in terms of multifunctionality, *I think* allowed successful learners to hedge their statements. Hedging performs an important sociolinguistic function as it helps to maintain relationships between speakers by making their utterances sound less “blunt and assertive” (Carter & McCarthy, 2006: 223). Whilst *I think* is capable of softening statements on its own, its collocation with *maybe* at each level, in particular at B2, emphasised the intentional uncertainty or lack of strength in learner speech. Analysis of the word *think* has therefore demonstrated that knowledge of individual words is not the only influencing factor in success; the flexibility that can be achieved using them and exploiting their multifunctional nature is also vital.

6.3.3 *So*

So was earmarked for additional investigation in the main study following the findings presented in Chapter 4. The pilot study, based on 12 C1 exams identified that nearly half of all its usage was as a subordinating conjunction for incorporating positive, negative or hypothetical ideas. Its second-most common use was as a discourse marker (38%) often for concluding remarks or signalling the end of a turn. Analysis in the main study therefore sought to discover whether its usage was similar or different across B1 and B2.

Tables 42 and 43 present data for *so* at all three levels and across all three exam parts; they are followed by a list of summarising bullet points before more detailed discussion is given.

Table 42: Frequency of *so* across Parts A, B and C at B1, B2 and C1

Exam Part	Raw freq.	% of sub-corpus	Position in sub-corpus freq. list	Position in sub-corpus keyword list	
				BNC	LINDSEI*
B1					
A	27	0.97	17	29	----
B	73	1.23	13	13	----
C	71	1.23	13	16	----
B2					
A	49	1.25	12	15	----
B	87	1.09	17	17	----
C	95	1.38	12	11	----
C1					
A	52	1.29	14	15	----
B	89	1.04	17	16	----
C	114	1.41	12	10	----

**so* was not found in any LINDSEI keyword lists

Table 43: Normalised frequencies and dispersion of *so*

Corpus	Raw freq.	%	Normalised freq.	Mean freq.	Range	SD	Juillard's D
B1	171	1.2	119.94	11.4	14 (93%)	7.1	0.83
B2	231	1.2	122.74	15.4	15 (100%)	7.9	0.86
C1	255	1.2	123.59	17	15 (100%)	9.3	0.85
BNC	236,884	0	23.69	----	----	----	----
BNC spoken	60,580	0	60.8	----	----	----	----
LINDSEI	10,257	0	129.35	----	----	----	----

- Range and Juillard's D show that *so* was used by learners throughout the exams at all levels.
- The mean frequency of *so* increased as proficiency developed. Successful learners needed to employ it at a similar dispersion rate but with higher frequency to complete interaction.
- Only at B2 and C1 did *so* occur less key and at lower percentages in Part B in comparison with Part A. This demonstrated a change in the nature of questions asked in Part A after B1.
- Part C at all levels saw a high usage of *so*. It allowed successful learners to answer questions designed to test the ceiling of their proficiency.
- *So* appeared in BNC keyword lists at all levels but not in LINDSEI lists. Clearly the research tool used was similar to that used in LINDSEI but it did not fully represent the dynamic spoken contexts captured by the BNC. The use of exams in data collection influenced this finding.
- The previous point is supported by B1, B2 and C1 normalised frequencies which were double that of the BNC spoken, but which corresponded with LINDSEI data.

This condensed summary of *so*'s quantitative aspects confirm that it was significant to the success of learner speech at all three proficiency levels. The next step involved extracting the functions that *so* fulfilled, as in the pilot study, and the shared and uncommon collocations across the levels. The procedure from the pilot study was followed in which descriptions from Carter and McCarthy (2006: 140-144 & 734) were used to analyse KWIC concordance lines. Exams were treated as a whole as opposed to being treated per exam part.

Table 44: Function of *so* at B1, B2 and C1

Categories	B1		B2		C1	
	Combined Freq.	Combined %	Combined Freq.	Combined %	Combined Freq.	Combined %
Adverb of degree	23	13.45	15	6.49	16	6.27
Substitute	4	2.34	11	4.76	8	3.14
Subordinating conjunction	76	44.44	87	37.66	119	46.67
Discourse marker	10	5.85	17	7.36	14	5.49
Incorrect use of 'so'	7	4.09	32	13.85	12	4.71
And so on	0	0.00	1	0.43	5	1.96
Miscellaneous	51	29.82	68	29.44	81	31.76
TOTAL	171	100.00	231	100.00	255	100.00

- Once again, the use of *so* as a subordinating conjunction to link clauses of result or consequence dominated at each level. (See Extract 34)
- B1 students displayed the use of *so* as adverb at a proportion larger than B2 and C1 combined (See Extract 35)
- B2 students showed a much higher rate of inaccurate uses of *so*. This sometimes included unnecessary repetitions of *so* in utterances but it also saw *so* being used in place of *and* or *because* (See Extract 36)
- At B1 and C1, discourse markers were used to summarise or clarify other speakers' utterances more frequently than they were used to open a conversation which was highest at B2 (See Extracts 37 and 38)

Extract 34: C1, Exam 13

<\$26F> <\$O40> You could <\$G?> </\$O40> What about interviewing people?
 <\$27M> <\$=> Well if you </\$=> <\$=> if </\$=> when when you want to find out about history <\$=> a few people about the </\$=> few people are left in the world **so** you can't really find a lot of people to interview and some of them might have forgotten some of the things that happened and so on.

Extract 35: B1, Exam 14

<\$28M> Comedy er because the er life is **so** hard so we need to spend something for fun so I love the comedies to watch with my friends or with my family.

Extract 36: B2, Exam 10

<\$20M> <\$O67> Yes we have to </\$O67> study **so** [because] my our <\$O68> English </\$O68> is not that good

Extract 37: C1, Exam 3

<\$6F> **So** y= **so** <\$=> you mean erm </\$=> you mean we we shouldn't to change the countryside.

Extract 38: B2, Exam 10

<\$20M> *So what do you think of lifestyle? Healthy lifestyle?*

Also as can be seen from Table 44 above, the ‘Miscellaneous’ category’s figures were rather high, occupying nearly a third of utterances at all three levels. This category uncovered that *so* functioned as a discourse marker for changing the topic of a conversation, for intervening in discussion to take or retake a turn and for signalling the end of a turn. In comparison with the pilot study, the use of *so* as a discourse marker for learners summarising their own utterances rather than those of other speakers was once again documented (see Extract 39). This formed approximately a sixth of all uses.

Table 45: *So* as a summarising discourse marker

Occurrence of discourse marker for learners summarising themselves			If added to existing discourse marker figures	
Level	Freq.	%	Freq.	%
B1	29	16.96	39	22.81
B2	37	16.02	54	23.38
C1	47	18.43	61	23.92

Extract 39: C1, Exam 15

<\$31F> *Yeah and also erm for <\$G1> we've been talking is erm also it will be really good for safety <\$O65> cos </\$O65> neighbours can help each other <\$O66> to protect them </\$O66> yeah you feel safe that the neighbour's not gonna <\$O67> rob you </\$O67> or gonna do something bad to you <\$O68> but </\$O68> there to protect each other so I think it's really good to have a good relationship.*

Examination of the collocates elicited from the B1, B2 and C1 data unveiled similar findings to those already identified. *So* was often used as a subordinator so as to explain or add to opinions. The identification of personal pronouns, *think*, and lexical and auxiliary verbs amongst the collocates common at the three levels is unsurprising as learners often had to add reasons to expand on their views (see Table 45). However, more intriguing is the list of collocates appearing only at particular proficiency levels (see Table 46).

Table 46: Common collocates of so at B1, B2 and C1 levels

Collocates of <i>so</i> common at all levels							
Collocate	Level	T-score	MI score	Collocate	Level	T-score	MI score
I	B1	11.772	3.849	DO	B1	2.725	3.449
	B2	9.333	3.419		B2	3.325	3.168
	C1	8.78	3.406		C1	3.104	3.266
IS	B1	5.626	3.735	THEY	B1	2.545	3.321
	B2	7.306	4.137		B2	4.045	3.09
	C1	5.588	3.102		C1	6.004	3.765
THINK	B1	5.234	3.743	MAYBE	B1	2.392	3.383
	B2	6.615	3.954		B2	2.958	3.209
	C1	6.132	3.541		C1	2.822	3.218
IT'S	B1	4.582	3.949	NOT	B1	2.365	3.239
	B2	5.08	4.141		B2	3.453	3.697
	C1	5.952	3.826		C1	3.611	3.362
LIKE	B1	4.214	3.298	ARE	B1	2.243	3.567
	B2	4.795	3.19		B2	2.941	3.838
	C1	5.82	3.458		C1	4.761	3.927
WE	B1	3.769	3.542	BE	B1	2.185	3.212
	B2	5.379	3.976		B2	2.372	3.274
	C1	4.804	3.727		C1	3.68	3.129

Table 47: Collocates of *so* specific to particular CEFR levels

Words found only at B1	Words found only at B2 and C1	Words found only at C1
KNOW	IT	YEAH
LONG	HAVE	YOU
WAS	IF	REALLY
MOST	WHEN	THINGS
WATCH	WHAT	ACTUALLY
BORING	THAT'S	
WILL		
LOVE		
IMPORTANT		

B1 collocates demonstrated combinations comprising *so* and topic-specific lexis. Additionally, though, the inclusion of *long*, *boring*, *important* confirm that B1 learners do use *so* as an adverb for modifying adjectives more than B2 and C1. The fact that both B2 and C1 collocates included *if* and *when* also add strength to statements here and later in Section 6.5 (RQ3 interaction) that success at higher levels necessitates an ability to explain and extend thoughts to incorporate explanation, hypothetical ideas or relevant

examples from everyday life. Finally, collocates identified only at C1 level demonstrate a more naturalistic character. Whilst *yeah* was often combined with *so* to indicate that a turn was ending (see Carter & McCarthy, 2006: 214-215), success at C1 also related to the ability to produce vague language (*things*), and utilise stance markers (*really*, *actually*) to convey additional attitude (see Figures 28 and 29).

Figure 28: C1 concordance lines for *so* and *things*

N	Concordance
1	erm to get reasonable things that they might need so if the hotel is in a place where like erm it's far
2	my degree my career that is business and marketing so probably have things like better opportunities for
3	a house you value it more when you know the history. So you value things more when you know the history
4	er people have their own perspective of of things so I think that the hotel rating is not like a measure for
5	improving like many aspects of things there in Dubai so as a result erm er tourists like like now consider
6	have forgotten some of the things that happened and so on. Mhm. So I think for me of knowing history is by
7	have people who describe and explain the things you so it's a good way of knowing history. Do you think if

Figure 29: C1 concordance lines for *so* and *really*

N	Concordance
1	is a lot of people at the same time period in the same place so it's really really harmful for the environ
2	the price of the hotel and it really really surprised me. laughs So er yeah it= it's over it's too expensive t
3	the weather is really good and it's really nice with rainy weather so yeah they really work on improving that
4	er tourism er in China er you know China is a really large country so in China tourism is really one of the im
5	there because it's a seasonal er weather like it comes from India so the weather is really good and it's reall
6	this comfortable hotel but we don't know if we really have this one so maybe we we will concern about it erm
7	that they just don't care about that they just care about money so that really sucks. I think it should be a
8	. Hot? Is really hot + Yes. + in summers and it's bit cold in winter so it's a really good a good weather to to
9	history a few people about the few people are left in the world so you can't really find a lot of people to in
10	actually one in erm Bahrain and one in Oman in the middle east so yeah that's really good and a lot of er l
11	where= wherever you are coming with probably would sustain you so the hotel amenities don't really matter.
12	Erm well for me I er really like er Chinese food so I always go to the Chinese restaurants
13	places like mountain deserts and beaches and yeah so on so we are really working on improving the
14	with really high ratings are probably going to be really expensive so I wouldn't consider going to a five star
15	geographical places like mountain deserts and beaches and yeah so on so we are really working on improv
16	or gonna do something bad to you but there to protect each other so I think it's really good to have a good r
17	they've got swimming pools they've got gym and suites and stuff so it's it's really nice you c= it attracts lots
18	internet people like us puts puts the information on the internet so we're not really sure if it's are correct
19	and er thinking something you never think be= er in a daytime so I think it's really good to er the life me
20	Yeah. + all the pictures you don't know er what it really looks like so erm this one maybe is not the most si
21	of tradition they follow. Now people don't really follow that tradition so I think it's more interesting to know ab
22	like the sports because I think I can watch the match on the tv so I don't really want to er spend my mon

Figure 30: C1 concordance lines for *so* and *actually*

N	Concordance
1	this? So reviews at the hotel you can't actually control about it so let's forget about it. Okay laughs Yeah. Uh-hu
2	for girls er I own I don't like do some sports actually laughs yeah so sometimes I will go shopping with my friend
3	to attract people and because Qatar won 2022 for football and yeah so now becoming growing actually from like to
4	sm= small country and my neighbourhood is actually in it's in Doha so it's like any normal neighbourhood in the wo
5	Fried Chicken but like erm all of us we call it SFC cos it's shorter so yeah erm it's not actually a restaurant it's a f
6	that erm we are the one that has to protect ourselves + Yeah. + so Actually er my house in Saudi Arabia is er or

6.3.4 Discussion of the keyword *so*

Once again, *so* embodied a keyword used throughout the B1, B2 and C1 data; though its dispersion remained relatively unchanged, *so*'s main occurrence did increase with proficiency. Carter & McCarthy (2006: 143) state that *so*'s "most common use" is as a discourse marker. The UCLanESB data, however, revealed that it functioned more frequently as a subordinating conjunction for expressing the results or consequences of clauses. As its collocation with *think* was again highlighted in the analysis, it is clear that *so* helped learners to be successful in the integration of additional detail and sub-themes (see CoE, 2001); a function illustrated further at B2 and C1 by the collocates *if*, *when* and *actually*.

When its use as a discourse marker was analysed more closely, differences emerged across the levels. B2 learners used *so* as a way of opening or initiating discussion whilst B1 and C1 learners used it to clarify or summarise others. In addition to aiding discourse via the sequencing of information, it was clear that success also depended on the strategic function it helped to realise. These results closely link to CEFR descriptors, albeit across different activities. B2 learners utilised *so* for turntaking to "initiate discourse", a feature of discourse competence, whereas B1 and C1 learners employed it strategically for clarification purposes (CoE, 2001: 124). Another use which emerged from the UCLanESB data relates to *so*'s utilisation for closing or concluding topics (see Carter & McCarthy, 2006: 214), especially at C1 whose collocates include *yeah*. However, a distinction between existing definitions of summarising B1, B2 and C1 language was that approximately a sixth of all summaries concentrated on an utterance they, rather than their conversational partner had produced. When this occurred, learners were often employing their strategic competence to signal that a turn was ending (see CoE, 2001; 86) or that a turn was concluding, an ability specified only at C1 but evidenced at all three levels (see CoE, 2001: 56).

6.3.5 *So, very, really*

This brief section builds on the discovery made in 6.3.3 that the use of *so* as an adverb of degree more than halved after B1. This prompted investigation (focussing on adjectives in R1 position) into other adverbs of degree and their collocates at B1, B2 and C1. The results are summarised in the bullet points below:

- Three main adverbs of degree identified: *so*, *very* and *really*
- *So* appeared in the top 20 most frequent word lists at all levels; it was also a keyword but several uses of *so* were combined in this figure
- *Very* fell beyond the 20 most frequent words (B1 = 24, B2 = 36, C1 = 28) but it was key at B1 and B2 (keyword list position = 19 and 23, respectively)
- *Really* showed dramatic differences across B1, B2 and C1; it did not appear as a keyword at B1 and whilst it appeared in 135th position in the B2 keyword list, it stood at 20th in the C1 lists (all keyword lists used in this analysis came from comparisons using BNC data)

Collocational data, using the top 20 results from words in R1 position, revealed interesting results as shown in Table 48. The table displays that:

- *So* was only used as an adverb of degree with adjectives in R1 position at B1, despite it being the highest level according to EVP
- *Very* was the most common adverb of degree across the levels; it also corresponded with LINDSEI data
- *Really* was common at C1. Though used at B2, it was more indicative of C1 learner speech
- The use of adverbs coincided with a slight increase in negative connotations as proficiency rose. 14.5% of statements at B1 (6/6 = *so*, 3/56 = *very*, 0/0 = *really*), 24.1% of statements at B2 (0/0 = *so*, 11/45 = *very*, 2/9 = *really*) and 20.4% of statements at C1 (0/0 = *so*, 13/66 = *very*, 10/47 = *really*) conveyed a negative idea or opinion
- All adjectives came from A1 to B1 according to EVP (CUP, 2015a). Again, evidence supports RQ1's findings that complexity of vocabulary is not the key to increased proficiency.

Table 48: Adverbs of degree and their collocates

Adverb of degree	B1 collocates (freq., T-score, MI)	B2 collocates (freq., T-score, MI)	C1 collocates (freq., T-score, MI)	LINDSEI (freq., MI)
SO	BORING (4, 2.406, 5.824) EXPENSIVE (2, 1.953, 5.409)	n/a	n/a	n/a
VERY	GOOD (12, 4.026, 5.405) IMPORTANT (10, 3.253, 5.7) BIG (5, 2.117, 5.251) NICE (4, 2.213, 6.613) FRIENDLY (4, 2.42, 6.362) BEAUTIFUL (4, 1.971, 6.099) INTERESTING (3, 1.648, 4.362) HAPPY (3, 1.694, 5.514) FAMOUS (3, 1.956, 5.514) USEFUL (2, 1.694, 5.514) HIGH (2, 1.363, 4.777) COLD (2, 1.993, 8.099) CLOSE (2, 1.964, 5.777)	GOOD (11, 3.753, 5.008) IMPORTANT (10, 3.264, 5.969) EASY (3, 2.425, 6.631) BEAUTIFUL (3, 2.801, 6.668) SERIOUS (2, 1.722, 7.368) NOISY (2, 1.406, 7.368) LONG (2, 1.701, 5.783) INTERESTING (2, 1.652, 4.429) HIGH (2, 1.967, 5.909) EXCITING (2, 1.718, 6.953) DANGEROUS (2, 1.384, 5.561) CUTE (2, 1.406, 7.368) COMFORTABLE (2, 1.384, 5.561)	IMPORTANT (28, 5.818, 5.921) GOOD (7, 2.996, 4.249) NICE (5, 2.161, 4.889) HARD (5, 2.621, 6.752) FAMOUS (5, 2.193, 5.682) USEFUL (3, 1.717, 6.852) TERRIBLE (2, 1.396, 6.267) FAST (2, 1.68, 5.045) DIFFICULT (2, 1.377, 5.267) DIFFERENT (2, 1.226, 2.91) CLEAN (2, 1.71, 6.267) BEAUTIFUL (2, 1.702, 5.852) OLD (1, 1.391, 5.945)	GOOD (342, 5.239) NICE (281, 5.778) BEAUTIFUL (198, 5.22) INTERESTING (159, 5.642) HAPPY (138, 5.589) DIFFERENT (114, 4.176) DIFFICULT (103, 5.258) HARD (75, 5.465) FRIENDLY (66, 6.469) IMPORTANT (65, 5.164) PROUD (61, 6.188) SMALL (59, 5.169) BIG (59, 4.834)
REALLY	n/a	BEAUTIFUL (3, 2.227, 8.015) WARM (2, 1.996, 8.809) INTERESTING (2, 1.39, 5.87) HOT (2, 1.992, 7.934)	NICE (13, 3.709, 6.825) GOOD (12, 3.774, 5.285) IMPORTANT (8, 3.029, 4.565) HIGH (3, 1.718, 6.981) HOT (2, 1.387, 5.718) DIFFICULT (2, 1.71, 6.303) COMFORTABLE (2, 1.71, 6.303) BAD (2, 1.36, 4.718) TERRIBLE (1, 1.401, 6.718) SURE (1, 1.401, 6.718)	n/a

6.3.6 Discussion of the keywords *so*, *very*, *really*

Although not expanded upon in Section 6.3.3, collocational analysis of *so* revealed that its use as an adverb of degree was more prevalent at B1. This prompted investigation of degree adverbs used more widely across the three learner levels. *Very*, *so*, and *really*, in that order from most- to least-frequent, are given in Biber et al. (1999) as the most common adverbs of degree in speech. In UCLanESB's data, however, learner preference rather than commonality seemed to be the main distinguishing factor: B1 learners used *so*, B2 learners used *very* and C1 learners used *really* more often than the other levels. Though collocations in R1 position were restricted in number at B1, *very* and *really* showed the increasing flexibility they afforded learners. As can be seen in the case of *really*, significant collocates increased alongside proficiency. This finding generates two propositions for success. The first is that, as was alluded to in Section 6.3.3, individual words themselves may not be a good indication of increased proficiency. Though RQ1 showed that the number of types increases from B1 to C1, it is instead the range of functions they enable and thus the flexibility they afford learners that is crucial to success at different CEFR levels. As Nation (2001), Qian (2002) and Schmitt (2008) assert, one should not prioritise the number of words known to the detriment of how *well* they are known. Similarly nor should complexity be prioritised since R1 collocation analysis found that all adjectives were learned between A1 and B1. Secondly, since amplifiers such as *so* and *very* (Biber et al., 1999), and *really* in this study, “increase [the] intensity” of gradable and non-gradable adjectives, learner success in the UCLanESB data has been able to show that one particular CEFR ability extends further than C1 and C2. Prepositional use at C1 relates to the ability to “qualify opinions and statements in relation to degrees of, for example, certainty/uncertainty, belief/doubt, likelihood, etc.” (CoE, 2001: 129). The varied use of degree adverbs, at all three levels, shows that such an ability can be found at other levels as well. Therefore, whilst learner preference or choice in terms of which adverb to use can be one indicator of success, adverbs of degree can still provide a more global picture of success that is not confined to the nuances identified beyond B2 in the CEFR only.

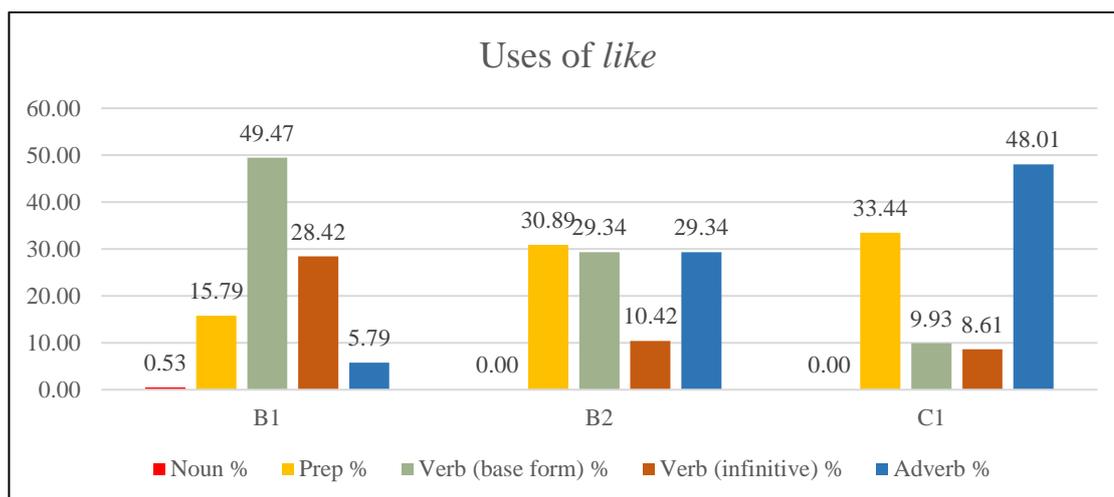
6.3.7 Like

Like ranged by 10 occurrences in comparative frequencies from B1 to C1 (B1 = 130.79; B2 = 137.61; C1 = 146.44). Its dispersion rate also showed that despite the drop in frequency across levels, the learners needed to use it throughout the interaction (Juillard's D = 0.88 at B1, 0.76 at B2 and 0.73 at C1). However, the inclusion of *like* in frequency or keyword lists can be misleading due to the numerous ways in which it can be employed. The Spoken BNC frequency list for example includes it on four separate occasions as a preposition, verb, adverb and conjunction. Examination of *like*, facilitated via the use of UCREL's (n.d.) Constituent Likelihood Automatic Word-tagging System [CLAWS] was completed to uncover its usage at each of the three levels (see Table 49 for raw frequencies and Figure 31 for percentage proportions).

Table 49: Uses of *like* (raw frequencies)

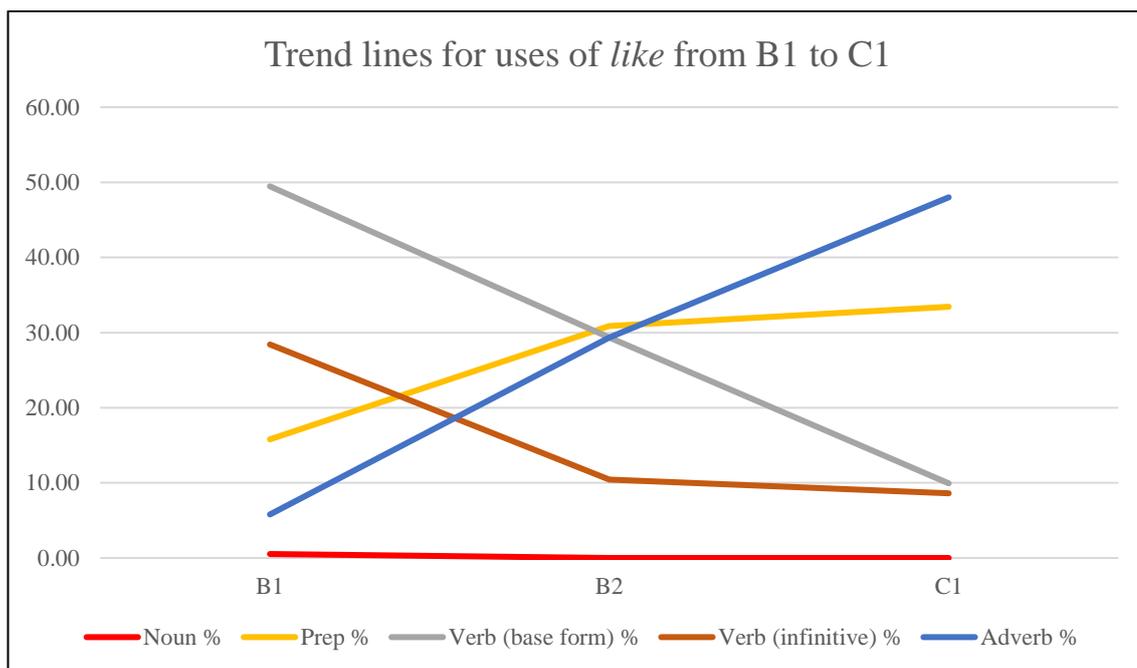
Level	Noun	Prep	Verb (base form)	Verb (infinitive)	Adverb	Total
	NN1	like_II	like_VV0	like_VVI	like_RR	
B1 Total	1	30	94	54	11	190
B2 Total	0	80	76	27	76	259
C1 Total	0	101	30	26	145	302

Figure 31: Uses of *like* (% proportions)



In addition to the chart above, trend lines depicting the percentage rises and falls in the usage of *like* demonstrate more clearly the changes across the three levels.

Figure 32: Like trends across UCLanESB data



- More than three quarters of its use at B1 constituted a verb form; this reduced to approximately 60% and 20% at B2 and C1 (see Extract 40)
- B2 did not exhibit any prominent usage but preposition use did hold the majority (see Extract 41)
- C1 students used *like* as an adverb on nearly 50% of all occasions. However, there were examples which could be attributed to overuse (see Extract 42)
- As the trends in Figure 32 show, verb base forms reduced from B1 to C1 at an almost identical rate to the use of adverbs rising
- No evidence was found of *like* being used as a conjunction at any level

Extract 40: B1, Exam 5

<\$9F> Okay. What type of film films do you **like** to watch? I **like** thriller yes erm and I like romance romance drama and this mm yes.

Extract 41: B2, Exam 14

\$0> Okay and what can young people do in your area?
<\$28F> In here? Er like there's all of activities they can do erm the garden I like the garden here because erm it's so huge with a lot of type of activities **like** erm tennis erm football er playing area for the children er also erm +

Extract 42: C1, Exam 2

<\$4M> Er I think that erm the major problems **like** transportation is the traffic. Cars and buses and all these things. Er I think that the perspective of the nation about <\$=> the

buses </\$=> using the buses you might see **like** only **like** the lower class of the nation use just the bus and erm however **like** I can see that <\$=> my </\$=> the transportation in my country is developing nowadays the government acts like responds wisely to the problem.

Statistical analysis also revealed interesting insights:

- Verb usage decreased, not between each level, but from B1 to C1 ($p < 0.01$). B1, B2 and C1 learners had mean scores of 9.87, 6.80, 3.73 for *like* as either a base or infinitive verb
- Usage as a preposition increased significantly from B1 to B2 and from B1 to C1 ($p < 0.01$ in both tests). There was no statistically significant increase in this respect from B2 to C1
- Adverb use did also increase significantly but once again, this was only significant from B1 to C1 ($p < 0.05$).

Collocational analysis also provided support for these findings:

- The importance of *like* for expressing learners' likes and dislikes, an item learnt at A1 (EVP, CUP, 2015a), could not be refuted. *I like* had an MI score of 3.922, 3.603 and 3.362 at B1, B2 and C1 and *don't like* had an MI score of 4.585, 4.326, 4.423 at the three respective levels
- However, the amount of evidence for *I like* did fall level-by-level: 12.633(B1 – 187 occurrences in collocate window), 10.663(B2 – 135 occurrences) and 9.391(C1 – 116 occurrences)
- *Don't like* on the other hand stayed rather stable: T-scores were 4.980 (B1 - 27 occurrences), 4.665 (B2 – 24 occurrences) and 4.954 (C1 – 27 occurrences)
- Collocates also arose at B2 and C1 to suggest how *like* was used as both a preposition and *adverb*
 - At B2, *something* collocated rather strongly with *like* to express similarity, a use learnt at A2 level or to give examples as a substitute for *such as*, learnt at B1 level (EVP, CUP, 2015a): t-score = 4.570, MI = 5.286.

Figure 33: Concordances for *something like* (B2)

N	Concordance
1	but she's very nice. And she just talk about something like hello how are you and how do you do like
2	. + and business + Yes. Er you know some= something like er some unique stories erm some prof=
3	older people er find help from er his sons or something like that or for young people that er use er s
4	to er stay in home yeah watch er television something like that. Mm yes because in every holiday m
5	just er like I want to do er and I want to do something like my best. I think er if I know what will ha
6	the culture you can like watch on TV + Yes. + something like natural geographic something like that +)
7	technology about some t= + Maybe microwave something like that yeah. Also maybe this is yeah I agree
8	those modern technology in their past life and something like that. They find it difficult. They used to N
9	pub or the have fun in the shopping centre or something like else. Yes I agree with that because peopl
10	TV + Yes. + something like natural geographic something like that + Yes. + so you can understand from
11	cou= like cities they have in every city they have something like unique about them about the same city ye
12	the people and never keep your head down or something like that yeah. I don't know I agree with er th

- At C1, *like* was combined with *if* to introduce an example that would illustrate a previous point: t-score = 4.582, MI = 3.952

Figure 34: Concordances for *like if* (C1)

N	Concordance
1	hotel was really good it had this and it had that like if someone came up to me and
2	the most important cos you waste money going like if you are going or coming to yo
3	go through the wording itself I just see erm like if going to state it out of ten wl
4	you have to be er clever like with the planes like if there's emergency or someth
5	if the rooms are no good you have to to the Like if there was like a hotel in Lon
6	more than that. Yeah definitely more than that er like if you go to some you know I d
7	it depends where the location of the hotel like if we're talking about a four sta
8	of the major problems in Riyadh is the traffic. Er like if I want to go somewhere like

- At C1, *like* as an adverb was used in a variety of ways, often as a filler, but examination of collocation revealed that it was sometimes employed to show that the learner was not entirely sure of their word choice e.g. *have like* (t-score = 5.124, MI = 3.648). In a sense, it acted as a hedging device, not for the meaning conveyed, but for the means of conveying it

Figure 35: Concordances for *have like* (C1)

N	Concordance
1	convenience. I don't think that students can have like a heal= well you can do if you rea
2	be my sister. Er maybe because is we have like the same interests so we share
3	will be true because I don't know they have like in rural areas they have like the s
4	who is running into something and doesn't have like the safety sense of driving. Yeah
5	on that but it's really difficult cos you don't have like much time. Most of the students
6	know they have like in rural areas they have like the sense of of helping each othe
7	are very nice yeah. Well I think that mm have like erm safety systems + Yeah. + ma
8	than international like in my country we have like an amaz= erm we have a lot of th
9	drive + Yeah. + cos some teenagers don't have like the sense of responsibility + Yeal
10	erm like the hotel is so costly you should have like amen= amenities that like er they

6.3.8 Discussion of the keyword *like*

Similar to *so*, the appearance of *like* in keyword lists did not tell the whole story in that its many functions were merged under one list item. Use of CLAWS identified four main categories: *like* as a preposition, verb base form, verb infinitive and adverb. The prevalence of each usage fluctuated across levels: at B1, *like* as a verb in base form dominated, at B2 prepositional usage narrowly occupied the majority and at C1, *like* as an adverb clearly overshadowed the other uses. Observation of trends identified that the fall of *like* in verb base form, from B1 to C1, mirrored the rise in its use as an adverb, both of which were found to be statistically significant changes across the levels.

Collocational analysis supported these key findings: *I like* and *don't like* were found to significant collocations despite the former clearly falling after B1 in the amount of evidence found or it. At B2, *something like* collocated with an MI of 5.286 to emphasise the similarity between items or to substitute *such as*. At C1, collocations such as *like if* and *have like* were identified. Though once again this variation in usage supports previous conclusions that success at higher levels is attributed to flexibility, rather than simply to the size and complexity of vocabulary, the latter collocations revealed other insights. The chunk *like if* can be easily called upon as a communication routine for adding detail or subthemes. However, *something like* and *have like* resonate strongly with the use of vague language. Vague language often denotes an assumption that speakers in a dialogue share a base of common knowledge. Whilst important for enhancing the efficiency of speech, it can otherwise be viewed “wrongly...as a sign of careless thinking of sloppy expression” (Carter & McCarthy, 2006: 202). While vague language is a feature of spoken grammar

and though here it enabled B2 and C1 learners to be successful, the strictest of assessors may see it as an avoidance strategy or a sign that learner vocabulary is lacking. This may be true in some instances, but it would be a generalisation to make this assumption every time. Since spoken grammar works towards the simplification of speech (see Section 3.2), it would be inaccurate or unfair to penalise learners for using it. As a final note, the collocate *have like*, used at C1, highlighted another distinction of relevance to success. Hedging, introduced earlier, relates according to Carter and McCarthy (2006) to the softening of utterances. The concordance lines presented evidence, however, that hedging can be an important tool or strategy for showing that lexical choices may not be the most accurate. This study therefore lends support to Biber et al.'s (1999) definitions that explicate the use of hedges to modify lexis. This study has demonstrated the pragmatic and strategic importance of a word such as *like*. It expresses likes and dislikes, it introduces examples, it highlights similarity, it combines with collocates to exploit vague language and finally, it acts as a hedging device for showing uncertainty not only of the content of utterances, but the means employed to realise them.

6.3.9 RQ2b summary

Analysis for RQ2b has shown that:

- Keyword data, when compared against the spoken BNC showed that the research tool did increase the keyness of exam-like lexis.
- To be successful, analysis using LINDSEI as a reference corpus highlighted that B1, B2 and C1 learners required topic-specific lexis also.
- The verb *think* was used throughout in a range of colligations illustrating how communicative routines for giving opinions could be realised.
- Collocational analysis of *think* revealed a variety of functions. It enabled learners to successfully sequence utterances, shift focus, express stance and hedge language.
- *So* similarly enabled a range of language functions either in discourse, linking cause and effect, or as a discourse marker for summarising, clarifying or opening discussion.
- Adverbs of degree at the different proficiency levels showed that for this study at least, B1 students are characterised via their use of *so*, B2 learners are

characterised via their use of *very* and C1 learners are characterised by their use of *really*.

- Success once again was related not to the complexity of vocabulary but another finding from analysis of degree adverbs could be that learner preference for the words they use could also be a potential indicator of proficiency.
- *Like* significantly increased as an adverb as its usage as a verb base form decreased
- Prepositional use also increased significantly after B1.
- Success at higher proficiency levels once again related to the flexibility in functions an individual word could satisfy. Just because a word is used frequently does not mean that it is always repetitive or used in the same way.

6.4 RQ2c: What were the most frequent 3- and 4-word chunks at B1, B2 and C1 and their notable collocations and functions?

The three- and four-word chunks for the B1, B2 and C1 data are presented below. In total, there was a rise in the three-word chunks from 469 at B1, 481 at B2 and 538 at C1. Four-word chunks did not display such uniform growth. Despite B2 having 17 more chunks than B1, B1 and C1 learners were rather similar in the chunk frequency they displayed. It was also evident, as Carter and McCarthy (2006) remark, that many of the four-word chunks were extensions of three-word chunks. Statistical significance was not reached in any of the comparisons so the rise in proficiency for this study did not relate to a rise in the number of chunks. Statistical significant tests did highlight, however, that learners at all levels produced significantly fewer four-word chunks than three-word chunks ($p < 0.01$ at all levels).

In Tables 50 and 51 below, it is first striking to see how many chunks at all levels centre on the verb *think*. Whereas at B1, emphasis was placed on giving opinions, *er I think, I think er, I think it's*, at B2 and C1, chunks were employed both to give and seek opinions: *I think it's, so I think, what do you think, do you think about*. It is also notable that the fillers *er* and *erm*, discussed in Section 6.2.3 also featured in such chunks at all levels but that their inclusion also decreased across the three- (B1 = 8, B2 = 4, C1 = 3) and four-word chunks (B1 = 6, B2 = 4, C1 = 0). Closer examination of RQ2c focused on three specific areas: *a lot of, I agree with (you)* and topic-related lexis.

Table 50: B1, B2 and C1 three-word chunks

	B1			B2			C1		
	CHUNK	F	R	CHUNK	F	R	CHUNK	F	R
1	I DON'T	68	14	A LOT OF	53	11	I THINK IT	47	15
2	I THINK ER	43	11	I DON'T	38	13	IN MY COUNTRY	41	12
3	A LOT OF	35	11	I THINK ER	35	12	I DON'T	36	14
4	ER I THINK	33	12	I AGREE WITH	29	11	THINK IT'S	36	13
5	I THINK IT	28	10	ER I THINK	28	11	A LOT OF	35	11
6	THINK IT'S	27	10	I THINK IT	28	9	SO I THINK	32	12
7	I WANT TO	23	8	AGREE WITH YOU	27	12	IT'S A	27	10
8	DON'T LIKE	22	8	IN THE FUTURE	24	6	DO YOU THINK	25	12
9	ER IT'S	20	9	DO YOU THINK	23	9	OF THE HOTEL	25	4
10	AND ER I	19	12	SO I THINK	22	10	I THINK THE	24	11
11	ER AND ER	16	6	THINK IT'S	21	9	I THINK THAT	23	9
12	ER YOU KNOW	16	6	ER IT'S	19	8	IT'S NOT	23	13
13	I CAN'T	16	7	I WANT TO	19	8	I AGREE WITH	22	10
14	AND ER ER	15	9	LOT OF TIME	19	4	IT'S VERY	22	11
15	DON'T KNOW	15	7	AND ER I	18	10	YEAH IT'S	22	10
16	ER I LIKE	15	9	DON'T HAVE	17	7	ER I THINK	21	9
17	IT'S VERY	15	6	WHAT DO YOU	17	10	ERM I THINK	21	10
18	SO I THINK	15	6	I THINK I	15	11	WHAT DO YOU	20	10
19	GO TO THE	14	8	THEY DON'T	15	5	DON'T HAVE	18	12
20	I LIKE TO	14	7	I LIKE TO	14	8	I THINK ERM	18	5
	Total	470		Total	481		Total	538	

Table 51: B1, B2 and C1 four-word chunks

	B1			B2			C1		
	CHUNK	F	R	CHUNK	F	R	CHUNK	F	R
1	I THINK IT'S	24	9	I AGREE WITH YOU	21	9	I THINK IT'S	35	13
2	I DON'T LIKE	21	8	I THINK IT'S	20	9	WHAT DO YOU THINK	15	9
3	ER I THINK ER	13	8	A LOT OF TIME	19	4	I AGREE WITH YOU	11	9
4	I DON'T KNOW	10	6	WHAT DO YOU THINK	15	9	I DON'T KNOW	11	7
5	GO TO THE CINEMA	9	5	SPEND A LOT OF	14	3	LOCATION OF THE HOTEL	11	4
6	TO BE IN A	9	5	THEY DON'T HAVE	10	3	YEAH I AGREE WITH	11	7
7	AND ER AND ER	9	4	DO YOU THINK ABOUT	9	5	IT'S IT'S	10	6
8	ER A LOT OF	8	6	I DON'T THINK	9	7	THE LOCATION OF THE	10	3
9	I WOULD LIKE TO	8	4	LOT OF TIME WITH	9	1	DO YOU THINK ABOUT	8	4
10	BE IN A FILM	7	5	AGREE WITH YOU BUT	8	4	A LOT OF PEOPLE	7	6
11	ER I DON'T	7	5	YES I AGREE WITH	8	6	IN MY COUNTRY IS	7	3
12	HAVE A LOT OF	7	4	ER I THINK ER	7	4	SO I THINK IT	7	6
13	LIKE TO BE IN	7	4	I DON'T LIKE	7	3	A LOT OF THINGS	6	5
14	WANT TO BE A	7	5	WITH YOU BUT ER	7	3	I THINK THAT THE	6	4
15	I I DON'T	6	4	AND ER I THINK	6	4	MOST OF THE TIME	6	4
16	I LIKE TO WATCH	6	4	AND ER IT'S	6	4	THINK IT'S VERY	6	5
17	I THINK ER ERM	6	4	HAPPEN IN THE FUTURE	6	2	TOURISM IN MY COUNTRY	6	2
18	I THINK ER THE	6	4	I DON'T KNOW	6	5	YEAH IT'S VERY	6	4
19	I WANT TO BE	6	3	SO I THINK IT	6	3	A FOUR STAR HOTEL	5	1
20	AND I DON'T	5	4	DON'T THINK SO	5	3	BUT I THINK IT	5	5
	Total	181		Total	198		Total	189	

6.4.1 *A lot of*

A lot of, introduced at A2 (EGP, CUP, 2015b) was chosen due to its high frequency at all three levels and for the fact that it appears in the Spoken BNC and LINDSEI chunk data. It also shared a similar dispersion rate across levels in that Juilland's D for B1, B2 and C1 was 0.72, 0.68 and 0.72, respectively. Since, chunk occurrence was much reduced in the UCLanESB corpus, *a lot of* was explored in terms of its influence on accuracy and collocation. It was not the intention of this study to analyse errors, but since gains in accuracy affects judgements of success and since accuracy rises with the use of chunks (see Section 3.3.2), a brief error analysis was conducted. *A lot of* was also compared with *many* and *much* to see if any parallels could be drawn.

- All countable and uncountable nouns were classed as either correct or incorrect. When turned into a mean percentage, *a lot of* had accuracy scores of:
 - 86.32% at B1
 - 67.81% at B2
 - 98.23% at C1
- At B1, all uses of uncountable nouns (money, art, food, technology, and time) were accurate. Countable nouns proved more difficult with the plural 's' often missing e.g. *a lot of experience~~s~~, building~~s~~, park~~s~~, place~~s~~, film~~s~~, book~~s~~ and word~~s~~*
- At B2, again, countable nouns posed more problems than uncountable (the latter being incorrect only on three of 27 occasions). In terms of countable nouns, again, plural 's' was omitted: *a lot of element~~s~~, definition~~s~~, film~~s~~, culture~~s~~, tradition~~s~~, thing~~s~~, programme~~s~~ and source~~s~~*
- At C1, all uses of uncountable nouns, such as *food, frustration, benefit, money, and entertainment*, were accurate. Only two examples of countable nouns, *problem~~s~~ and point~~s~~*, were incorrect out of 33.
- Countable nouns were a challenge at all levels despite their instruction and use at A1 level according to English Grammar Profile (CUP, 2015b)
- Though statistical significance was reached in comparing means from B2 to C1 ($p < 0.01$), it is difficult to categorically state that use of *a lot of* becomes more accurate as proficiency increases since B1 learners outperformed the B2 learners.

- Analysis of *much* and *many* showed that accuracy did improve across the B1 and C1 levels. Though combinations of *many* and correctly formed countable nouns were the common source of error, the error rate related to *much* and an uncountable noun did slightly increase
- Though *much/many* improved in accuracy across B1, B2 and C1, in comparison with *a lot of*, accuracy was still lower (except in the case of B2)
- *A lot of* arguably helped most learners to improve in their overall accuracy

Table 52: Accuracy of *much* and *many* at B1, B2 and C1

B1					
	FREQ	CORRECT	%	INCORRECT	%
MUCH	2	0	0	2	100
MANY	31	21	67.74	10	30.3
TOTAL	33	21	63.64	12	36.36
B2					
	FREQ	CORRECT	%	INCORRECT	%
MUCH	9	6	66.67	3	33.33
MANY	24	17	70.83	7	29.17
TOTAL	33	23	69.7	10	30.3
C1					
	FREQ	CORRECT	%	INCORRECT	%
MUCH	9	8	88.89	1	11.11
MANY	49	43	87.76	6	12.24
TOTAL	58	51	87.93	7	12.07

In terms of collocates, a wide variety of items in R1 position was not found due to the small frequencies involved. T-scores were often lower but MI scores were still able to show that collocations comparable with BNC Spoken and LINDSEI did exist.

Table 53: Common collocates of *a lot of* across UCLanESB, Spoken BNC and LINDSEI data

Corpus	R1 collocate	Frequency	T-score	MI score
B1	Time	3	1.924	4.72
	People	3	2.379	5.112
	Er	3	4.941	3.351
	Channels	3	1.995	8.697
	Things	3	1.378	5.305
B2	Time	19	4.746	6.587
	Things	4	2.198	5.888
	Er	3	4.748	3.078
	People	2	1.827	3.53
C1	People	7	3.234	5.323
	Things	6	2.618	6.585
	Erm	3	4.528	4.85
	Er	3	3.689	3.686
	Money	2	1.366	4.882
Common collocates across UCLanESB				
BNC spoken	People	611	24.222	5.638
	Things	139	11.232	4.403
	Er	68	2.081	0.42
LINDSEI	People	115	13.299	5.676
	Things	61	9.253	5.701
	Er	23	14.611	3.111

- Interesting to note, despite the low frequencies, is the use of vague language by UCLanESB learners. The MI scores for *a lot of things* are higher than BNC Spoken and LINDSEI at B2 and C1; at B1, the difference is still not that large.
- The appearance of *er* and *erm* after *a lot of* could lead some to believe that UCLanESB learners exhibited lower levels of fluency but this was also reflected in LINDSEI data
- As Extract 43 shows, such a chunk can help learners be more efficient during real-time speech

Extract 43: B2, Exam 13

<\$0> Er <\$27F> how might your culture be different in 100 years' time?
 <\$27F> In 100 years' time in the future. Erm <\$E laughs /\$E> er I think there is huge huge change will happen in the future. If er God give me a longer life to see that I will see my my my my kids to grow in the in that time so I think **a lot of things** will change.

Nobody have s= er the own or local culture all the culture in the world er the same I think so.

6.4.2 Agree with

In Tables 50 and 51 earlier, *agree with* appeared in multiple forms. Whilst it did not feature in the B1 lists at all, it formed five chunks at B2: *I agree with*, *agree with you*, *I agree with you*, *agree with you but*, and *yes I agree with*, and three chunks at C1: *I agree with*, *I agree with you*, *yeah I agree with*. Although in concordance data, *agree with you* did appear 4 times at B1, chunk data suggested that interaction based on reactions to others' utterances was not a priority at this level. However, analysis of *agree with* at the remaining two levels raised questions as to the authenticity of listenership.

- B2 learners *appeared* to show genuine listenership during the interactions. The four-word chunk *Yes I agree with* appeared 8 times at B2 and though it did not appear on the lists, *Yes I agree with* was also used 4 times. At C1, *Yes I agree with* did not occur whereas *yeah I agree with* occurred 10 times.
- B2 learners also occasionally employed stance markers, such as *totally* (3 times, MI = 8.763), *absolutely* (3 times, MI = 8.500), *completely* (1 time), and *definitely* (1 time), to add a degree of strength to agreements (see Extracts 44 and 45). At C1, only *totally* appeared once with *I agree with you*.

Extract 44: B2, Exam 7

<\$13M> *Okay er I am **totally** agree with you but if er the family er make timetable for the the girls go to maybe her week er outside and er they maybe don't feel bored +*

Extract 45: B2, Exam 15

<\$32M> *Yeah **absolutely** I agree with you but er they less communication with each other.*

However, the use of *I agree with* did not necessarily perform a purely opinion-based role. Analysis revealed that it was multifunctional in nature, especially at B2. Take Extract 45 as an example. Despite the emphatic use of *absolutely* to imply explicit agreement with the previous statement, it is followed by the conjunction *but*. This was found in a number of exams at B2:

Figure 36: *I agree with you but* B2

N	Concordance
1	you can dr= drive to the city centre well . Yes I agree with you but I think I still enjoys life in er
2	time working at the computer. Yeah absolutely I agree with you but er they less communication
3	er + Yes. + good teacher. I I agree with you I agree with you but er it depends with the
4	+ Yeah. + they have to control this. Yeah Yes I agree with you but er as you know everything it
5	to receive it from teacher. How about you ? Er I agree with you but er smartphone at this
6	. And er + Mm yeah I + Sorry. + abso= absolutely agree with you but er mm if if cold weather er I
7	mopping and er something. Okay er I am totally agree with you but if er the family er make
8	er to Arabic a lot of time when we study. Yeah I agree with you but er I also may use

Whilst *but* was a significant collocates at B2 ($t = 3.383$, $MI = 5.420$), it did not appear in the C1 or LINDSEI collocates list at all. At C1, the conjunction *but* was instead replaced with *because* ($t = 1.946$, $MI = 5.091$). Though only occurring on four occasions, and once being used to show disagreement (see Figure 37), C1 learners employed the chunk more for its interpersonal nature and more for its ability to extend and build on utterances. The B2 finding, however, indicated that, *I agree with you* performed a different function: that of a strategic stalling device for gaining time whilst students formulated their own ideas. As can be seen in the figure above, *agree with you but* was nearly always accompanied by *er* which accentuated the hesitation between learners using the chunk and joining it with their own opinion. Though giving the impression of listenership, KWIC analysis revealed otherwise at the B2 level.

Figure 37: *I agree with you because* C1

N	Concordance
1	will do that such as recycling erm rubbishes to the bins. Yeah I agree with you because erm people use educati
2	many years if we don't help the environment. No this time I don't agree with you because it's not only the cars +
3	they all help to clean the environment and give us a green life. I agree with you because also the cars er pollute
4	and smoking cancer like throat cancer tongue cancer well. Yes I agree with you because junk food is also food a

As a final note in this section, the occurrence of *I agree with*, *agree with you*, and *I agree with you* in the B2 and C1 chunk lists prompted another line of enquiry. Since chunks were used to infer or express agreement, data was analysed to see how disagreement was expressed at all three levels. Whilst this uncovered another strategic use of *agree*, *agree* for prompting a listener to take the turn (*do you agree [with me]?* occurring once at B1, three times at B2 and once at C1), *I don't/didn't agree* was only

used five times across all levels. What tended to happen instead, was that learners used the verb *disagree*: five instances at B2 and in four instances at C1. In spite of this, the data also showed that the expression of disagreement via a chunk or verb had a potentially negative impact on the rapport between speakers. At C1, for instance, a speaker disagreed with his partner twice in quick succession. Whilst this made the disagreement much more explicit, thus prompting a joke from their partner (see Extract 46), it highlights the potential influence disagreement can have on this type of discourse.

Extract 46: C1, Exam 14,

<\$29M> *I again don't agree with you.*

<\$28M> *Why?*

<\$29M> *Cos +*

<\$0> <\$E> *laughs* </\$E>

<\$28M> *He's not agreeing with me* <\$E> *laughs* </\$E>

The much more superior frequency of *I agree with* in comparison with *I don't agree* or *disagree* implies that the chunk also held a sociolinguistic function of maintaining relationships between speakers during the interaction.

6.4.3 Topic-related chunks

To very briefly conclude the presentation of results for RQ2c, the focus shifted away from more generic, neutral chunks to an exploration of those that related specifically to a particular topic. Three-word chunk data did not reveal many chunks of this nature, but in the four-word chunk lists, 6 were found at B1, 1 was found at B2, and 4 were found at C1.

- The chunks identified largely related to one topic at B1 and C1; the topic that was found to have been included on several occasions in Section 5.1.4
 - B1 = Cinema: 5 out of 15 exams
 - C1 = Travel and tourism: 7 out of 15 exams
- At B1, cinema related chunks accounted for 45 chunks (24.8% of all four-word chunks)
- At C1, travel and tourism chunks accounted for 32 chunks (16.9% of all four-word chunks)
- Given that at all levels, there was a statistically significant drop from 3-word chunk totals to 4-word chunk totals ($p < 0.01$ at all three levels), the data could

suggest a clear ability of learners to ‘pick up’ on chunks in written prompts or interlocutor questions (see Extracts 47 and 48): /.../ indicates first chunk, //...// indicates second chunk)

Extract 47: B1, Exam 11

<\$21F> *Erm er I prefer television than cinema cos er there is a different channels that I can choose and different and it's available all the day so I can can watch any time I want er no specific time <\$G?> like in cinema. <\$E> reads prompt sheet </\$E> <\$=> Would you like to be </\$=> no I don't /like to //be in/ a film// <\$=> because I think I don't </\$=> /like to //be in/ a film// you mean to +*
<\$0> *To be a star in a <\$O32> film yes </\$O32>*
<\$21F> <\$O32> *film </\$O32> No cos I don't like to act. Yeah so I don't /like to //be in/ a film//.*

Extract 48: C1, Exam 1

<\$E> *Candidates read sheet </\$E>*
<\$E> *\$2M whispers </\$E>*
<\$2M> *Erm alright fine er <\$=> how's it important that /the //location of the/ hotel// </\$=> how important is it to you /the //location of the/ hotel//?*
<\$1F> *Erm /the //location of the/ hotel// has to be close to the tourist places.*
<\$2M> *Yeah.*
<\$1F> *So I don't have to spend much.*

6.4.4 B1, B2 and C1 lexical chunk discussion

The data demonstrated that UCLanESB learners did use formulaic chunks in their speech. Averaging between 30 and 35 three-word chunks and approximately 12 four-word chunks per exam, formulaic language allowed learners a degree of success in their interactions. However, correlation between rising chunk numbers and rising proficiency could not be established. For instance, though three-word chunk numbers grew by 60 from B1 to C1, statistical significance was not reached for either the three- nor four-word categories. The frequency of chunks was therefore not a reliable indicator of growing proficiency in the UCLanESB data. One reason explaining this could be the differences exhibited in individual exams. Three-word chunk data showed considerable variation from exam to exam: B1 frequencies fluctuated from 11-46 chunks, B2 figures wavered between 18-59 chunks and at C1, chunk frequency varied from 12-68. Four-word chunk data was also similar in this respect 5-22 chunks at B1, 4-47 chunks at B2, and 4-31 chunks at C1. As the CEFR explains, each learner’s communicative competence is formed by their previous language experiences. It could be possible that learners within levels had not had much experience of lexical chunk tuition whereas others had.

Attention thus turned to the nature of the chunks produced to see whether this had an effect on learner success at B1, B2 and C1. Relating chunks to Carter and McCarthy's (2006) categories revealed that the vast majority of chunks contained subject-verb forms comprised of lexical and auxiliary verbs. Conjunction-verb structures were also used quite often across the levels and occasional preposition and noun phrase expressions were also identified. Again, regardless of three- or four-word constructions, not much difference was found across the levels but for success, the data demonstrates that some of the chunks relate to previously discussed functions, for instance, *I think it's, so I think, I think er/erm, I don't like, and er, and it's very* (e.g. important). Previous discussion has highlighted the pragmatic, discourse, strategic and sometimes sociolinguistic functions that individual words contained within them facilitate. What this chunk data highlights is that their use also impresses a degree of fluency or ease given the benefits for online processing and retrieval chunks are said to have (see Section 3.2). Whilst *er* and *erm* did appear in some chunks, especially at B1, the fact of the matter remains that salient lexicogrammatical items highlighted earlier as having an impact on success, added another dimension to successful learner language use.

The final point to be raised here regarding chunk type was able to pinpoint one clear difference between levels. Intrapersonal chunks such as *er you know, I don't know, but I think, I think er* are useful as they "reflect...meanings (meanings which build and consolidate personal and social relations) created between speakers and listeners" (Carter & McCarthy, 2006: 835). Some of these, especially those centred on the verb *think* showed that B1 learners were more concerned with expressing opinions while learners at B2 and C1 distributed usage between expressing and seeking opinions. The degree of interaction is not only exemplified via chunk data, but it is also realised via chunks in learner language; simultaneously this finding stresses the more strategic nature that turn-taking chunks can have (see Carter & McCarthy, 2006: 836). To reinforce this final point, other chunks relating to agreement *I agree with you*, reveal a greater attention paid to interacting in the discourse at B2 and C1 rather than discourse being potentially composed of two separate monologues with little joint construction.

6.4.5 RQ2c summary

Analysis for RQ2c has shown that:

- Learners at all levels did use chunks in their speech.

- Though increases in frequency were evident across the levels, there were no statistically significant gains in usage as proficiency grew.
- Many chunks centred on the verb *think* and were useful for expressing opinions and also seeking them.
- Chunks such as *a lot of* identified that chunks can have gains in grammatical accuracy. Correct usage is not guaranteed but for lower level learners, *a lot of* may prove an easier to use option than *many* and *much*.
- The chunk *I agree with you* also demonstrated a real attempt to interact at B2 and C1; it did not appear in chunk lists at B1. However, B2 learners did use it more as a stalling device while they prepared their own utterances. It performed a sociolinguistic, pragmatic and strategic function.
- Multifunctional chunks can facilitate success as proficiency rises.
- Topic-related lexis should also be taught in chunks so that learners can benefit from their advantages for fluency.

6.5 RQ3: What Common European Framework of Reference for Languages (CEFR) indicators are present in terms of spoken interaction, spoken production and strategies at B1, B2 and C1 and how are they realised?

Analysis of learner language at the B1, B2 and C1 levels was able to pinpoint the occurrence of the selected can-do statements identified in Appendix 12. The prevalence of production, interaction, production strategy and interaction strategy statements will therefore be presented here by level. Summaries of findings will then be offered per level before Section 6.5.4's brief comparison of the higher-occurring statements. Section 6.5.5 to 6.5.12 will then explore particular statements in-depth but due to the number of statements analysed, this more detailed discussion will have to be limited to those with clear findings for learner success.

6.5.1 B1 can-do occurrence

In figures 38-41 the frequency of B1 statements is displayed for production, interaction, production strategies and interaction strategies.

Figure 38: B1 production statement occurrence

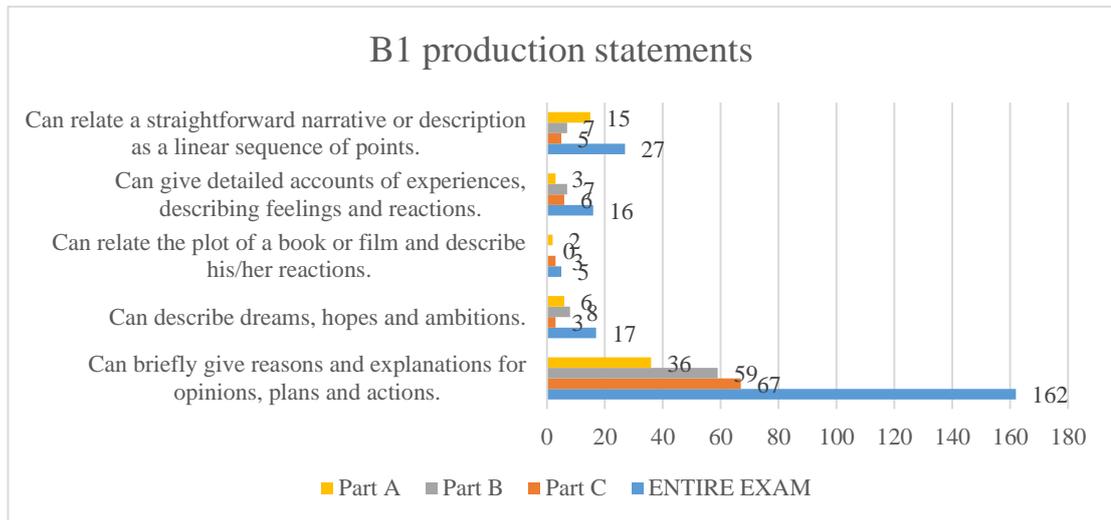


Figure 39: B1 interaction statement occurrence

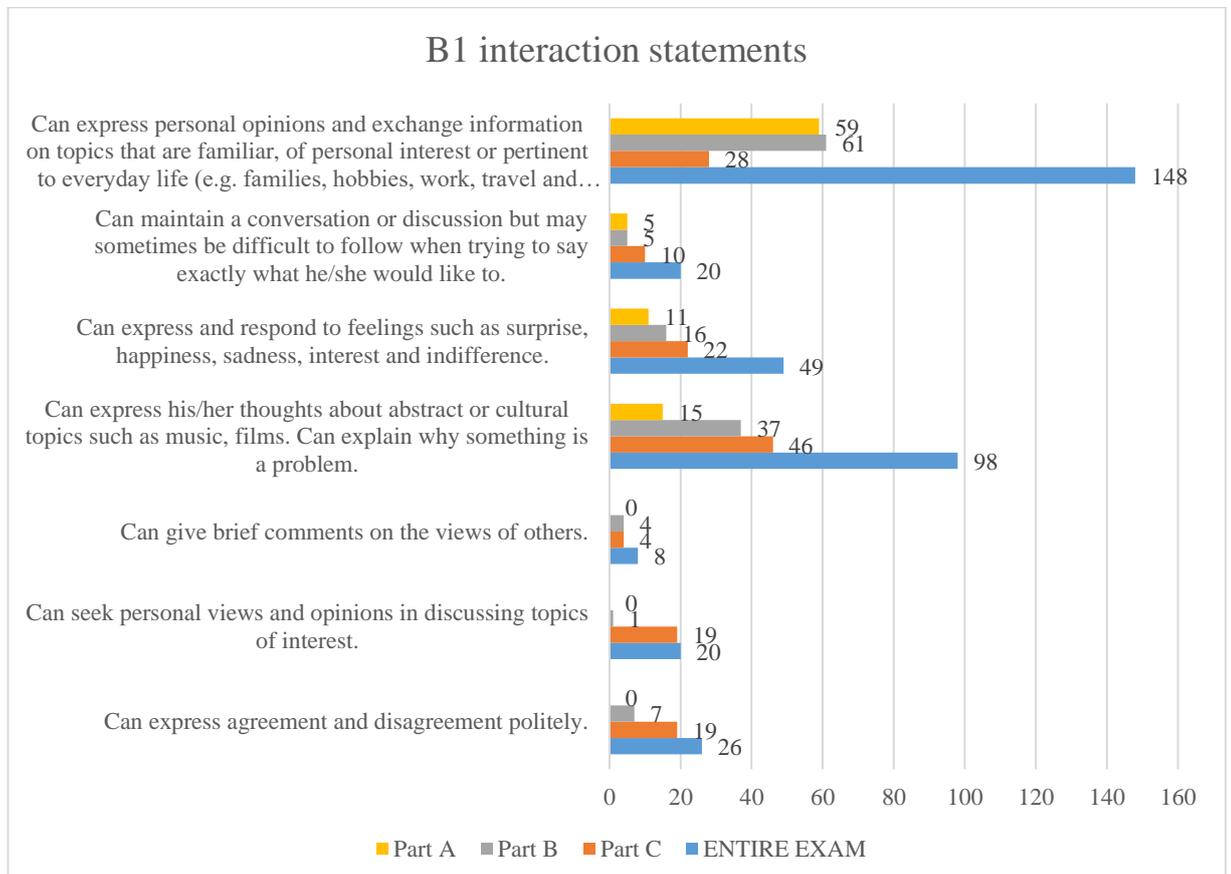


Figure 40: B1 production strategy statement occurrence

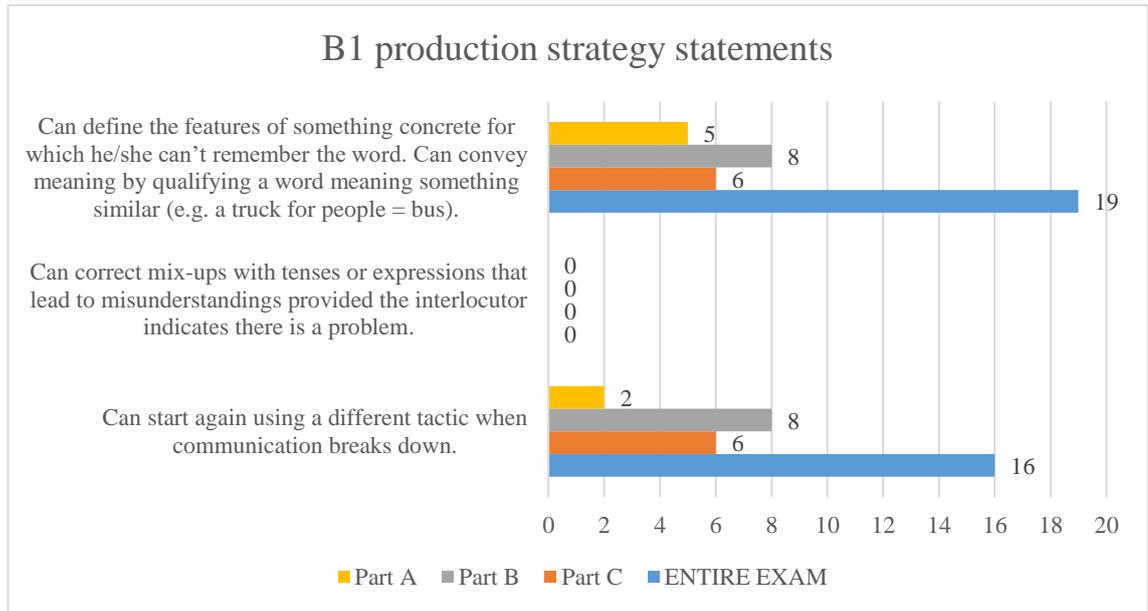
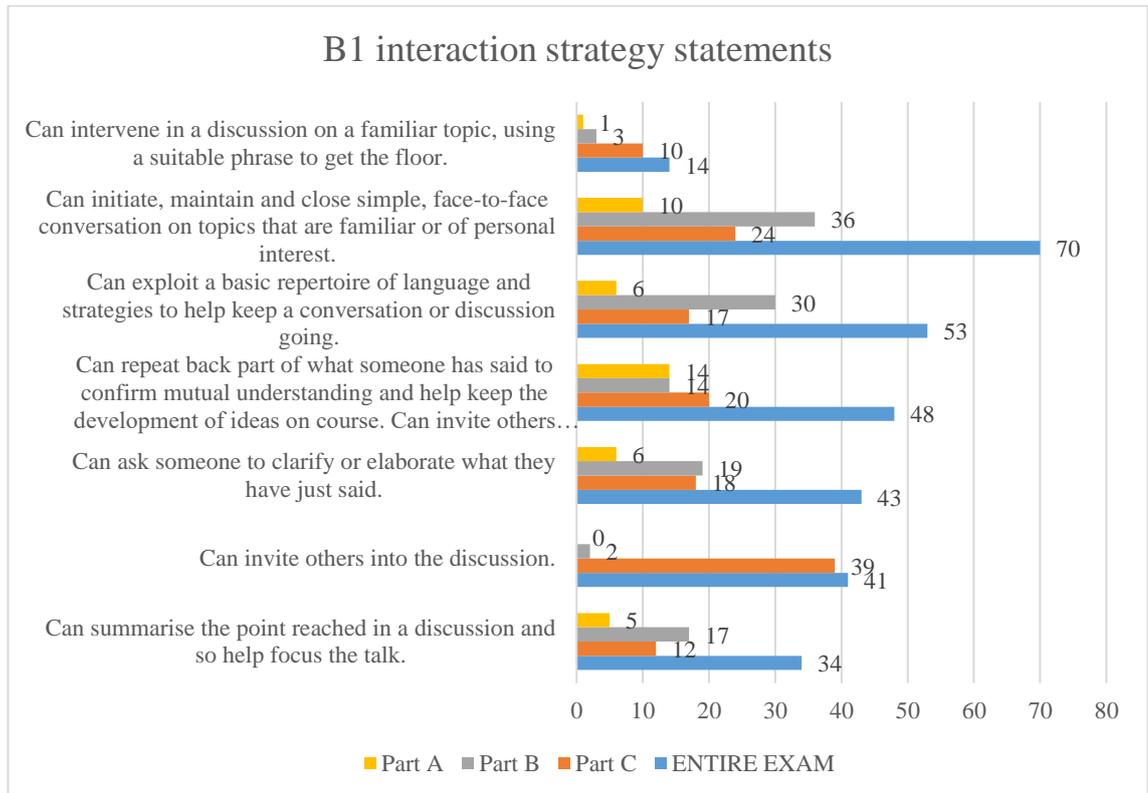


Figure 41: B1 interaction strategy statement occurrence



B1 learners provided a total of 934 statements across the four categories. Of these, 24.3% belonged to the production category, 39.5% related to interaction, only 3.8% constituted a production strategy and 32.4% associated with interaction strategies. Given

that one of two key characteristics of B1 learners is their ability to “maintain interaction and get across what [they] want to” (see CoE, 2001: 34), it is perhaps unsurprising that nearly three quarters of statement occurrence concerned interaction or interaction strategies. However, whilst the ability to express ‘main points’ or personal opinions was demonstrated on 148 occasions across the 15 exams, the ability to seek personal opinions was exhibited in only 20 instances (see Figure 39 interaction). The learners in this study, therefore, achieved a degree of success due to their overwhelming capacity to *convey* information about their personal views, but less due to *eliciting* those of others. This initial finding is further supported by the high numbers of reasons and explanations given for opinions (see Figure 38 production) and the relatively high amount of accounts offered for sustaining such views (see Figure 41 interaction strategies).

6.5.2 B2 can-do occurrence

B2 can-do occurrence is displayed in figures 42-45.

Figure 42: B2 production statement occurrence

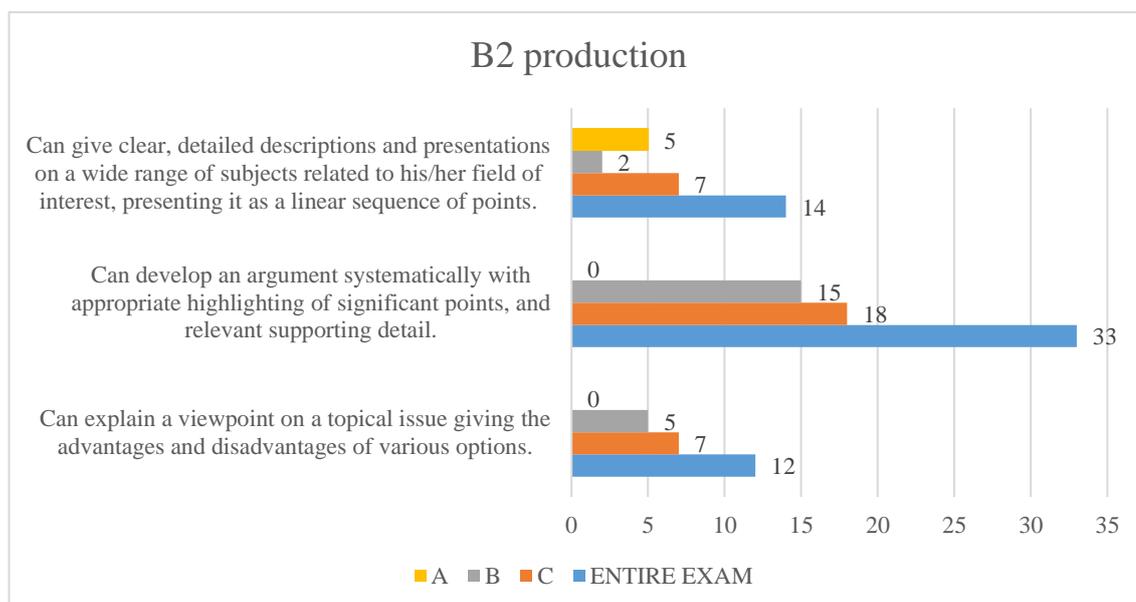


Figure 43: B2 interaction statement occurrence

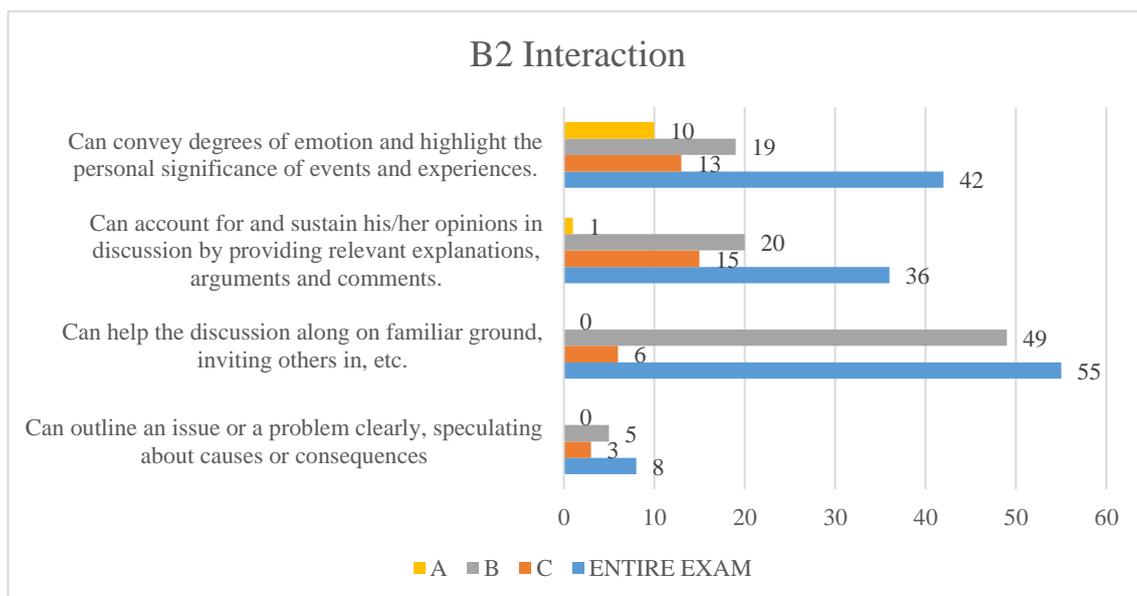


Figure 44: B2 production strategies statement occurrence

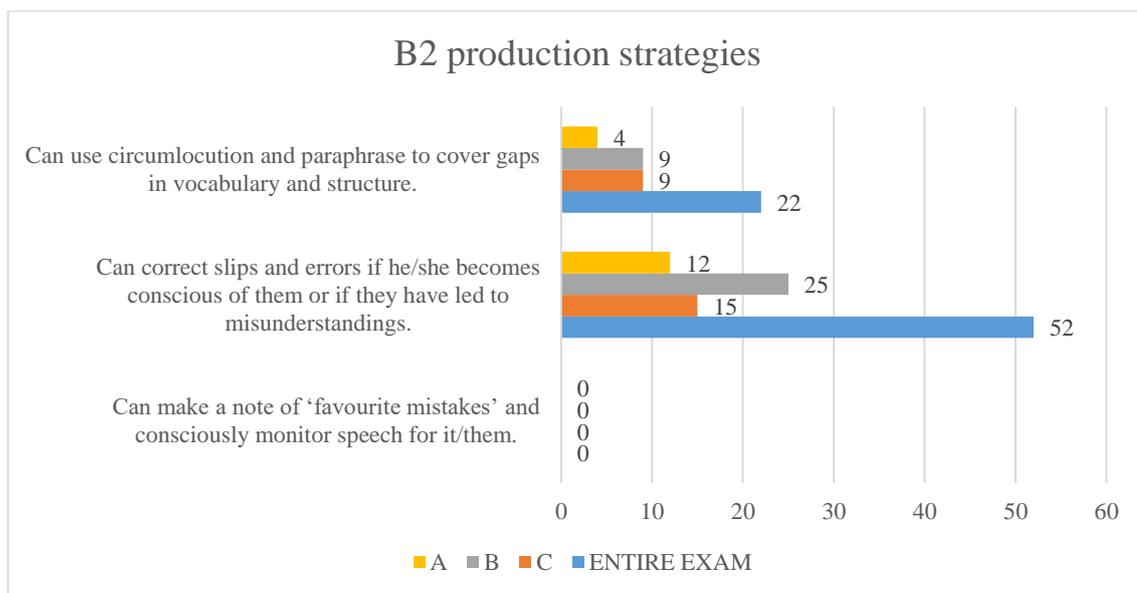
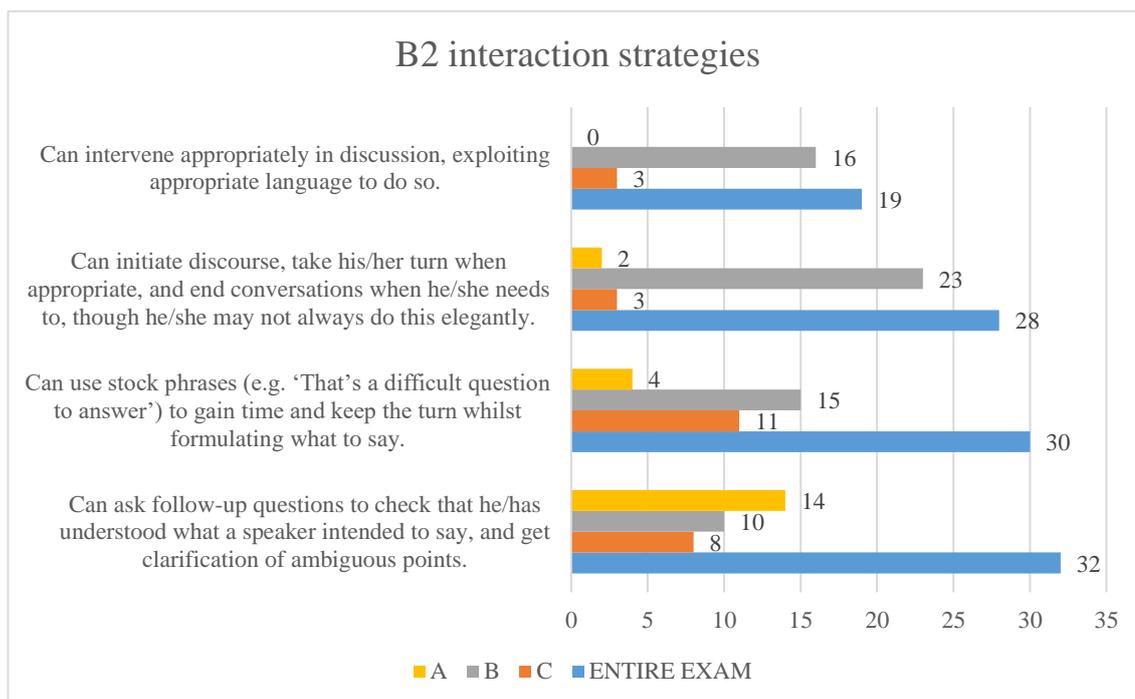


Figure 45: B2 interaction strategies statement occurrence



B2 learner language carried out the selected CEFR statements on 383 occasions. These represented 15.4%, 36.8%, 19.3% and 28.5% of the production, interaction, production strategy and interaction strategy categories, respectively. Once again, as was seen at B1, interaction-related statements dominated figures by occupying two-thirds of all statements evidenced in learner language. The ability to ask follow-up questions (32, a third of all interaction strategy occurrence, see Figure 45) and the ability to develop discussion on familiar ground (55. 39% of interaction statements, see Figure 43) further illustrated the importance of learner capabilities to engage in reciprocal, jointly constructed conversation for success. However, a key difference between B1 and B2 was the percentage of production strategies found. Increasing by 15.5%, B2 learners showed how they do, in fact, exhibit a “new degree of language awareness” (CoE, 2001: 35). Though no ‘favourite mistakes’ were identified, learners did correct slips and errors on 52 occasions which could signify one way in which they demonstrate the ‘new found’ perspective on the target language which is responsible for their labelling as ‘vantage’ learners (CoE, 2001). Finally, as appropriacy of conversational abilities are also emphasised at this level, the data was also able to show that over 40% of production strategy occurrence concentrated on intervening appropriately in discussion, and on initiating, participating and ending conversation. This analysis initially confirms that

successful B2 learners in this study complied with the descriptions of typical B2 learners provided by the CEFR.

6.5.3 C1 can-do occurrence

Figures 46-49 present can-do occurrence at C1 level.

Figure 46: C1 production statement occurrence

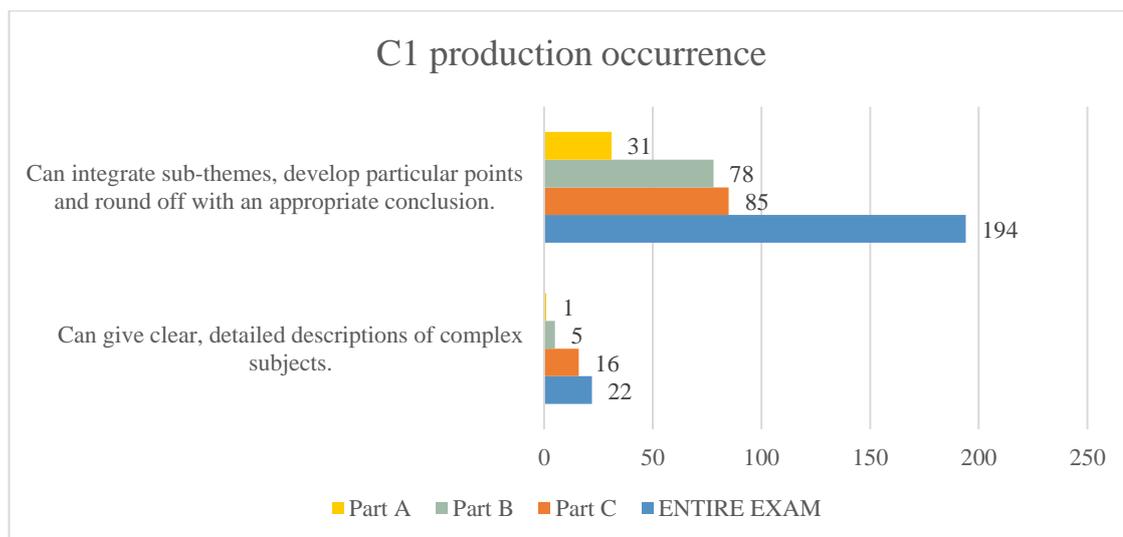


Figure 47: C1 interaction statement occurrence

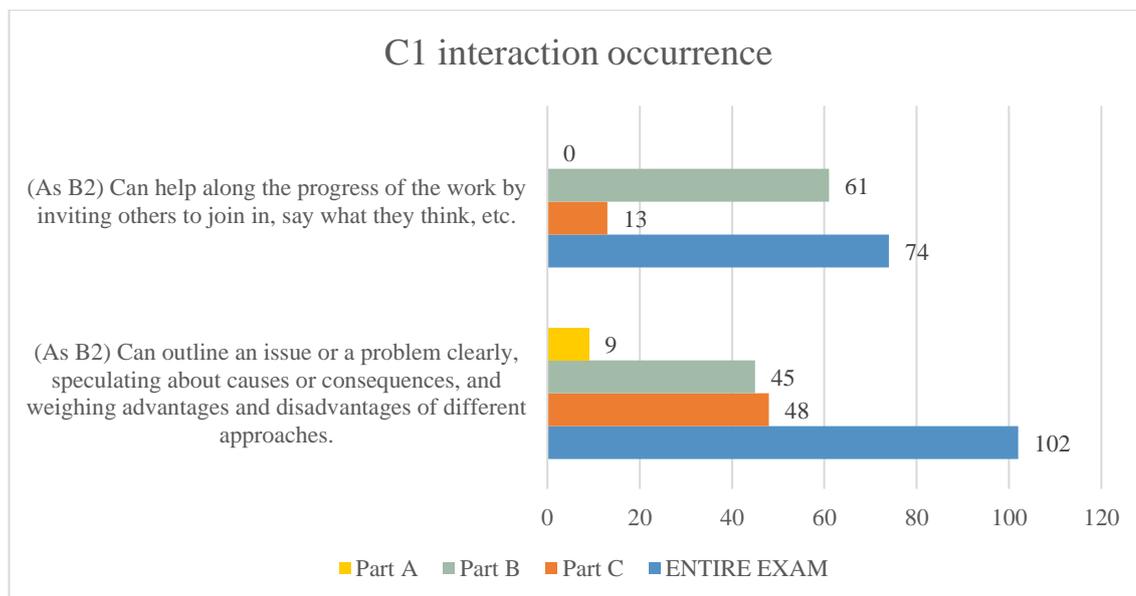


Figure 48: C1 production strategy statement occurrence

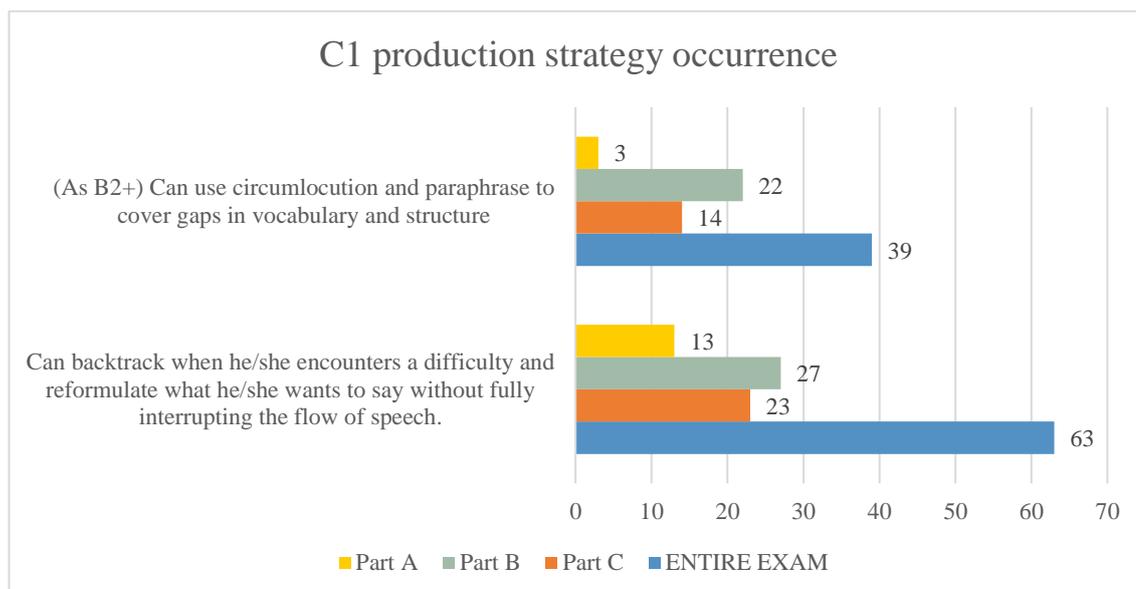
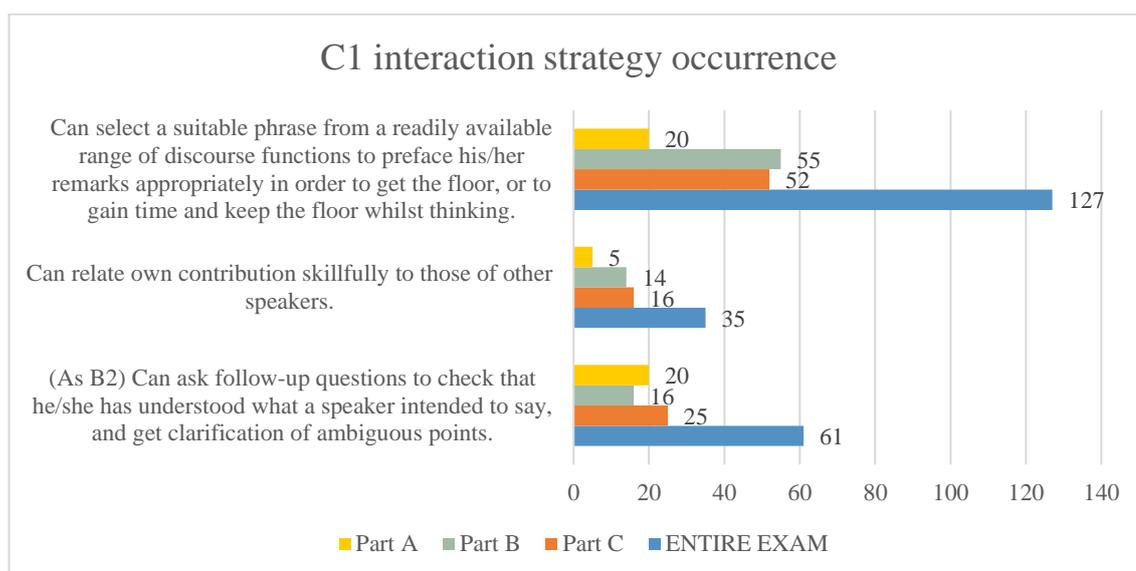


Figure 49: C1 interaction strategy statement occurrence



At C1 level, production formed 30.1% of all 717 instances. Interaction, production strategies and interaction strategies comprised 24.1%, 14.2% and 31.1%, respectively. Interestingly, despite the fact that this level had the lowest numbers of selected statements for analysis, the learners within it produced the highest amount of statement evidence. Immediately, therefore, it could be argued that the statements dominating each of the four categories were fundamental to learners’ overall spoken success. For instance, descriptions of C1 ability in the CEFR highlight the use of suitable phrases for prefacing remarks and the “controlled use of organisational patterns, connectors and cohesive

devices” (CoE, 2001: 36). In fact, 127 of interaction strategy occurrence, over half of that category, related to such an ability whilst 194 of the production statements, over 90%, related to the integration of sub-themes, development of point and the use of conclusions. With such high proportions, these abilities were arguably central to C1 learner success in this study. On the other hand, unlike CEFR descriptions of B1 and B2 learners, C1 level descriptions focus more on aspects of quality, such as fluency and spontaneity, than the can-do statements specified (see CoE, 2001: 34-36). Though this study can relate abilities to these qualities, thorough examinations of qualitative features may require a change in research approach beyond the scope of this study.

6.5.4 A closer comparison of abilities and language at B1, B2 and C1

Sections 6.5.1-6.5.3 ascertained the central traits of successful learners in accordance with CEFR descriptions. B1 learners were characteristic in their interaction, albeit in *expressing* more than *seeking* opinions; B2 learners demonstrated an ability to engage in discussion on familiar ground and demonstrated a heightened awareness of accuracy and appropriacy; and C1 learners had the capacity to preface remarks suitably during interaction whilst enhancing utterances via the integration of additional information. The next stage of analysis involved comparing the most frequent statements at each of the three levels to reveal whether learners evidenced similar abilities in production, interaction, production strategies and interaction strategies. As well as exposing which descriptors were of vital importance to learner success at each level, this step would also act as a precursor to comparisons of the learner language used to realise them as, due to the amount of data, only selected statements could be analysed in such depth. Tables 54-57 display the frequency and proportion of prevalent B1, B2 and C1 statements per CEFR category. The trends in this data will be quickly summarised before Sections 6.5.5-6.5.12 discuss them more thoroughly.

Table 54: Most frequent production statements

CEFR category	Level	Highest-occurring statement	Freq.	% of category
Production	B1	Can briefly give reasons and explanations for opinions, plans and actions.	162	71.37%
		<\$5F> I feel lonely and er and er a bit of scared mm and er because er this is a this is a er unfamiliar city to myself so erm I er I can't go out myself because I think er I will be lost in the way [Exam 3]		
	B2	Can develop an argument systematically with appropriate highlighting of significant points, and relevant supporting detail.	33	55.93%
		<\$11M> I I er as far as I'm concerned both of music and er and sports are very important subject in school er er first of all music can devel= er can develop your mental your mental and your psychological and sports can exercise your body makes your f= erm h= er makes your body makes your figure some awesome. [Exam 6]		
	C1	Can integrate sub-themes, develop particular points and round off with an appropriate conclusion.	194	89.81%
		<\$16F> Oh yeah erm I think foreign language we start in the school is very mm important because in our college we have er international students if we know the foreign language we can communication with the others er who are come from other countries we can know their culture and can make friends with him so I think learn the foreign language is very useful in our study life. [Exam 8]		

Table 55: Most frequent interaction statements

CEFR category	Level	Highest-occurring statement	Freq.	% of category
Interaction	B1	Can express personal opinions and exchange information on topics that are familiar, of personal interest or pertinent to everyday life (e.g. families, hobbies, work, travel and current events).	148	40.11%
		<\$25F> If I want to practise my speaking or listening I would choose listening to music but if I just do some reading or writing essay I have to in silence place and think about mm more more things yes I need silence when I do writing and reading. [Exam 13]		
	B2	Can help the discussion along on familiar ground, inviting others in, etc.	55	39.01%
		<\$4M> what advice would you give er to a foreign student who is suffering from culture shock? [Exam 2]		
	C1	(As B2) Can outline an issue or a problem clearly, speculating about causes or consequences, and weighting advantages and disadvantages of different approaches.	102	57.95%
		<\$2M> I said if it's near to the attraction point then that's good but if it's far away then you know it will be costly to travel up. [Exam 1]		

Table 56: Most frequent production strategy statements

CEFR category	Level	Highest-occurring statement	Freq.	% of category
Production strategies	B1	Can define the features of something concrete for which he/she can't remember the word. Can convey meaning by qualifying a word meaning something similar (e.g. a truck for people = bus).	19	54.29%
		<\$10F> <i>They costumes like Harry Potter and Hermione so it was very nice. Also there are there were magic sticks</i> [Exam 5]		
	B2	Can correct slips and errors if he/she becomes conscious of them or if they have led to misunderstandings.	52	70.27%
		<\$29F> <i>Erm. I think this er this is erm I I think I agree with you on the first part but I disagree with you in the er sorry I disagree with you in the first part but I agree with you in the interesting job</i> [Exam 14]		
	C1	Can backtrack when he/she encounters a difficulty and reformulate what he/she wants to say without fully interrupting the flow of speech.	63	61.76%
		<\$29M> <i>It depends. It depends on many things. First of all it's <\$=> it's about the <\$=> like some trains now they've got er lock= locks on the er on the doors and every door there's lock they they can't control the door unless they stop.</i> [Exam 14]		

Table 57: Most frequent interaction strategy statements

CEFR category	Level	Highest-occurring statement	Freq.	% of category
Interaction strategies	B1	Can initiate, maintain and close simple, face-to-face conversation on topics that are familiar or of personal interest.	70	23.10%
		<\$24F> <i>if you don't work hard you can't do your best and er you can't er er be the erm one in your job so that's why you should work hard.</i> [Exam 12]		
	B2	Can ask follow-up questions to check that he/has understood what a speaker intended to say, and get clarification of ambiguous points.	32	29.36%
		<\$21F> <i>So in your opinion you prefer do some indoor sports?</i> [Exam 11]		
	C1	Can select a suitable phrase from a readily available range of discourse functions to preface his/her remarks appropriately in order to get the floor, or to gain time and keep the floor whilst thinking.	127	56.95%
		<\$12F> <i>Okay I take your point but erm this is some problem about the bins</i> [Exam 6]		

This analysis uncovered several themes. Relating to production, being successful at B1, B2 and C1 necessitated an ability to expand on opinions, admittedly to varying degrees: *briefly* at B1, *systematically* at B2 and *appropriately* both at B2 and C1. The latter two levels also denote a degree of continuation. Whereas B1 learners were successful in supporting thoughts or actions, B2 and C1 learners were able to evidence a sustained effort to develop or extend a particular thought or theme with supporting detail. Regardless of level, the essence of these data signifies that it is simply not sufficient to

just give opinions at B1, B2 or C1; to be truly successful, learners must be able to employ language that reinforces or enhances them.

In terms of repeated interaction, can-do statements did diverge in terms of goal. B1 students tended to concentrate on the delivery of information whilst B2 learners changed tack by making attempts to uphold the interactional nature of conversation. The emphasis on turn-taking at B2 (see CoE, 2001: 28) is evidenced by this data. At C1, on the other hand, interaction centred once again on the content of the dialogue. Perhaps due, to some extent, to the ‘complex’ nature of the subject-matter involved, greater critical thinking was demonstrated regarding advantages, disadvantages, or causes and consequences of issues or problems. In short, the most numerous statements in the interaction category infer that success at B1, B2 and C1 relies on the ability to *give* opinions, *seek* opinions and *elaborate* on opinions, respectively, though it is evident that the nature of the research tool would have also influenced this finding.

Production strategies once again differentiated B1 learner performance from the other two levels. As is acknowledged in Taylor (2011), lower level learners more typically attend to linguistic rather than pragmatic features. However, all three levels utilised production strategies for unknown vocabulary (B1) or for self-correction (B2 and C1) suggesting that they did, in fact, conform to CEFR descriptors: B1 learners were able to ‘get by’ via the use of circumlocution whilst B2 and C1 learners demonstrated an ability to correct errors, without the interlocutor’s assistance, to avoid misunderstandings. Turning attention to interaction strategies, B1 and C1 learners also both exhibited their pragmatic abilities in the sense that turns were started, ended and, when necessary for C1 students, maintained during conversation, thus coinciding with CEFR qualitative aspects for both levels. On the other hand, B2 learners’ success seemed to rely more on the ability to confirm comprehension of both the native speaker interlocutors and fellow learners in the exam. Though this initially corresponded to CEFR descriptions of interactive ability, Figure 45 on page 217 does demonstrate that a range of interaction abilities were in fact evidenced by this group of learners.

6.5.5 Production

The most common production statements identified in Table 54 related to the reasons, arguments and sub-themes learners used to reinforce discussions of opinions, issues and problems. However, since analysis into this ability would overlap considerably

with the most frequent *interaction* statements, the decision was taken to focus on the second-most common production descriptors which were found 27, 14 and 22 times at B1, B2 and C1, respectively:

Figure 50: Second-most frequent production statements (CoE, 2001: 58-59)

<p>B1 <i>Can relate a straightforward narrative or description as a linear sequence of points.</i></p> <p>B2 <i>Can give clear, detailed descriptions and presentations on a wide range of subjects related to his/her field of interest, presenting it as a linear sequence of points.</i></p> <p>C1 <i>Can give clear, detailed descriptions of complex subjects.</i></p>
--

Previous analysis in RQ2a-c has shown how ideas were linked together at B1, B2 and C1, but it did not focus on the growing levels of detail across proficiency levels. Can-do statement analysis therefore allowed insights to be gained into how such detail was realised in learner speech.

Using corpus-driven techniques once again, each example of the individual statements in Figure 50 above was read and a list of potential cohesion devices was created from the data. Linking devices were then checked for similarity and refined into 26 categories:

Figure 51: Cohesion device list

<ul style="list-style-type: none"> • <i>Although</i> • <i>And</i> • <i>And also</i> • <i>And then</i> • <i>As</i> • <i>Because</i> • <i>But</i> • <i>Cos</i> • <i>So</i> • <i>So yeah</i> • <i>Moreover</i> • <i>Basically</i> • <i>Also</i> 	<ul style="list-style-type: none"> • <i>Personal pronouns</i> • <i>Relative pronouns</i> • <i>Demonstrative pronouns</i> • <i>Adverbs of place</i> • <i>Repetition</i> • <i>Related vocabulary</i> • <i>Lists</i> • <i>Such as</i> • <i>For example</i> • <i>Like</i> • <i>The thing is/The one this</i> • <i>Even if</i> • <i>If</i>
---	--

Each statement was then analysed to see which, and how many, cohesion devices were utilised, see Table 58.

Table 58: Cohesion devices across B1, B2 and C1

Level	Can-do statement freq.	No. of device types identified	Total no. of devices	Freq. of device (descending order)			Mean words per utterance (to the nearest whole word)
				Device	Freq.	%	
B1	27	11	87	And	44	50.57	40 (SD = 20.15)
				Personal pronouns	10	11.49	
				Related vocabulary	7	8.05	
				Adverb of place	6	6.90	
				Demonstrative pronoun	6	6.90	
				Repetition	6	6.90	
				But	3	3.45	
				If	2	2.30	
				Because	1	1.15	
				For example	1	1.15	
				Relative pronoun	1	1.15	
B2	14	15	104	And	30	28.85	91 (SD = 28.99)
				Because	13	12.50	
				Demonstrative pronoun	13	12.50	
				But	11	10.58	
				Personal pronouns	11	10.58	
				So	10	9.62	
				Repetition	8	7.69	
				Adverb of place	1	0.96	
				Although	1	0.96	
				Also	1	0.96	
				And also	1	0.96	
				For example	1	0.96	
				Like (for giving examples)	1	0.96	
				So yeah	1	0.96	
Such as	1	0.96					
C1	22	19	159	Personal pronouns	49	30.82	82 (SD = 43.96)
				So	21	13.21	
				And	15	9.43	
				But	12	7.55	
				Because	11	6.92	
				Like (for giving examples)	11	6.92	
				Repetition	10	6.29	
				Demonstrative pronoun	9	5.66	
				Relative pronoun	3	1.89	
				So yeah	3	1.89	
				And also	2	1.26	
				As	2	1.26	
				Basically	2	1.26	
				Cos	2	1.26	
				Even if	2	1.26	
				For example	2	1.26	
And then	1	0.63					
If	1	0.63					
The/One thing is	1	0.63					

This analysis revealed two main differences. Firstly, the conjunctions employed changed notably across the levels. The most frequent conjunction *and* represented over half of the cohesive devices in the statements at B1. At B2 it comprised a portion of less than 30% and at C1, it constituted less than 10%. As previously highlighted in RQ2a (Section 6.2.4), spontaneous speech often likens the addition of utterances or phrases to links in a chain (Carter & McCarthy, 2006). Whereas at B1 these links equated to the addition of ideas via *and*, at B2 and C1, relationships such as contrast or cause-and-effect, e.g. *but*, *because* and *so*, were utilised (see Extracts 49-51). Secondly, the final column of Table 58 depicts a rise in mean utterance length. Though not all of the words will have contributed fully or successfully to the message being conveyed, statistical tests showed that mean length increased significantly from B1 to C1 ($p < 0.01$) but not between B2 and C1. The findings indicate that being successful at B2 or C1 level required not only a change in the relationships of combined ideas, but it also required a rise in utterance length. In this study, therefore, the straightforward, coordination of points at B1 were rather short and focussed mostly on addition, whereas detailed descriptions of topics, at B2 and C1, required different cohesive links, such as subordination, and more words.

Extract 49: Exam 3, B1

*<\$5F> Erm er <\$G2> <\$=> in </\$=> there are many lakes and rivers in <\$G2> and er **and** er my my hometown is famous for its delicious food such as hot dry noodles and er duck neck **and** er a building named and a building named erm yellow erm yellow tower yellow tower and crying tower.*

Extract 50: Exam 8, B2

*<\$16F> I'd like to go to South Africa. Yeah **because** I have I have been I have been here at two years ago **and** I think there is the most beautiful beautiful country I ever go **but** erm but the guy told I cannot go out alone at er afternoon **because** it's very dangerous **but** the view is really really beautiful. It just near the sea **and** the ocean is very blue I never see the blue like that. It's very really beautiful.*

Extract 51: Exam 9, C1

*<\$25F> Erm I guess another problem in Cairo is the smoke. We've got lots of factories <\$=> which are </\$=> really they are at the boundaries **but** they are so close to the tow= to the city itself **so** it it made that black cloud **so** they're trying to fix it **but** it's not yet fixed by putting filters on the um pipes and stuff **and** er try to reduce they carbon footprint.*

Finally, analysis of these production statements uncovered another difference. Between B2 and C1, distinctions were less prominent but a distinction did exist in terms of the pronouns used. C1 learners displayed a total of 61 pronouns (49 personal pronouns, 9 demonstrative pronouns and 3 relative pronouns), which constituted 38% of all devices

used. B2 students displayed a total of only 24 pronouns (13 demonstrative pronouns and 11 personal pronouns), forming 23% of all devices. At C1, such pronoun usage fluctuated between subject and object pronouns while at B1 and B2 levels, only one instance comprised an object pronoun. Whereas B1 and B2 students used subject pronouns in all but one instance to refer to the main point of the utterance, C1 learners were able to flexibly broaden pronoun usage to refer to aspects beyond the subject being discussed, for example countries' laws and the problems they generated. To be successful, C1 learners in this study needed to demonstrate greater control and once again flexibility, this time in respect to their pronoun usage. They displayed greater variation in the pronouns used, as well as the item being referred to (see Extracts 52 and 53).

Extract 52: Exam 5, B2

*<\$IOM> Of course. Yes erm maybe maybe you don't know my project is film production so erm and maybe you don't know er a Chinese director called er <\$G3>. I think he er **he** is my er a= erm how to say erm idol yes idol because erm this person erm finish a lot of successful f= er film in when **he** was young. I think this is er this is talent yes and er also when I er watched **him** film erm **it** can change my mind I get some inspiration and the creation yes.*

[he and him clearly refer to the film director]

Extract 53: Exam 2, C1

*<\$3M> Basically it's not allowed in the er law in my country but even if you're fifteen or thirteen **they** could drive around in a car and if like a policeman should stop you or a road safety person should stop you you could bribe **them** like really low amounts like anybody could afford **it** and **they** will let you go so **it** causes a lot of er accidents **it** causes people driving recklessly and everything and er yeah **they** are not strict at all with the laws so **it** causes a lot of problem and also the roads are really bad and the government hardly does anything about **it** just like a few states try to fix the roads and make **them** good most of the other states are really bad which cause other accidents as well.*

[‘they’ refers to young people, ‘them’ and ‘they’ refer to policemen, ‘it’ refers to offering bribes, ‘it’ also refers to the problem caused, and ‘them’ refers to roads]

6.5.6 Production statement summary

Analysis of the second-most frequent can-do statement for production has demonstrated that successful learner speech at all three levels relates to CEFR descriptions in three ways. Firstly, the linear or detailed sequences mentioned at B1, B2 and C1 were shown to exhibit a change in conjunction usage and the meaning conveyed. O’Keeffe and Mark’s (forthcoming) paper on the methodology involved in calibrating learner grammar competencies for English Grammar Profile presents a list of criteria for

establishing whether descriptions can be accepted and are thus representative of learners at different levels. Amongst others, these include the frequency, accuracy, and spread of uses. Whilst this study cannot be representative of *all* B1, B2 and C1 learners, the lists in RQ2a and RQ2b demonstrate that conjunction usage was very frequent, sometimes key and high in range across the UCLanESB levels. The change from simple addition of ideas to the incorporation of contrast and cause-and-effect utterances was therefore found to be a noticeable difference from B1 to C1. Though discourse may be simplistic in successful spoken grammar, the meanings conveyed can vary. Secondly, successful learner speech has been able to shed light on the CEFR's vague use of the word 'detailed' at B2 and C1. It was shown that after B1, detail was exemplified through the length of utterances which significantly increased in the number of words they contained. Thirdly, though utterance length was found to be a distinguishing feature between B1 and B2, the flexibility demonstrated via cohesive devices such as subject and object pronouns was distinct between B2 and C1. Not only did object pronouns at C1 occur significantly more frequently, but they also varied in the items referred to.

In one way, the findings have confirmed that B1 learners' coherence is evidenced by "a series of shorter, discrete simple elements" and for B2 learners who are said to employ a "limited number of cohesive devices", Table 58 has shown that learner abilities should not be underestimated (CoE, 2001: 28-9). However, beyond the production statement analysed, the coherence descriptor for C1 (see CoE, 2001: 28) does not explicitly refer to the number of devices nor length as the CEFR does for the other two levels. It instead highlights the controlled use C1 learners should display. The findings from successful learner speech in this study are unable to comment on accuracy but the degree of flexibility evidenced shows that C1 learners did have more control than B1 and B2 learners.

6.5.7 Interaction

Due to the nature of the language tool used (see Section 5.2.1) it was clear that the degree of success achieved by learners would be linked to their ability to give opinions. Table 55's discussion earlier revealed that a progression was evident from simply giving opinions (B1), to seeking opinions (B2) and to elaborating more fully on the reasons behind them. However, since opinions were integral to both production and interaction statements, further investigation was required. The statements below (B2 amended to

increase comparability) were used as a basis for examination, in particular, to ascertain how opinions developed from being expressed, to being accounted for and to being speculated upon:

Figure 52: Opinion-related interaction statements (CoE, 2001: 74, 77, 79)

B1

Can express personal opinions and exchange information on topics that are familiar, of personal interest or pertinent to everyday life (e.g. families, hobbies, work, travel and current events).

B2

Can account for and sustain his/her opinions in discussion by providing relevant explanations, arguments and comments.

C1

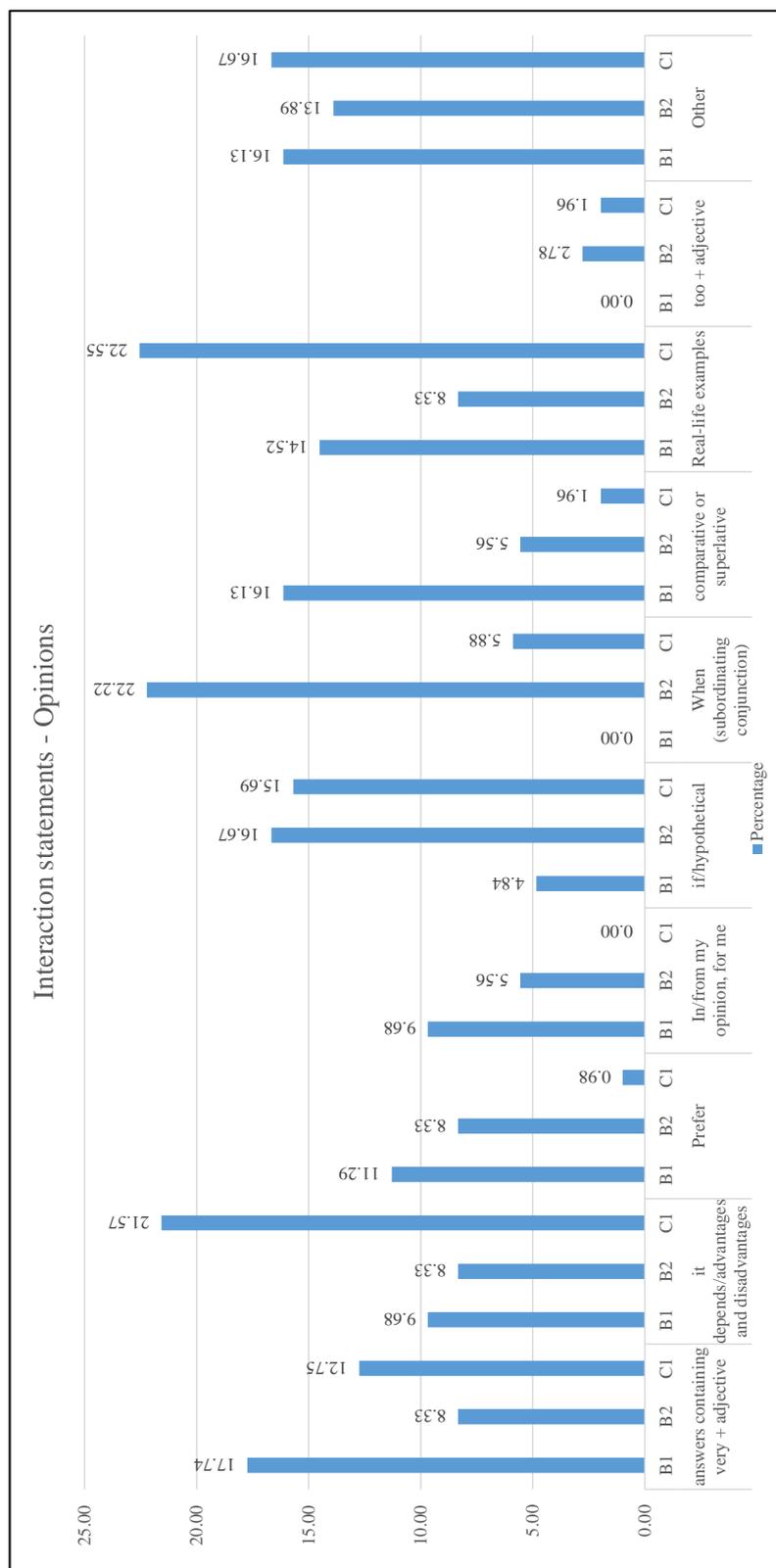
Can outline an issue or a problem clearly, speculating about causes or consequences, and weighing advantages and disadvantages of different approaches.

A similar approach to that adopted in the earlier category was followed. Utterances at all levels were read once so that a list of salient structures or patterns could be elicited. Once established, utterances were assigned to one of the categories; an ‘other’ category was added for nondescript examples. Since B1 contains two slightly different abilities, utterances were also separated according to whether they concerned the giving of opinions or the exchange of information; only the former set were used in this analysis. Table 59 below shows the numbers assigned to each category at the three levels whilst Figure 53 displays the proportions each category composed.

Table 59: Opinion patterns at B1, B2 and C1

Pattern	Level	Number	Example
Answers containing very + adjective	B1	11	<i>Er the experience I think is very important so we need learn some some skill we er another work (B1, Exam 1)</i>
	B2	3	
	C1	13	
It depends/advantages and disadvantages	B1	6	<i>they deal with them in a way that they are so rich and this is like in some ways offensive to them. However, the er the good things are that you know as I said that they are many cultures in one place (C1, Exam 2)</i>
	B2	3	
	C1	22	
Prefer	B1	7	<i>Er actually er I prefer smartphone and er other devices which is small than computer (B2, Exam 15)</i>
	B2	3	
	C1	1	
In/from my opinion, for me	B1	6	<i>because I from my opinion er I think all materials is er significant erm important to to study (B1, Exam 12)</i>
	B2	2	
	C1	0	
If/hypothetical	B1	3	<i>Yeah because if you're a millionaire you can stay at home and not be social sociable so if you work you're gonna have social life you can talk to people (B2, Exam 14)</i>
	B2	6	
	C1	16	
When (subordinating conjunction)	B1	0	<i>When you want to find out about history a few people about the few people are left in the world so you can't really find a lot of people to interview (C1, Exam 13)</i>
	B2	8	
	C1	6	
Comparative or superlative	B1	10	<i>Er I think er the better is house because the house is very big and er can I live with my all my family er and can to able more communicate with them in er apartment er (B1, Exam 8)</i>
	B2	2	
	C1	2	
Real-life examples	B1	9	<i>I'm not agree with the this sentence because I think that young people find it difficult most of young people find it difficult too for me I find it difficult. I asked my sister every time er how to do that how to do this... (C1, Exam 9)</i>
	B2	3	
	C1	23	
Too + adjective	B1	0	<i>as there are too many attraction points for tourism in our country (C1, Exam 1)</i>
	B2	1	
	C1	2	
Other	B1	10	<i>Erm people er people go to the shop or café they just want to relax theirself and er and this place can provide their er their relaxing atmosphere (B2, Exam 6)</i>
	B2	5	
	C1	17	
Total	B1	62	
	B2	36	
	C1	102	

Figure 53: Proportions of opinion patterns at B1, B2 and C1



B1, B2 and C1 students utilised a variety of patterns when giving opinions. In terms of majorities, B1 students prefaced remarks (as they did in Section 6.3.1) rather often with the *very + adjective* structure, stating that something was *very important* or

very useful, for example, but doing very little in terms of expanding upon why they held that opinion. B2 students on the other hand opted most often for subordinating conjunctions beginning with *when* (see Carter & McCarthy, 2006) so as incorporate a real or imaginary condition for when an opinion would be valid, e.g. *but sometimes when you travel and you ever never know any any er word..you can't cope with the new culture* (B2, Exam 12). C1 students, demonstrated very similar usage in two different categories: that of saying that something depended on something else or highlighting advantages and disadvantages, for instance *Cos junk food well it's bad but you know it's still giving you something* (C1, Exam 12), or that of using a real life example to illustrate a point, e.g. *the environment was destroyed almost but unfortunately my country's government didn't recognise that* (C1, Exam 7). Interestingly, some of the patterns also increased or decreased across levels. Figure 53 shows how overt opinion-related lexis such as the verb *prefer* or discourse markers such as *in my opinion* reduced heavily from B1 to C1. By contrast, *it depends* (along with advantages and disadvantages) and conditional structures involving *if* saw dramatic changes. This suggests that increasing proficiency could be affected by shifts towards or away from certain structures

The interaction statements in Table 55 highlighted that there was also a need to seek opinions. As B2's most frequent interaction statement (see Figure 43), investigation would reveal whether B1, B2 and C1 learners employed very similar question types or whether a change was evident at either of the levels. Appendix 13 displays question types and frequencies for all utterances satisfying the statements. Here, however, only a summary of the trends identified will be given,

An immediate observation concerned the numbers involved in the analysis. From B1 to C1, the total of all questions asked doubled from 36 to 72; intriguingly, the total grew by an interval of 18 both times from B1 to B2, and from B2 to C1. Clearly, therefore, success at higher proficiency levels demanded a greater attempt to interact with others involved in spoken communication via the use of questions. Similarly, as well as frequency, the *range* of question types asked may offer another indication of increased proficiency. Though question type figures dropped to 11 at B2, B1 learners evidenced 16 question types while C1 learners employed 20. Given that only 4 questions occurred on more than 2 occasions at B1, compared with 8 at C1, there may be some evidence to argue that questions for maintaining and achieving interaction with interlocutors should demonstrate more flexibility at higher levels. This, however, should not be confused with a greater level of accuracy. Though the CEFR (CoE, 2001: 28) declares that C1 learners

should maintain “a high degree of grammatical accuracy” with “rare, difficult to spot” errors, analysis revealed that some questions were ill-formed, *how do you think?* occurring 4 times at C1 being a case in point. A final observation involved the responses sought using the two most common techniques at the three levels. The can-do statements earlier in Figure 39 revealed that B1 learners were able to obtain opinions from others, or involve them in the conversation, whilst B2 and C1 learners were able to ‘help the discussion along’. The two most frequent question types, see Table 60, used at each level hence provide some explanation of i) the way this was achieved and, ii) the difference between the two. As can be seen, *do/did you...?* was the most common at B1 and B2 comprising a quarter of all questions asked; at C1, a quarter of all questions asked used the form: *what do you think about/of...?* Though admittedly, learners asked the former did offer more than a one-word answer, it nevertheless is formed as a yes/no question. The latter, however, clearly asks more of learners in that it cannot be answered yes or no, or with a one-word answer; it clearly asks for thoughts behind an opinion. B2 learners incorporated this question type into more of their speech, but it was only at C1 that it dominated as an interaction feature. Since it was used infrequently at B1, analysis of this study’s learner language has shown that *frequency* and *type* of question can signal a rise in proficiency.

Table 60: Two most common question types at B1, B2 and C1

Question	Freq. (%)		
	B1	B2	C1
<i>Do/Did you...</i>	9 (25%)	15(28%)	7 (10%)
<i>What do you think about/of...?</i>	3 (8.33%)	14 (26%)	17 (24%)

6.5.8 Interaction statement summary

Providing explanations for thoughts and opinions saw a range of approaches at B1, B2 and C1. However, whereas B1 learners exhibited the use of adjective phrases, B2 learners employed real-life examples and C1 learners expressed advantages, disadvantages and weighed up different sides of particular arguments. B1 learners’ most common tactic bears some resemblance to spoken grammar in that sufficient meaning

was conveyed using basic, more general adjectives, e.g. *important* or *useful* (McCarthy, 1999). B2 learners, however, demonstrated once again, as the discussion of *we* in Section 6.2.1 highlighted that discourse often implicated the learners' own contexts and experiences. At C1, more evaluative language such as 'it depends' itself showed a type of communicative routine that allowed learners to delve deeper into the justifications behind their thoughts. Similarly, interaction questions corroborate the earlier chunk finding that B1 learners did not seek opinions very much. The appearance of chunks not only suggests that the delivery of questions would have been smoother, but it also indicates that seeking opinions often followed the same 'routine' as was remarked by Bygate (1987). In addition, it is the researcher's opinion that questions evidently performed a strategic role. Though learners needed to interact with partners in the tasks, some learners clearly 'passed the buck'. Before the discussion, for example in Part B of the exam, had started, a question was asked that immediately put the partner under pressure. As a strategic function, speakers using this tactic would have more time to formulate a response. Just as the chunk *I agree with you* was not always fully genuine at B2, the use of questions across B1, B2 and C1 sometimes did not act as a genuine request for opinions.

6.5.9 Production strategies

Analysis of production strategies explored learner correction of their speech, one of the compensation strategies identified by Bygate (1987). Table 56 demonstrates that both B2 and C1 learners had the majority of their productive strategies in this category. Though B1 learners initially seemed to employ circumlocution and paraphrase more than correction, analysis proved otherwise.

For B1 level, the CEFR seems a little contradictory on the notion of repair. Though in freer production at this level repair is said to be "very evident" (CoE, 2001: 34), in the corresponding can-do statements, the CoE (2001: 65) states that corrections only occur "provided the interlocutor indicates that there is a problem". For this reason, no evidence of this strategy was identified in the preliminary analysis stage. However, ignoring this condition allowed greater comparison across the levels. In fact, had the statement been amended previously, correction in Figure 40 would have formed 49% of all production strategies, a much more comparable figure with B2 (70%) and C1 learners (62%) despite it still being lower. These percentages demonstrate that whilst B1 learners

did not correct their utterances as often as learners in the higher levels (in number nor proportion), the CEFR seems to do them a slight disservice by remarking that they were unable to monitor their speech independently. In terms of success, corrections and potential recasts are therefore to be expected at all three levels; though the degree of success in the correction itself remains undetermined.

The next stage of analysis investigated the areas learners sought to repair in their speech; it would simultaneously help to make CEFR descriptions a little more specific. Table 61 below displays the three main categories that emerged in the data: word choice, tense and missing words. Whilst others were identified, e.g. pronunciation slips, they were small in number and so were omitted.

Table 61: Repair across B1, B2 and C1

Level	Total can-do occurrence	Error focus							Total	Total %
		Word choice		%	Tense	%	Missing words	%		
B1	34	17	verb = 6, subject pronoun = 3; object pronoun = 2; adverb = 2; and adjective, auxiliary verb, conjunction, and preposition = 1	50.00	10	29.41	1	2.94	28	82.35
B2	52	26	subject pronoun = 10; verb = 5; auxiliary verb = 3; adverb and noun = 2; adjective, conjunction, determiner and object pronoun = 1	50.00	4	7.69	6	11.54	36	69.23
C1	63	27	verb = 8; auxiliary verb = 5; noun and subject pronoun = 4; adverb = 2; conjunction, determiner, preposition and object pronoun = 1	42.86	4	6.35	13	20.63	44	69.84

The data highlight that the three main categories comprised the majority of all repairs at B1, B2 and C1. Relating this to the CEFR, whereas B1 descriptors emphasise “mix-ups with tenses or expressions”, B2 and C1 statements respectively highlight the need to i) repair “slips and errors”, or ii) backtrack upon encountering a “difficulty” (CoE, 2001: 65). Though little expansion is offered as to slips or difficulties that B2 and C1 may constitute, the data from successful learners has shown that B1 learners do attend to tenses (29%) and errors in expression (50%), mostly related to verb choice, e.g. *he record he he make the movie* (B1, Exam 5). At B2 level, tense correction fell rather dramatically to 8% but changes to word choice remained high. The data suggests that the ‘slips’ mentioned in the CEFR may relate to subject pronouns or indeed missing words that were quickly and easily corrected (see Extracts 54 and 55). For C1 learners, however, focus shifted considerably to the issue of missing words in utterances. Whilst they did show a tendency to repair tenses (6%) or amend word choices (43%), they showed a greater capacity to rather quickly identify an error and amend their utterance (21%) (see Extract 56). What

should be pointed out, is the fact that this analysis has not aimed to assess the degree of success in accurately and appropriately fixing language errors, but instead it has aimed to demonstrate that successful learners do monitor their speech despite the processing demands placed on speakers operating in real-time conversation. Assessing accuracy would have been difficult since learners corrected inaccurate utterances both successfully and unsuccessfully, and sometimes, they corrected utterances that were correct in the first place. While this analysis aimed to delve deeper into the occurrence of repairs, it must be emphasised that reformulations and corrections are based on what learners *perceive* to be, and not what is or is not, inaccurate or ineffective.

Extract 54: B2, Exam 10

<\$20M> <\$O67> Yes we have to </\$O67> study so **my our** <\$O68> English </\$O68> is not that good.

Extract 55: B2, Exam 9

<\$18M> + and and the poor people they can't they can save the money I know my friends they don't have the erm er economic s= they don't **have financial situation good financial** <\$O54> **situation** </\$O54>

Extract 56: C1, Exam 9

<\$18M> <\$=> I heard **nine year old** </\$=> **like Dubai is nine** this is the ninth year like how old Dubai is.

6.5.10 Production strategy summary

Table 56 revealed that the majority of production strategies evidenced related to some form of repair. Though B1 learners initially appeared to use paraphrase and circumlocution more frequently, it was discovered that their most often used strategy involved correction. CEFR descriptors state that repair is very evident at B1 but that it is also often signalled by the interlocutor; this was not found to be the case. All learners corrected their speech for word choice matters. The second most common repairs then differed across the levels: B1 learners concentrated on tenses whereas at B2 and C1, focus turned to missing words. Though learners rephrased their speech, this study did not ascertain whether the changes were necessary or whether they were successful. For this reason, the appearance of repair, though a common feature of spoken grammar (see Section 3.2), may have influenced interlocutor impressions of both fluency and control during the interaction.

6.5.11 Interaction strategies

B2 learners' most frequent interaction strategy related to asking follow-up questions and seeking clarification from others. Comprising almost a third of all their interaction strategies, it warranted further analysis across the levels. Clarification encompassed 29.36% at B2, 27.35% at C1 and only 14.19% at B1. Though B1 statements also included confirmation via repetition, here only statements specifically labelled in the CoE as clarification were investigated further.

Figure 54: Can-do statements for clarification (CoE, 2001: 87)

<p>B1 <i>Can ask someone to clarify or elaborate what they have just said.</i></p>
<p>B2 <i>Can ask follow-up questions to check that he/has understood what a speaker intended to say, and get clarification of ambiguous points.</i></p>
<p>C1 <i>(As B2) Can ask follow-up questions to check that he/she has understood what a speaker intended to say, and get clarification of ambiguous points.</i></p>

The first step involved ascertaining at whom the clarification requests were aimed since the research tool included both a NS interlocutor and fellow learners. Perhaps unsurprisingly, the majority of clarifications were posed to the NS interlocutor (88%, B1; 91%, B2; and 84%, C1). It would be easy to assume that this was a sign that learners found it difficult to understand the interlocutor but this would be unfair to learners as 'understanding a native speaker' represents another ability to be evidenced (see CoE, 2001: 75). Clarifying the NS therefore represented an additional strategy permitting successful interaction in the UCLanESB data. The next stage uncovered the reasons as to why clarification had been sought. As Table 62 demonstrates, this revealed nine distinct categories.

Table 62: Reasons clarification was sought at B1, B2 and C1

Clarification purpose	B1	%	B2	%	C1	%
Task/Instructions	15	34.88	6	18.75	7	11.48
Whether a task should begin	3	6.98	2	6.25	1	1.64
Whether they should expand on their response	1	2.33	0	0.00	0	0.00
Meaning of vocabulary/question	13	30.23	4	12.50	13	21.31
Vague examiner question	3	6.98	13	40.63	10	16.39
Fellow student's question	1	2.33	0	0.00	1	1.64
Fellow student's response	5	11.63	3	9.38	7	11.48
Repetition	2	4.65	4	12.50	21	34.43
Unsure about their own statement	0	0.00	0	0.00	1	1.64
Total	43	100.00	32	100.00	61	100.00

As can be seen, two-thirds of clarification at B1 was attributed to task instruction and the meaning of vocabulary or questions. Admittedly, task instructions in Part B were not straightforward and indeed this was where most students checked what they needed to do (see Extract 57). At B2, however, focus shifted to clarifying examiner questions in which it was possible to give a range of answers. For instance, questions about learners' circumstances could be answered with information about the learners' L2 country or learners' home countries (see Extract 58):

Extract 57: B1, Exam 2

<\$3M> *Er excuse me we're same topic +*

<\$0> *Yes.*

<\$3M> *Okay but I think it's different.*

<\$0> *No your prompts are different <\$04> the topic </\$04> is the same.*

Extract 58: B2, Exam 13

<\$0> <\$26M> *er what do you like about the place where you live?*

<\$26M> *Mm here or in my country?*

<\$0> *Any.*

Interestingly, while checking the meaning of vocabulary and making interlocutor questions more explicit were both rather frequent at C1 level, the majority of learner requests related to asking for repetition. In fact, the high frequency of this reason uncovered a disguised interaction strategy. The use of repetition by learners often prompted the interlocutor to repeat the same word as if the learner had not heard it. However, it is the researcher's opinion that this was, in fact, a stalling technique which allowed learners valuable time in which to formulate their utterance before it was

articulated. Finally, though clarification of vocabulary decreased between B1 and B2, as perhaps would be expected, it increased from B2 to C1 by nearly 10%. Comparison of learner tactics revealed that whereas B1 learners explicitly requested explanation of vocabulary (see Extracts 59 and 60), C1 learners made attempts to paraphrase statements in order to confirm comprehension (see Extract 61). Success at higher levels can therefore be illustrated via attempts to recast sentences in order to clarify vocabulary, a strategy which may result from increased confidence and experience with a target language.

Extract 59: B1, Exam 3

Sorry. What's the word's meaning? (B1, Exam 3)

Extract 60: B1, Exam 14

Put a video what it mean put a video? (B1, Exam 14)

Extract 61: C1, Exam 5

<\$0> Mhm okay and how important do you think cultural awareness is when you're travelling?

<\$10F> Er so you mean cultural awareness?

<\$0> Mhm.

*<\$10F> **Mm it means er the people in the other countries tell their feelings?***

<\$0> Yes yes.

Since this thesis aims to enhance descriptions of learner success with examples of language to be evidenced, it is necessary to outline how clarification was achieved by learners. Table 63 has grouped responses according to categories that emerged in the data. B1 learners, though attempting to formulate accurate sentences had the majority of their remarks in the 'inaccurately formed question' category. This was mostly due to the omission of auxiliary verbs or inaccurate word order. B2 and C1 learners, however, showed a greater proportion of questions containing accurate structure, though this may have been due to mastery of strategic questions such as '*could you repeat that?*'. As alluded to earlier, C1 students simply used repetition phrases such as *sorry* or *excuse me* to request that the speaker repeat the utterance. Interestingly, this analysis shows some crossover with some of the lexical features highlighted in the previous research questions. The use of *you* or *so* for example can be implemented to perform a clarification function. However, a question that this analysis raises relates to how accuracy is truly judged in spoken interaction. Grammatically speaking, a total of 22 questions from B1 to C1 were grammatically incorrect. However, due to increased understanding of spoken grammar, some may argue that the questions used (see Figure 55), containing features such as

ellipsis, were in fact accurate and would pose no challenges nor impede understanding in real-time speech. The notion of success in a strategy such as clarification may not be straightforward as it would clearly be open to interpretation. Similarly, it would query the approach taken in teaching materials as to whether the structures taught would be deemed natural in interaction speech.

Table 63: Language used for clarification at B1, B2 and C1

Language used	B1	%	B2	%	C1	%	Example
Statement	5	12	0	0	0	0	<i>I don't know what's the meaning here.</i> (B1, Exam 12)
Inaccurately formed question	12	28	4	13	6	10	<i>Which one we'll talk about the one or we'll talk about all three?</i> (B2, Exam 15)
Fully formed question	3	7	6	19	12	20	<i>Er I think the plane is the er sorry can you explain the topic?</i> (C1, Exam 8)
You/It mean[s]	3	7	3	9	5	8	<i>Yeah I know er but you you mean I er if I download something.</i> (B1, Exam 14)
What is X?	3	7	0	0	3	5	<i>Maybe what's the allocate?</i> (C1, Exam 3)
Repetition of words in question	3	7	1	3	8	13	<i>Successful art?</i> (B2, Exam 5)
Sorry/Pardon/Excuse me	3	7	1	3	11	18	<i>Er excuse me?</i> (B1, Exam 9)
Why?	3	7	0	0	0	0	<i>Er why?</i> (B1, Exam 11)
So _____	3	7	3	9	4	7	<i>So you are not the type of person who travels a lot?</i> (C1, Exam 1)
Other	5	12	14	44	12	20	<i>Erm I live in in UK or China?</i> (B2, Exam 1)

Figure 55: Questions from B1, B2 and C1 data which may sound more natural

Erm er we have to begin? (B1)
Yes. Keep going on? (B1)
Both of them or I choose any one? (B1)
This means my hometown? (B2)
We should answer two questions or just choose one? (B2)
It's pets or all animals? (C1)

6.5.12 Interaction strategy summary

Clarification and follow-up questions were identified as being the most often used interaction strategy. Throughout the exams, students at different levels needed to confirm understanding or obtain more details so that they could continue with the interaction. It was found that the majority of clarification requests were aimed at the NS interlocutor. However, whilst some may question the students' ability to understand a NS, an ability identified in the CEFR, additional analysis proved this not always to be the case. B1 learners often sought clarification of task instructions, B2 learners clarified ambiguous terms which made questions rather vague and C1 learners used repetition. In the latter, the interactive strategy was believed to be of dual-purpose. Whilst it did enable learners to seek clarification or confirmation of understanding, some students did appear to use it as a stalling device. Analysis of the questions posed also observed that that B1 learners

lacked the same amount of control in word order or their use of auxiliary verbs. Conversely, at B2 and C1, learners employed stock phrases such as *could you repeat that?* Whilst this demonstrated that at these levels, learners could employ chunks of language to fulfil their needs, it simultaneously raised questions regarding the impressions of success that would be generated. Some clarification requests worked effectively and presented learners with the information they required; their use of features typical of spoken grammar whilst denoting a proficient speaker may not be considered acceptable to those judging learners according to traditional ‘writing based’ grammars.

6.5.13 RQ3 summary

Closer examination of the language realising can-do statements at the three levels was able to establish what made learners successful according to the CEFR scales. Put simply, it revealed the focus of learners’ speech and how this differed between B1, B2 and C1. The data has been able to generate an ‘overview’ or summary of the abilities demonstrated at each level:

B1: B1 learners were characteristic in their interaction. Said to be able to ‘get their message across’, three quarters of all their can-do statements related to interaction or interaction strategies. However, such interaction often denoted a focus on giving opinions and providing brief explanations (production). Often achieved via the use of *very + adjective*, as was found in Section 6.3.1, learners concentrated more on expressing their own thoughts than eliciting those of others. With regards to strategy usage, B1 learners were able to define features of unknown words, initiate, maintain or close discourse, and though not specifically stated in the CEFR statements, were able to correct their own discourse, regardless of interlocutor signals, when errors relating to word choice or tense arose.

B2: B2 learners saw interaction-related statements dominate language use once again. Though their productive abilities enabled them to systematically develop an argument in support of their statements, there were notably more attempts to invite others into the discussion, an observation previously seen in RQ2c with the frequent use of *what do you think?* Production strategies, rather than focussing on unknown vocabulary instead saw

correction increase, especially with respect to word choice. For interaction strategies, B2 learners felt the need to use questions so as to clarify another speaker's meaning.

C1: In a change from B1 and B2, this level saw production-related statements overshadow those related to interaction. The ability to expand on thoughts and ideas via the integration of sub-themes was exhibited on numerous occasions. In a similar vein, interaction did not seek the views of others, nor simply give an opinion; C1 learners saw a real need to elaborate and speculate on the problems or issues discussed, most probably in response to the 'more complex' (CEFR) topics set. To enable such elaboration, production strategies centred on the backtracking to rectify errors in particular missing words, and to facilitate discourse, C1 learners showed a frequent ability to suitably preface remarks in order to get the floor, maintain the turn and gain time.

The main differences identified across successful B1, B2 and C1 speech were therefore:

- The objective of interaction: B1 learners gave thoughts and opinions, B2 learners sought thoughts and opinions from others, and C1 learners elaborated on personal opinions.
- The combination of utterances: In terms of production, all levels displayed a typical spoken grammar trait in that ideas in utterances were joined together much like links in a chain. B1 learners were successful mostly through the addition of clause coordination using *and*, B2 learners used more subordination of clauses language with the incorporation of real-life examples to illustrate points and C1 learners often weighed up options, such as advantages and disadvantages using language like *it depends...*
- Correction: B1 learners, following further analysis, were found to correct their own language but rather less frequently than B2 and C1 learners. The proficiency rise did also not result in learners attending to more pragmatic features of language as has been observed in previous studies (see Taylor, 2011).
- The techniques used to request clarification: All learners were successful in seeking confirmation or clarification when necessary. Nearly all examples of these can-do statements related to the interlocutor's utterances rather than those of fellow learners. However, B1 learners focussed on task instructions, B2 concentrated on vague questions and C1 simply sought repetition, perhaps with a view to stalling and buying time. The appearance of clarification in the

UCLanESB corpus did not mean that learners were therefore less successful; it fulfilled other purposes than simply trying to understand a NS.

7. CONCLUSION:

7.0 Chapter introduction

In this chapter, summaries of the study's main findings will be given in accordance with the aims proposed at the beginning of the thesis. Once conclusions to the research questions have been given, the chapter will then proceed by highlighting the study's limitations, its implications for teaching and assessment, and its avenues for future research.

7.1 The aims and main findings of this study

The purpose of the current study was to determine:

- RQ1) What percentage of the words used by successful B1, B2 and C1 learners come from the first 1000 and second 1000 most frequent words in English?
- RQ2a) What were the 20 most frequent words at B1, B2 and C1 and their notable collocations, colligations and functions?
- RQ2b) What were the 20 most frequent keywords at B1, B2 and C1 and their notable collocations, colligations and functions?
- RQ2c) What were the most frequent 3- and 4-word chunks at B1, B2 and C1 and their notable collocations and functions?
- RQ3) What Common European Framework of Reference for Languages (CEFR) indicators are present in terms of spoken interaction, spoken production and strategies at B1, B2 and C1 and how are they realised?

With regards to the first question, earlier research had ascertained that a core vocabulary of the 2000 most frequent words in English would satisfy most language users' needs. However, mostly concentrating on the NS or on the skills of reading, writing and listening, less research had been conducted into speech and vocabulary profiles across proficiency levels. This study has been able to show that despite CEFR descriptions of vocabulary changes, B1, B2 and C1 learners demonstrated very little difference in their overall vocabularies, i.e. the words utilised from this 2000 most frequent group. Comparable portions of K-1 and K-2 words were produced and at all levels only one in

every 33 words originated from beyond this threshold. This finding confirmed that the 2000 most frequent words were fundamental to learner success at B1, B2 and C1 and that little difference was exhibited across the levels. Though word family coverage did show a reduction when compared with NS findings, words from the K-1 and K-2 bands still fulfilled the vast majority of UCLanESB learner needs.

Alternatively, when attention turned to the different word types used and the total number of words employed, subtle distinctions were revealed. B1, B2 and C1 CEFR descriptions of vocabulary range were found to be supported not across all word bands but within this 2000 word group. Findings established that B2 learners used significantly more K-1 and K-2 word types than B1 learners and that C1 learners likewise used more K-1 and K-2 word types than B2 and B1 learners. C1 and B2 learners also produced significantly more individual words than their B1 counterparts. The transition in the CEFR from 'enough language to get by' to a 'sufficient' then 'broad' vocabulary was thus exemplified via word types, not within a learner's *entire* vocabulary but instead within this *core* vocabulary. Related to this result, the final objectives of the first research question included testing Laufer and Nation's (1999) claim that a rise in proficiency would result in a reduced usage of high frequency words, and ascertaining from which level according to the EVP (CUP, 2015b) lexis used by learners came from. Combined token and family coverages showed that B1, B2 and C1 learners were almost identical in their vocabulary profiles but when K-1 and K-2 lexis was treated separately, K-1 usage did fall by 5% from B1 to C1; K-2 lexis conversely rose by nearly 3%. Hence, although there was a shift towards less frequent vocabulary, the difference was small and only occurred in the first 1000 words. With two-thirds of all lexis used across the levels identified as being from A1 or A2 level, this study demonstrated that successful learner speech at higher levels does not necessarily display a marked use of less frequent vocabulary nor the use of more difficult lexis, a misconception that can unfairly bias some teachers' and assessors' judgements of learners' overall success.

RQ2a, the first in a triad of research questions centring on collocation, colligation and function, aimed to uncover insights into successful learner speech via the 20 most frequent words used. Typical of many spoken corpora, all three levels' lists consisted of articles, conjunctions, pronouns, prepositions and fillers, as well as some verb forms. It was found that the 20 most frequent words comprised 45% of the B1 corpus, 41% of the B2 corpus and 39% of the C1 corpus. Corresponding with LINDSEI data, this initial analysis stressed the significance that these words had on success: without these words, a

large chunk of learner speech would be missing. The cumulative figures simultaneously identified that despite differences being small, success at higher levels resulted in a decreasing dependency on the words identified in the lists. The study lends a little support, therefore, to a lesser need for learners to cling on to lexical ‘teddy bears’ (Hasselgren, 1994) at higher proficiency levels.

Through closer examination of *we*, *er* and *erm*, RQ2a detected other subtle differences across B1, B2 and C1. *We* denoted a change in the meanings conveyed across the levels and an ability to modify meanings according to the task. Further to its pragmatic importance, it also enabled learners at B2 and C1 to create strategies clarifying and confirming understanding. Relating once again to CEFR descriptors for vocabulary range, collocations of *we* showed a tendency for B1 learners to rely on the auxiliary verb *can* whereas B2 and C1 learners were able to display a little more variety. Examination of the seemingly simplistic pronoun thus emphasised the range of functions it performed and how it enabled learners to be successful. Similarly, *er* and *erm*, expected to feature in a frequency list for speech, corroborated the importance of fillers not only for strategic needs but also for discourse. Though learners are often penalised for using such fillers, they are an essential characteristic of spontaneous, unplanned speech; reducing their usage would not only heighten the demands placed upon speakers, but also it would reduce the naturalness of their speech. The fillers’ high frequencies in the UCLanESB data instead highlighted that *er* and *erm* did not deny learners success in their speech at any level. Whilst a significant decrease in *er* at C1 provided support for the CEFR’s acknowledgement of fewer hesitations as proficiency rises, *er* and *erm* played an important role prior to lexical choices and during the addition of ideas via conjunctions.

Moving from frequent to keywords, RQ2b delved into lexis which appeared significantly more often than in the BNC World and LINDSEI. Lists based on the BNC elicited words more typical of the research tool used (i.e. the speaking exam). LINDSEI instead pinpointed the topic-specific lexis which enabled learners to discuss the selected subjects. In particular, this question’s findings highlighted the flexibility and multifunctionality afforded via the use of seemingly ordinary words; it also demonstrated how these two elements broadened as proficiency rose. *Think* clearly performed a more pragmatic role either in the realisation of learner opinions, the hedging of their comments or in the emphasising of stance but it also formed part of strategies for learners needing to buy time during the communication and for learners needing to maintain turns or complete cohesive links during the addition of ideas. *So*, on the other hand and in contrast

to Carter and McCarthy's (2006) findings, mostly acted as a conjunction for expressing the results or consequences of an action or idea. Also influential as a discourse marker for summarising speakers' and interlocutors' comments, *so* was visibly used as a strategic device fulfilling CEFR can-do statements for confirming comprehension and for signalling the end of a turn. *Like* functioned not only as a lexical verb, but also as a preposition and adverb for expressing likes and dislikes, examples, comparisons and for filling pauses in speech much like *er* and *erm* would. This latter keyword analysis importantly presented a significant change that illustrated a key difference of successful speech at B1, B2 and C1. Just as the adverbs of degree (*so*, *very*, *really*) had been shown to change in usage across levels, keyword analysis drew attention to the decreasing use of *like* as a lexical verb and its increasing use as an adverb across levels. Taken together, RQ2b contributed to learner success by revealing how learners evidenced varying and expansive usage of individual words across the levels. Taking the keywords in this study at face value, therefore, would have overlooked some of the potential, and significant, differences that can be uncovered and which lead to increasing success at higher proficiency levels. Ultimately, RQ2b showed that rather than broadening learners' knowledge of new or unknown vocabulary, significant gains can be made simply by exploiting learners' current vocabularies to achieve more with the same set of words.

The third and final question concerning collocations, colligations and functions shifted attention away from individual lexis to chunks of formulaic language. As a component of communicative competence models, RQ2c sought to identify which chunks were used by learners and the roles they performed in their speech. Many of the three- and four-word chunks did not occur in great frequencies across the UCLanESB data. Though numbers per level did rise from 469 to 538 for three-word chunks and from 181 to 189 for four-word chunks from B1 to C1, the only statistically significant gains were those found between changes in cumulative three- and four-word chunk frequencies within levels; no differences were established across levels. Whilst occurrence was often rather low, they did still form part of learner's speech. Chunks were therefore an identifiable feature of successful speech at all levels. The study did not investigate the actual effects of formulaic chunks on learner language but their appearance in the data did suggest that UCLanESB learners benefitted to some extent in terms of fluency, processing, word retrieval and accuracy.

Initial observation of the chunks used established that combinations containing *er* and *erm* reduced as proficiency increased and that the verb *think* often appeared. At B1,

the verb expressed personal opinion; it was only at B2 and C1 that it assumed a more interactive role. Other differences between learners at the three levels were identified but only once particular chunks had been examined in detail. The chunk *a lot of*, for instance, was used at all three levels but whilst collocational information seemed similar, it was the gains in accuracy that highlighted its relevance to learner success. In comparison with the quantifiers *much* and *many*, *a lot of* resulted in percentage accuracy gains at B1 and C1 of 23% and 10% respectively; B2 actually showed a slight drop. Nevertheless, the data was able to show that some of the chunks used would have had a good impression on learner accuracy in their speech. The chunk *I agree with you*, on the other hand, drew particular attention due to its frequency and evolving usage. First of all, it did not appear in the B1 lists but it did appear at B2 and C1. B2 level saw its highest usage but it did not always perform an interactive function. Its collocation with *but* revealed that in this study, the expression of agreement could also perform a strategic role for buying time, an ability detailed in B2 can-do statements. Rather than demonstrating true listenership on all occasions, therefore, B2 learners were able to use the chunk as a stalling device so they could think about or prepare their response. At C1, however, it illustrated how meaning could be co-constructed between speakers and how it was useful as a sociolinguistic device. While some learners did disagree with their partners, the use of *I agree with you* clearly maintained relationships between speakers. The chunks employed, including topic-specific lexis, allowed learners to exhibit a range of pragmatic, strategic and sociolinguistic functions whilst helping them, at times, to display a greater degree of control in their speech. Drawing parallels with the findings from RQ2a and RQ2b, successful learner speech once again relied upon the use of multifunctional chunks which can build on and exploit lexis already known to speakers.

This study's final research question aimed to address a drawback of the CEFR that had been highlighted in much previous literature. Never aiming to be language specific, the CEFR outlines the abilities of learners at different levels without actually documenting how they can be achieved or how they change from level to level. Although the EVP (CUP, 2015a) and EGP (CUP, 2015b) contribute greatly to knowledge of lexical and grammatical structures at different CEFR levels, this investigation of can-do occurrence exemplified the main characteristics that determine or lead to success in real spoken encounters at B1, B2 and C1 and the language that realises them. It established that B1 learners' can-do statements related mostly to some form of interaction in spite of their goal often being to express their own views than elicit those of others. They were

then able to briefly give reasons for these thoughts via the simple addition of ideas. It is unsurprising, therefore, that their interactive strategies concentrated on managing their own turns. In contrast to CEFR statements regarding correction, B1 learners were also found to be able to correct their own speech without prompting from the interlocutor and for production strategy statements, they most often compensated for unknown vocabulary. B2 learners, on the other hand, demonstrated what could be considered to be a more noticeable attempt at interaction. Whereas RQ2c revealed that agreement in interaction was not always genuine, B2 learners often sought to invite others into the discussion, seeking their views and clarifying them when necessary. When giving their own opinions, learners employed real-life examples, amongst other tactics, to help develop a systematic argument or highlight particular information. Slips and errors were once again the dominant productive strategy. Finally, C1 learners were identified as being distinct from learners at the previous two levels as instead of the majority of their can-do occurrence relating to interaction, they instead related to production. To be successful, therefore, C1 learners had to be able to integrate sub-themes, often using speculation, and round-off with an appropriate conclusion, though this could also have been a by-product of their more 'complex' subject matter. Again learners were often found to 'backtrack' and reformulate erroneous utterances and preface remarks to take the turn or gain time during interaction.

In closely examining the way learner language was used and of what it comprised, several conclusions for success in speech were drawn in RQ3. Firstly, the words and chunks presented in earlier RQs were clearly implicated in many of the abilities highlighted across levels. Conjunctions, fillers and chunks, for example, illustrated the simplified nature that spontaneous speech adopts and the pragmatic and strategic lexis which can aid learners in their productive 'routines' (Bygate, 1987) or in their needs to buy time during speech. It also reinforced the view that learner error, repair or reformulation is not always a sign of a lack of control but that it is a natural characteristic of spontaneous unplanned speech. Furthermore, the use of questions or repetition for clarification requests should not always be taken as a sign of a lack of understanding. It could be that learners were genuinely puzzled by the information given or the question posed, or it could be that they were less conspicuously stalling for time. RQ1, 2a, 2b and 2c were thus able to shed light on the lexis employed at B1, B2 and C1 and the functions they fulfilled in the pursuit of being successful according to the CEFR. What this final

research question has also addressed are the aims underpinning learner speech at the three levels and the manner in which lexis was called into action when learners required it.

The conclusion has so far provided a summary of each research question's key findings but this independent treatment of the RQs fails to demonstrate how their conclusions contribute collectively to the definability of success for this study. As was acknowledged in the literature review, proficiency has previously been likened to the ability to function or make use of competence in natural language situations. However, what this rather broad definition does not address is the interplay of the knowledge and skills which comprise an individual's proficiency. Though proficiency evaluations are said to be notoriously "difficult" and are often confined to analyses of vocabulary size or lexical diversity (Leclercq & Edmonds, 2014: 11), this study has demonstrated how knowledge of the first 2000 English words was utilised within and across the B1, B2 and C1 levels to realise a vast range of productive, interactive and strategic functions in learner speech. It also coincides, to some extent, to updated definitions of proficiency (see Hulstijn, 2007; 2011) as it pinpointed how lexis, common to all the learners played a fundamental role in performing differing linguistic routines and formulating differing strategies across the three levels. As Carlsen (2012: 3) remarks, variation within proficiency levels is inevitable, but it is the shared characteristics which emerge that can differentiate a particular level from those above and below it. With CEFR scales used to track what language learners *could do* at B1, B2 and C1, this study has therefore illustrated how spoken success could be distinguished via prominent productive, interactive and strategic statements *within* levels, and an increase in the multifunctionality of lexis used to realise them *across* levels.

What this study has ultimately shown is that judgements of success in second language learning and speech should therefore not be restricted to a score on a test or to broad comparisons of learners against native speaker models. Learners instead need to be viewed as language users in their own right who are able to adapt their linguistic, sociolinguistic, pragmatic and strategic knowledge in accordance with varying task demands or language use contexts. Though corpus research is able to elaborate on descriptions of what it means to be proficient in a language (Leclercq & Edmonds, 2014), success is clearly a multifaceted phenomenon requiring further extensive research.

7.2 Limitations

In all research, it is important to openly acknowledge study limitations so that the scope of findings can be fully understood. Before documenting the implications that this study has for teaching and research, its constraints must first be outlined:

1. In studies utilising corpora, especially such as the UCLanESB corpus which was solely constructed by the researcher, size will always be a limitation. While spoken, specialised corpora of learner language tend to be smaller, it must be accepted that their findings “cannot provide the basis of sweeping generalizations” about language (Carter & McCarthy, 1995: 143). Related to this is the caveat that corpora can only provide a snapshot of language. No matter how large they are, in nearly all cases, they cannot replicate the language itself, nor the infinite choices and combinations possible within it. Corpora provide a source of evidence about language but ultimately, they will never have exactly the same properties as the language itself. This means that spoken corpora, such as the UCLanESB corpus, are only able to capture what is evidenced in language use and not what the speaker is capable of. They cannot reveal what learners are able to understand and what they choose not to use in their speech. Put simply, just because particular words or structures do not appear in a corpus does not mean that learners are unaware that they exist; they may have simply decided not to employ them in their language production or the task may not have demanded them.
2. Also in terms of evidence, another limitation of this study is that it was based on learner performance. Though corpora have been said to offer insights into competence, due to the frequencies and recurrence of items found, this study of success cannot completely claim to be representative of the learners’ full communicative competence in English. Focusing specifically on an exam-based research tool, for instance, it is unable to fully report on the learners’ use of language in alternative tasks or contexts, or with different speakers. Likewise, as the previous point stated, it cannot be used to make assumptions about *all* learners’ abilities. Though it was representative of the learners at UCLan’s Preston campus, it clearly cannot describe all B1, B2 and C1 learners’ speech and nor, as some readers will point out, can it represent all first language backgrounds, nationalities,

or ages. With vocabulary range, and therefore the lexis contained with it changing from “one communication situation and register to another” (Götz, 2013: 64), the results of this study will not relate to all modes and genres of learner speech.

3. A third limitation concerns the analysis of speech through writing. In this study, spoken language was captured and represented in the form of written, broad transcriptions. Not only were multimodal features such as body language, gesture, and facial expression lost, but important prosodic information such as pronunciation, tone, timings and stress were also omitted. The transcripts allowed a permanent record of the speech to be kept but they will ultimately deny others from using the UCLanESB corpus for research into these alternate aspects. Similarly, the use of corpus analysis based on these transcripts can only highlight what is frequent in the written text files; they cannot shed light on what is absent in the data (Hunston, 2002) and they cannot be not an ideal tool for exploring infrequent, salient items that may instead hold the key to other nuances in successful speech.
4. The research tool comprising OPI speaking exams can also be seen as a limitation of this research into learner speech. As the second limitation and Section 5.2.1 addressed, exams such as OPIs are sometimes not an optimal tool in recreating speech with natural turn-taking and topic management. Similarly, using exam data would mean that the learners involved might have been concentrating on passing the test rather than satisfying the goals of the interaction. However, obtaining comparable samples of free, spontaneous speech would have posed a considerable challenge for the researcher in terms of ethics, analysis, ensuring that the spoken encounters were representative of successful speech and verifying that speakers were of B1, B2 or C1 level. Though it is felt that the best option was taken, it would be negligent not to acknowledge the research tool as a potential limitation of the study given the potential effects it may have on learners’ linguistic, sociocultural and pragmatic performances (see Norton, 2005).

7.3 Implications for teaching and learning

1. The results of this study firstly reinforce previous calls for classroom instruction to centre on the core, basic vocabulary of the first 2000 words in English (Adolphs & Schmitt, 2003; McCarthy, 1999). Though the overall size of a learner’s

vocabulary is intrinsically linked with L2 competence and assessments of proficiency (see Laufer, 1998; Meara, 1996; Stæhr, 2008; Taylor, 2011), the findings have stressed that success in speech cannot be achieved without knowledge of this specific group of words. The fact that the 2000 most frequent words accounted for 97% of successful B1, B2 and C2 speech, supports claims for learners to be taught vocabulary strategies so that they can learn words beyond this limit and compensate for unknown vocabulary when necessary (Schmitt, 2008). The use of high amounts of classroom time for words outside this vocabulary would not be entirely worthwhile given the small gains other bands yield in term of coverage (see Adolphs & Schmitt, 2003; Francis & Kucera, 1982). This has thus lead previous writers to advocate the use of wordlists for individual words and chunks. However, whilst wordlists have been likened to “gold dust for the language learner” (Harmer, cited by Longman Communication 3000, n.d.: 1), the additional insights obtained by looking into the context and usage surrounding particular words and chunks in this study have demonstrated that wordlist lexis should not be learnt in isolation. Learners should instead be given opportunities to explore words in context so that i) wordlists do not seem overwhelming, so that ii) vocabulary learning is not superficial and thus neglectful of the polysemic nature of vocabulary items, so that iii) further judgement can be offered as to the items of the most lexical content, and so that iv) they understand that not every word in the list is equally useful (see Lewis, 2012; McCarthy, 1999; Moon, 2010; Schmitt, 2008). The benefits of training learners in the use of corpora as a learning tool may therefore aid in this pursuit and have positive outcomes for their all-round success.

2. The next implication would be to consider the emphasis that is placed on the learning of vocabulary by CEFR descriptors. As pointed out on several occasions in this thesis, qualitative descriptions of vocabulary range in the CEFR clearly stress that higher amounts of vocabulary are required at higher proficiency levels. Even further into the document (see CoE, 110-112), whilst formulaic language is remarked upon, the overall impression given is that learners must learn more words if they are to progress and therefore be successful. This study by no means entirely refutes this statement. However, given that there was not much difference in vocabulary profile data and given that analysis of individual words and chunks revealed more about varying functions at B1, B2 and C1, this study argues that

broadening learners' vocabularies should not be carried out at the expense of teaching learners what can actually be achieved with the lexis they already know. Other writers have previously advocated that learners need to make use of a limited vocabulary which is continually repeated and recycled to satisfy a *range* of functions and meanings (Cobb, n.d.-a; Nation, 2001; Nation & Waring, 1997). Despite some trepidation as to the motives and definitions behind teaching vocabulary breadth or depth (see Read, 2004), it is my view that enhancing comprehensive word knowledge (ibid.) can be of more benefit to learners and have more influence on their spoken success. Undoubtedly, the first 2000 words in English provide a solid foundation that can then be built upon according to individual needs (Thornbury & Slade, 2006). By aiming to extend learners' understanding of form, meaning and use (see Nation, 2001: 27), therefore, I believe that more can be gained from enhancing *how well* they know the words they have already learnt (Nation, 2001; Qian, 2002; Schmitt, 2008).

3. Relating to vocabulary depth, Nation (2001) includes the notion of patterning under grammatical functions for use. The researcher believes that patterning in relation formulaic chunks as well as to colligation should also be implicated in this statement so as to once again extend learners' vocabulary knowledge. Though chunks did feature in learner speech, they were not always frequent in number. On the one hand, this finding supports the claim that chunks are "extremely difficult for the L2 learner to master" (Wray, 2000: 468) but on the other hand, this study may highlight that more needs to be done if learners are to use them in their speech and if they are to reap their benefits. Their inclusion in textbooks, the use of noticing and previous approaches in teaching as advocated by Erman and Warren (2000), Granger (1998), and Lewis (2012) needs to be augmented. Additionally, resulting from this study's findings, it is felt that one particular gap could be filled with learner corpora. The chunks in this study did fulfil a range of pragmatic, discourse, strategic and sociolinguistic functions but when CEFR can-do statements for production strategies and sociolinguistic functions were considered, no clear chunks could be identified. Perhaps drawing learners' attention to real language from NS or learner transcripts of speech could allow them to search for chunks of language that could be used to perform functions such as initiating, maintaining and ending turns, and seeking clarification. Though this could result in the use of 'wooden' stock phrases, such as *I take your point*

but, they could still be of benefit to learners who currently struggle during unplanned, real-time speech. If possible, the assembling of a bank of multifunctional chunks such as the ones highlighted in this study may also be useful.

4. Another implication of this study is, in fact, the UCLanESB corpus itself. Releasing it in an appropriate, user-friendly form would allow users to mine real examples of learner language according to their own needs. If this is made possible – with the UCLanESB and other learner corpora – more potential for teaching, material design and teacher training can be unlocked. Carter and McCarthy (1995), for instance, complain of the rather poor availability of spoken corpora. Though now this is not entirely the case given the release of the BNC online (BYN BNC, n.d.) and the variety of corpora available via platforms such as *Sketch Engine* (see Kilgarriff et al., 2014), other NS corpora, e.g. CANCODE, remain inaccessible. In terms of learner corpora, the English Profile Corpus can be explored using the EVP (CUP, 2015a) and EGP (CUP, 2015b) but concordances cannot be retrieved and in the case of LINDSEI, individual users instead have to pay a rather hefty licence fee. It is the researcher’s opinion, therefore, that releasing the UCLanESB corpus, particularly on a local level, may allow practitioners to gain a better understanding of learner speech. Writers have previously remarked that basing teaching on limited examples of learner data simply “goes against common sense” (Mark, 1998 cited in Granger, 2002: 6) as the value of teaching can be inhibited if practitioners fail to realise how their learners “are actually using the language” (Meyer, 2002: 27). For notions of success, the absence of real learner language or the unavailability of corpora containing it can have detrimental implications. If practitioners or indeed assessors are unaware of how learner language manifests itself and how it conforms to features of spoken grammar which may make it seem ‘simplistic’, unrealistic expectations may thus be placed on learners. As Hughes (2011: 60) observes, the gap between research and “teachers’ knowledge base” can be particularly evident where speech is concerned so it should be minimised as much as is possible.
5. Related to the use of the UCLanESB corpus for teaching, is also the potential role it could play as a study aid for learners. The introduction to this thesis stressed that the research aimed not to propose an alternative ‘model’ for language

learners. However, the language contained within the resulting corpus can itself be used to model how some functions can be realised in learner speech. As a supplementary tool, it can offer a substitute to resources containing NS speech which may seem unachievable or unrealistic. Giving learners access to the corpus or the texts it comprises could thus raise awareness of the relevance of learner language in learning aids. For instance, if the corpus was used to demonstrate how CEFR production and interaction strategies were realised in speech at different levels, it could draw learners' attention to features which make the process of speaking with fellow language users a more manageable and perhaps less daunting task. Raising awareness of features of learner language however is not sufficient on its own. To learn to be successful in speech, learners have to be given plentiful meaningful opportunities to practice speech in realistic tasks. Notably, the debate into accuracy and fluency is a complex one (see Hughes, 2011), but this study has confirmed Lewis' (2012: vi) assertion that "successful language is a wider concept than accurate language". Teachers ultimately have to decide in what ways various tasks will help learners to be successful in broader contexts outside the classroom.

7.4 Implications for research

Following on from highlighting implications for teaching and learning, this study's implications for teaching research also need to be outlined.

1. Firstly, an extension of this study is currently being carried out. Supplementing the UCLanESB corpus with additional learner data and a sub-corpus of NS data, further research will be conducted to ascertain whether additional findings can be identified. In *Successful Learner Speech: Findings from learner corpora* (Jones, Byrne & Halenko, forthcoming), analysis will be performed to see how learners' spoken English relates to communicative competence. In an extension of this current study, a corpus of apologies and requests will also be added so that specific learner pragmatic abilities can be explored.
2. Returning perhaps to a more traditional measure of success, current research is also being conducted into how sophistication and accuracy manifest in the speech of learners. Presented by Hunston (2016), a model of accuracy not based on

written grammar is being devised so that speech can be analysed from the bottom up. Once again advocating the view that learners need to be viewed and judged according to what they can do, rather than what they cannot, there is a potential for similar analysis to be conducted on the UCLanESB once more details are released.

3. Research into fluency is another way in which this current study could be extended. As previously mentioned in Section 7.2, the current composition of transcripts means that no in-depth study into fluency can be conducted. However, a preliminary test into learners' speech rates, albeit into overall exam rates rather than those specifically based on learner turns, showed that there are some potentially significant differences across learner speech at B1, B2 and C1. Fluency is of course a 'slippery notion' but it is inherently linked with impressions of learner success. As mentioned in the CEFR, for instance, at B2 learner should be able to interact "with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without imposing strain on either party" (CoE, 2001: 129) so fluency clearly has the potential to influence judgements based on learners' all round performances. Although research is being carried out at the University of Louvain into software for analysing fluency in learner language, it is felt that there is some scope here for research into learner speech rates and filled and unfilled pauses (see Gilquin & De Cock, 2013) in the UCLanESB corpus. Alternatively, if transcripts are to be made more detailed in order to research fluency, another research implication would be that the corpus could be used to investigate other spoken traits such as turn-taking and listenership.
4. In a rather considerable change to the methodology and focus of this study, it might also be relevant to explore *perceptions* of learner success in speech. Though perception-type studies sometimes do not yield any scientific, replicable facts, they can provide an excellent barometer for current beliefs or 'feeling' in the world of practice. Just as Timmis (2002) was able to provide insights into teacher and learner views towards NS models and spoken grammar in international contexts, a similar study into what makes learners successful could identify the barriers that currently exist. Research taking this approach could also investigate whether there are differences between novice and experienced teachers and

assessors, or whether learners from different cultural backgrounds judge themselves according to different criteria.

7.5 Concluding remarks

Earlier comparison of learners according to NS norms had resulted in perceptions of a lacking appreciation for what learners could successfully accomplish with their speech. Though native-like proficiency is often seen as the ultimate target in second language learning, literature acknowledges that it is insufficient in highlighting how learners can become more proficient in their speech at different stages of the second language learning process. Using descriptions of communicative competence from the CEFR as a basis, this study into B1, B2 and C1 speech has shown that learners demonstrate successful spoken language use in a number of ways. The use of a common vocabulary differing little in token and family coverages, frequency bands and difficulty was able to establish that it was the flexibility with which individual lexis could be used that most exemplified success at different levels. With particular words and chunks revealing that learner proficiency is in part reliant upon the manner that multifunctionality can be exploited and adapted in speech, this study verified that learners' production, interaction and strategies also exhibited features typical of spoken grammar. Its main implications concluded that to continue being successful, teaching should look to supply learners with lexis that can satisfy a range of functions and which enhances the value of their current vocabularies rather than simply extending them. It also accepted that attitudes to learner success are changing in language teaching, but that the use of learner corpora and learner language examples in the classroom can further enrich the abilities of learners by providing them with models which are more realistic and ultimately more attainable.

8. REFERENCES

- Adolphs, S., 2006. *Introducing electronic text analysis: A practical guide for language and literary studies*. Abingdon: Routledge.
- Adolphs, S. and Carter, R., 2013. *Spoken corpus linguistics: From monomodal to multimodal*. Abingdon: Taylor & Francis.
- Adolphs, S. and Knight, D., 2010. Building a spoken corpus: What are the basics? In: A. O’Keeffe & M. McCarthy, eds. 2010. *The Routledge handbook of corpus linguistics*. Abingdon: Routledge. pp.38-52.
- Adolphs, S. and Schmitt, N., 2003. Lexical coverage of spoken discourse. *Applied Linguistics*, 24(4), pp.425-438.
- Alderson, J.C., 2007. The CEFR and the need for more research. *The Modern Language Journal*, 91(4), pp.659-663.
- Alderson, J.C., Figueras, N., Kuijper, H., Nold, G., Takala, S. and Tardieu, C., 2004. *The development of specifications for item development and classification within the Common European Framework of Reference for Languages: learning, teaching, assessment. Reading and listening. Final report of the Dutch CEF construct project*. Unpublished document.
- Alptekin, C., 2002. Towards intercultural communicative competence in ELT. *ELT Journal*, 56(1), pp.57-64.
- Andreou, G. and Galantomos, I., 2009. The native speaker ideal in foreign language teaching. *Electronic Journal of Foreign Language Teaching*, 6(2), pp.200-208.
- Anthony, L., 2014. AntConc. (Version 3.4.3). [computer software] Tokyo, Japan: Waseda University. Available at: <<http://www.laurenceanthony.net/>> [Accessed 20 August 2014].
- Aston, G., 1997. *Small and large corpora in language teaching*. [online] Available at: <<http://www.sslmit.unibo.it/~guy/wudj1.htm>> [Accessed 25 May 2016].
- Aston, G., 2008. *It’s only human*. In: A. Martelli, V. Pulcini, eds. 2008. *Investigating English with corpora: Studies in honour of Maria Teresa Plat*. Monza: Italy. pp.343-354.
- Bachman, L.F., 1990. *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L.F., 2004. *Statistical analysis for language assessment*. Cambridge: Cambridge University Press.
- Bachman, L.F., 2007. What is the construct? The dialect of abilities and contexts in defining constructs in language assessment. In: J. Fox, M. Wesche, D. Bayliss, L. Cheng, C.E. Turner, & C. Doe, eds. 2007. *Language testing reconsidered*. Ontario: Ottawa Press. pp.41-72.

- Bachman, L.F. & Palmer, A.S., 1996. *Language testing in practice*. Oxford: Oxford University Press.
- Baker, P., 2006. *Using corpora in discourse analysis*. London: Continuum.
- Baker, P., 2010. Corpus methods in linguistics. In: L. Litosseliti, ed. 2010. *Research Methods in Linguistics*. London: Continuum International Publishing Group. pp.93-117.
- Baker, W., 2015. *Culture and identity through English as a lingua franca: Rethinking concepts and goals in intercultural communication*. Berlin: Walter de Gruyter.
- Barnbrook, G., 1996. *Language and computers: A practical introduction to the computer analysis of language*. Edinburgh: Edinburgh University Press.
- Barker, F., 2010. How can corpora be used in language testing? In: A. O’Keeffe and M. McCarthy, eds. 2010. *The Routledge handbook of corpus linguistics*. New York: Routledge. pp.633-645.
- Becker, J.D., 1975. The phrasal lexicon. Proceedings of the 1975 workshop on Theoretical issues in natural language processing. Association for Computational Linguistics, 1975.
- Biber, D., Conrad, S. and Reppen, R., 1998. *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E., 1999. *Longman grammar of spoken and written English*. Harlow: Longman.
- Birdsong, D., 2004. Second language acquisition and ultimate attainment. In: A. Davies and C. Elder, eds. 2004. *The Handbook of Applied Linguistics*. Oxford: Blackwell Publishing. pp.82-105.
- Bley-Vroman, R., 1983. The comparative fallacy in interlanguage studies: The case of systematicity. *Language Learning*, 33(1), pp.1-17.
- Bloomfield, L., 1933. *Language*. New York: Holt, Rinehart & Winston.
- Boers, F., Eyckmans, J., Kappel, J., Stengers, H. and Demecheleer, M. 2006. Formulaic sequences and perceived oral proficiency: Putting a Lexical Approach to the test. *Language Teaching Research*, 10(3), pp.245-261.
- Bondi, M., 2010. Perspectives on keywords and keyness: An introduction. In: M. Bondi and M. Scott, eds. 2010. *Keyness in texts*. Amsterdam: John Benjamins Publishing. pp.1-20.
- Bonnet, G., 2007. The CEFR and Education Policies in Europe. *The Modern Language Journal*. 91(4), pp.669-672.
- Brezina, V., 2014. *Statistics in Corpus linguistics: Web resource*. [online] Available at: <<http://corpora.lancs.ac.uk/stats>> [Accessed 2 June 2016].

Brigham Young University British National Corpus (BYU-BNC), n.d. *British National Corpus: BYU-BNC*. Available at: <<http://corpus.byu.edu/bnc/>> [Accessed 17 June 2016].

British Association for Applied Linguistics (BAAL), 2006. *The British Association for Applied Linguistics: Recommendations on Good Practice in Applied Linguistics*. [online] Available at: <http://www.baal.org.uk/dox/goodpractice_full.pdf> [Accessed 18 May 2015].

British National Corpus (BNC), 2004. *British National Corpus: What is the BNC?* [online] Available at: <<http://www.natcorp.ox.ac.uk/corpus/index.xml>> [Accessed 21 May 2014].

Brown, D.H., 2004. *Language assessment: Principles and classroom practices*. White Plains: Pearson Education.

Brumfit, C., 2001. *Individual freedom in language teaching: Helping learners to develop a dialect of their own*. Oxford: Oxford University Press.

Burnard, L., 2000. *Reference guide for the British National Corpus*. Available at: <<http://www.natcorp.ox.ac.uk/archive/worldURG/urg.pdf>> [Accessed 17 June 2016].

Bygate, M., 1987. *Speaking*. Oxford: Oxford University Press.

Callies, M. and Götz, S., 2015. *Learner corpora in language testing and assessment*. Amsterdam: John Benjamins Publishing.

Cambridge ESOL, 2011. *Using the CEFR: Principles of good practice*. Cambridge: Cambridge ESOL.

Cambridge University Press [CUP], 2015a. *English Profile: The CEFR for English: English Vocabulary Profile*. [online] Available at: <<http://www.englishprofile.org/wordlists>> [Accessed 7 June 2016].

Cambridge University Press (CUP), 2015a. *English Vocabulary Profile*. [online] Available at: <<http://www.englishprofile.org/wordlists>> [Accessed 21 January 2016].

Cambridge University Press (CUP), 2015b. *English Grammar Profile*. [online] Available at: <<http://www.englishprofile.org/english-grammar-profile>> [Accessed 21 January 2016].

Cambridge University Press (CUP), 2015c. *Compiling the EVP*. [online] Available at: <<http://www.englishprofile.org/wordlists/compiling-the-evp>> [Accessed 21 January 2016].

Canagarajah, S., 2007. Lingua franca English, multilingual communities, and language acquisition. *The Modern Language Journal*, 91(Supplement 1), pp.923-939.

Canale, M., 1983. From communicative competence to communicative language pedagogy. In: J.C. Richards and R.W. Schmidt, eds. 1983. *Language and communication*. New York: Longman. pp.2-27.

- Canale, M. and Swain, M., 1980. Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), pp.1-47.
- Carlsen, C., 2012. Proficiency level – A fuzzy variable in computer learner corpora. *Applied Linguistics*, 33(2), pp.161-183.
- Carter, R., 1998. Orders of reality: CANCODE, communication and culture. *ELT Journal*, 52(1), pp.43-56.
- Carter, R. and McCarthy, M., 2006. *Cambridge grammar of English: A comprehensive guide: Spoken and written English grammar and usage*. Cambridge: Cambridge University Press.
- Celce-Murcia, M., 2008. Rethinking the role of communicative competence in language teaching. In: E. Alcón Soler and M.P. Safont Jordá, eds. 2008. *Intercultural language use and language learning*. s.l.: Springer, pp.41-57.
- Celce-Murcia, M. & Dornyei, Z., 1995. Communicative competence: A pedagogically motivated model with content specifications. *Applied Linguistics*, 6(2), pp.5-35.
- Chalhoub-Deville, M. and Fulcher, G., 2003. The oral proficiency interview: A research agenda. *Foreign Language Annals*, 36(4), pp.498-506.
- Chomsky, N., 1965. *Aspects of the theory of syntax*. Cambridge: MIT Press.
- Chujo, K., 2004. Measuring vocabulary levels of English textbooks and tests using a BNC lemmatised high frequency word list. *Language and Computers*, 51(1), pp.231-249.
- Clark, H.H. and Fox Tree, J.E., 2002. Using uh and um in spontaneous speaking. *Cognition* 84, pp.73–111.
- Cobb, T., n.d.-a. *Why and how to use frequency lists to learn words*. [online] Available at: <<http://www.lex tutor.ca/research/>> [Accessed 2 March 2016].
- Cobb, T., n.d.-b. *Compleat Lexical Tutor v.8*. [online] Available at: <<http://www.lex tutor.ca/>> [Accessed 2 March 2016].
- Cobb, T., n.d.-c. *VocabProfile home*. [online] Available at: <<http://www.lex tutor.ca/vp/>> [Accessed 5 March 2016].
- Cohen, L., Manion, L. and Morrison, K., 2011. *Research Methods in Education*, 7th ed. Abingdon: Routledge.
- Cook, V.J., 1992. Evidence for multicompetence. *Language Learning*, 42(4), pp.557-591.
- Cook, V., 1999. Going beyond the native speaker in language teaching. *TESOL Quarterly*. 33(2), pp.185-209.
- Cook, V., 2002. *Portraits of the L2 user*. Clevedon: Multilingual Matters.

- Cook, V., 2008. *Second language learning and language teaching*, 4th ed. London: Arnold.
- Coperías Aguilar, M.J., 2008. Dealing with intercultural communicative competence in the Foreign Language Classroom. In: E. Alcón Soler and M.P. Safont Jordá, eds. 2008. *Intercultural language use and language learning*, s.l.: Springer, 59-78.
- Coppieters, R., 1987. Competence differences between native and near-native speakers. *Language*. pp.544-573.
- Corder, S.P., 1967. The significance of learners' errors. *International Review of Applied Linguistics*. 5(1-4), pp.161-170.
- Coste, D., 2007. Contextualising uses of the common European framework of reference for languages. In *Report of the intergovernmental Forum "The Common European Framework of Reference for Languages (CEFR) and the development of language policies: challenges and responsibilities"*. pp. 38-47.
- Council of Europe (CoE), 2001. *Common European Framework of Reference for Languages: Language, teaching, assessment*. Cambridge: Cambridge University Press.
- Council of Europe (CoE), 2014. *Education and languages, language policy: The Common European Framework of Reference for Languages: Learning, teaching and assessment (CEFR)* [online] Available at: <http://www.coe.int/t/dg4/linguistic/cadre1_en.asp> [Accessed 12 July 2015].
- Council of Europe (CoE), 2015. *Council of Europe in brief*. [online] Available at: <<http://www.coe.int/en/web/about-us/who-we-are>> [Accessed 12 July 2015].
- Council of Europe Committee of Ministers, 1982. *Recommendation no. R(82)18 of the committee of ministers to member states concerning modern languages*. [online] Available at: <<https://wcd.coe.int/com.instranet.InstraServlet?command=com.instranet.CmdBlobGet&Instr=anetImage=601630&SecMode=1&DocId=676400&Usage=2>> [Accessed 12 August 2015].
- Crowdy, S., 1994. Spoken corpus transcription. *Literary and Linguistic Computing*. 9(1), pp.25-28.
- Cullen, R. and Kuo, I.C., 2007. Spoken grammar and ELT course materials: A missing link? *TESOL Quarterly*, 41(2), pp.361-386.
- Data Protection Act, 1998. (c.29). London: Stationery Office.
- Davidson, F. and Fulcher, G., 2007. The Common European Framework of Reference (CEFR) and the design of language tests: A matter of effect. *Language Teaching*. 40(3), pp.231-241.
- De Bot, K., 1992. A bilingual production model: Levelt's 'speaking' model adapted. *Applied Linguistics*, 13(1), pp.1-24.

- De Cock, S., 1998. A Recurrent Word Combination Approach to the Study of Formulae in the Speech of Native and Non-Native Speakers of English. *International Journal of Corpus Linguistics*. 3(1), pp.59-80.
- De Cock, S., 2010. Spoken learner corpora and EFL teaching. In: M.C. Campoy-Dubillo, B. Belles Foruño and M.L. Gea-Valor, eds. 2010. *Corpus-based approaches to English language teaching*. London: Continuum. pp.123-138.
- De Jong, J., 2004. *What is the role of the Common European Framework of Reference for Languages: learning, teaching, assessment?* Invited lecture at the first Annual Meeting of the European Association for Language Testing and Assessment (EALTA), Kranjska Gora, Slovenia.
- Diniz, L., 2005. Comparative review: Textstat 2.5, Antconc 3.0, and Compleat Lexical Tutor 4.0. *Language Learning and Technology*. 9(3), pp.22-27.
- Dornyei, Z., 2007. *Research methods in applied linguistics*. Oxford: Oxford University Press.
- Dornyei, Z. & Skehan, P., 2003. Age. In: C.J. Doughty and M.H. Long, eds. 2008. *The handbook of second language acquisition*. Oxford: Blackwell.
- Dressler, R.A. and Kreuz, R., 2000. Transcribing oral discourse: A survey and a modal system. *Discourse processes*, 29(1), pp.23-36.
- Edge, J., 1988. Natives, speakers, and models. *JALT Journal*, 9(2), pp.153-157.
- Edwards, J.A. 1992. Design principles in the transcription of spoken discourse. In: J. Svartvik ed. 1992. *Trends in linguistics: Studies and monographs 65. Directions in corpus linguistics: Proceedings of Nobel Symposium 82 Stockholm, 4-8 August 1991*. Berlin: Mouton de Gruyter. pp.129-144.
- Ellis, R., 1994. *The study of second language acquisition*. Oxford: Oxford University Press.
- Ellis, N., Simpson-Vlach, R. and Maynard, C., 2008 Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics and TESOL. *TESOL Quarterly*, 42(3), pp.375-396.
- English Profile, 2016. *10th anniversary English Profile seminar*. Cambridge, Cambridge University Press. 5th February 2016.
- Field, J., 2011. Cognitive validity. In: L. Taylor ed. 2011. *Examining speaking: Research and practice in assessing second language speaking*. Cambridge: Cambridge University Press. pp.65-111.
- Figueras, N., 2007. The CEFR, a lever for the improvement of language professionals in Europe. *The Modern Language Journal*. 91(4), pp.673-676.
- Figueras, N., 2012. The impact of the CEFT. *ELT Journal*, 66(4), pp.477-485.

- Figueras, N., North, B., Takala, S., Verhelst, N. and Avermaet, P.V., 2005. Relating examinations to the Common European Framework: A manual. *Language Testing*. 22(3), pp.261-279.
- Firth, A., 1996. The discursive accomplishment of normality: On 'lingua franca' English and conversation analysis. *Journal of pragmatics*. 26, pp.237-259.
- Firth, A. and Wagner, J., 1997. On discourse, communication, and (some) fundamental concepts in SLA research. *The modern language journal*, 81(3), pp.285-300.
- Foot, M.C., 1999. Relaxing in pairs. *ELT Journal*, 53(1), pp.36-41.
- Fulcher, G., 2003. *Testing second language speaking*. Harlow: Pearson Education Limited.
- Fulcher, G., 2004. Deluded by artifices: The Common European Framework and Harmonization, *Language Assessment Quarterly*. 1(4), pp.253-266.
- Galambos, S.J. and Goldin-Meadow, S., 1990. The effects of learning two languages on levels of metalinguistic awareness. *Cognition*, 34(1), pp.1-56.
- Galaczi, E. & French, A., 2011. Context validity. In: L. Taylor ed. 2011. *Examining speaking: Research and practice in assessing second language speaking*. Cambridge: Cambridge University Press. pp.112-170.
- Gass, S.M. & Selinker, L., 2001. *Second language acquisition*. 2nd ed. Mahwah: Lawrence Erlbaum Associates.
- Gilquin, G. & De Cock, S., 2013. *Errors in disfluencies in spoken corpora*. Amsterdam: John Benjamins Publishing.
- Gilquin, G., De Cock, S. and Granger, S., 2010. LINDSEI: Louvain International Database of Spoken English Interlanguage. [CD-ROM]. Louvain: Presses Universitaires de Louvain.
- Götz, S., 2013. *Fluency in native and non-native English speech*. Amsterdam: John Benjamins Publishing.
- Goulden, R., Nation, P. and Read, J., 1990. How Large Can a Receptive Vocabulary Be? *Applied linguistics*. 11(4), pp.341-363.
- Goullier, F., 2006. *Council of Europe tools for language teaching: Common European Framework and portfolios*. [online] Available at: <http://www.coe.int/t/dg4/linguistic/Source/Goullier_Outils_EN.pdf> [Accessed 12 August, 2015].
- Graddol, D., 2006. *English next: Why global English may mean the end of 'English as a foreign language'*. [online] Available at: <<http://englishagenda.britishcouncil.org/sites/ec/files/books-english-next.pdf>> [Accessed 6 February 2014].

- Granger, S., 2002. The role of computer learner corpora in SLA research and FLT. In: S. Granger, J. Hung and S. Petch-Tyson, eds. 2002. *Computer learner corpora: Second language teaching and foreign language teaching*. Philadelphia: John Benjamins Publishing.
- Greaves, C. and Warren, M., 2010. What can a corpus tell us about multi-word units? In: A. O’Keeffe & M. McCarthy, eds. 2010. *The Routledge handbook of corpus linguistics*. Abingdon: Routledge. pp.212-226.
- Gries, S.T., 2010. Methodological skills in corpus linguistics: A polemic and some pointers towards quantitative methods. In: Harris, T. and J.M. Moreno, eds. 2010. *Corpus linguistics in language teaching*. Berne: Peter Lang AG, pp.121-146
- Han, Z.H., 2004. Fossilization: Five central issues. *International Journal of Applied Linguistics*, 14(2), pp.212- 242.
- Hasselgren, A., 1994. Lexical teddy bears and advanced learners: a study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics*. 4(2), pp.237-260.
- Hilgismann, P., n.d. *The International Corpus of Learner English v.2*. Available at: <<https://www.uclouvain.be/en-277586.html>> [Accessed 11 June 2016]
- Hoey, M., 2005. *Lexical priming: a new theory of words and language*. London: Routledge.
- House, J., 2003. English as a lingua franca: A threat to multilingualism? *Journal of Sociolinguistics*, 7(4), pp.556-578.
- Hughes, R., 2011. *Teaching and researching speaking*. 2nd ed. London: Pearson Education.
- Hulstijn, J., 2007. The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency. *Modern Language Journal*, 91(4), pp.663-667.
- Hulstijn, J., 2011. Language proficiency in native and nonnative speakers: An agenda for research and suggestions for second-language assessment. *Language Assessment Quarterly*, 8, pp.229-249.
- Hunston, S., 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Hunston, S., 2010. How can a corpus be used to explore patterns? In: A. O’Keeffe and M. McCarthy, eds. 2010. *The Routledge handbook of corpus linguistics*. Abingdon: Routledge, pp.152-166.
- Hunston, S., 2016. Measuring improvements in learner language: Issues of complexity and correctness. In: English Profile, 10th anniversary *English Profile seminar*. Cambridge, Cambridge University Press. 5th February 2016.

Hymes, D., 1992. The concept of communicative competence revisited. In: M. Putz, ed. 1992. *Thirty years of linguistic evolution. Studies in honour of René Dirven on the occasion of his sixtieth birthday*. Amsterdam: John Benjamins Publishing. pp.31-57.

IBM Corporation, 2015. *SPSS Statistics (23)*. [computer programme] IBM Corporation.

Jenkins, J., 2006. Current perspectives on teaching world Englishes and English as a lingua franca. *TESOL Quarterly*, 40(1), pp.157-181.

Johnson, M., and Tyler, A., 1998. Re-analyzing the OPI: How much does it look like natural conversation. In: R. Young and A.W. He, eds. 1998. *Talking and testing: Discourse approaches to the assessment of oral proficiency*. Amsterdam: John Benjamins Publishing. pp.27-51.

Jones, C., Byrne, S. and Halenko, N., (in press). *Successful spoken English: Findings from learner corpora*. Abingdon: Routledge.

Jones, C. and Waller, D., 2015. *Corpus linguistics for grammar: A guide for research*. Abingdon: Routledge.

Jones, C., Waller, D. and Golebiewska, P., 2013. Defining successful spoken language at B2 level: Findings from a corpus of learner test data. *The European Journal of Applied Linguistics and TEFL*, 2(2), pp.29-45.

Jones, N., 2002. *Relating the ALTE framework to the Common European Framework of Reference*. In: J.C.A. Alderson, ed. 2002. *Common European Framework of Reference for Languages: Learning, teaching, assessment: Case studies*. Strasbourg: Council of Europe. pp. 167-183.

Jones, N. and Saville, N., 2009. European language policy: Assessment, learning and the CEFR. *Annual Review of Applied Linguistics*, 29, pp.51-63.

Kachru, B.B., 1992a. *The other tongue: English across cultures*. Illinois: University of Illinois Press.

Kachru, B.B., 1992b. World Englishes: Approaches, issues and resources. *Language Teaching*, 25(1), pp.1-14.

Kennedy, G., 1998. *An introduction to corpus linguistics*. Harlow: Addison Wesley Longman Limited.

Kilgarriff, A., 2005. Language is never ever ever random. *Corpus Linguistics and Linguistic Theory*. 1-2, pp.263-275.

Kilgarriff, A., Baisa, V., Busta, J., Jakubicek, M., Kovar, V., Michelfeit, J., Rychly, P., and Suchomel, V., 2014. The Sketch Engine: Ten years on. *Lexicography*, 1(1), pp.7-36. Available at: <<https://www.sketchengine.co.uk/>> [Accessed 12 June 2016].

Koester, A., 2010. Building small, specialised corpora. In: A. O'Keeffe and M. McCarthy, eds. 2010. *The Routledge handbook of corpus linguistics*. Abingdon: Routledge, pp.66-79.

Kramsch, C., 2003. *The privilege of the non-native speaker. The sociolinguistics of foreign-language classrooms: Contributions of the native, the near-native, and the non-native speaker*. Boston: Heinle, pp.251-62.

Language Policy Division (LPD), 2006. *Plurilingual Education in Europe: 50 years of international cooperation*. [online] Available at: <http://www.coe.int/t/dg4/linguistic/Source/PlurinlingalEducation_En.pdf> [Accessed 12 August 2015].

Language Policy Division (LPD), 2009. *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR): A manual*. [online] Available at: <http://www.coe.int/t/DG4/linguistic/Source/ManualRevision-proofread-FINAL_en.pdf> [Accessed 12 August 2015].

Laufer, B., 1998. The Development of Passive and Active Vocabulary in a Second Language: Same or Different? *Applied Linguistics*. 19(2), pp.255-271.

Laufer, B. and Nation, P., 1995. Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*. 16(3), pp.307-322.

Laufer, B. and Nation, P., 1999. A vocabulary-size test of controlled productive ability. *Language Testing*. 19(33), pp.33-51.

Leclercq, P. & Edmonds, A., 2014. How to assess L2 proficiency? An overview of proficiency assessment research. In: P. Leclercq, A. Edmonds & H. Hilton, eds. 2014. *Measuring L2 proficiency: Perspectives from SLA*. Bristol: Multilingual Matters. pp.3-26.

Lee, J.J., 2005. The native speaker: An achievable model? *Asian EFL Journal*, 7(2), pp.152-163.

Leech, G.N., 1992. 100 Million Words of English: The British National Corpus (BNC). Available at: <<http://s-space.snu.ac.kr/bitstream/10371/85959/1/1.%202235197.pdf>> [Accessed 17 June 2016].

Leech, G., 2000. Grammars of spoken English: New outcomes of corpus-oriented research. *Language Learning*, 50(4), pp.675-724.

Leech, G., Rayson, P. and Wilson, A. 2001. *Companion website for Word Frequencies in Written and Spoken English: based on the British National Corpus*. Available at: <<http://ucrel.lancs.ac.uk/bncfreq/>> [Accessed 20 September 2015]

Levelt, W. J. M., 1989. *Speaking: From intention to articulation*, Cambridge, MA: MIT Press.

Lewis, M., 2012. *The lexical approach: The state of ELT and the way forward*. Hove: Language Teaching.

Litoselliti, L., 2010. *Research methods in linguistics*. London: Continuum.

Little, D., 2007. The Common European Framework of Reference for Languages: Perspectives on the making of supranational language education policy. *Modern Language Journal*, 91(4), pp.645-653.

Longman Communication 3000, n.d. *Longman Communication 3000*. Available at: <http://www.lex tutor.ca/freq/lists_download/longman_3000_list.pdf> [Accessed 13 October 2015].

Lyons, J., 1996. On competence and performance and related notions. In: G. Brown, K. Malmkjær and J. Williams, eds. 1996. *Performance & competence in second language acquisition*. Cambridge: Cambridge University Press. pp.9-32.

Mackey, A. & Gass, S.M., 2005. *Second language research: methodology and design*. Mahwah: Lawrence Erlbaum.

Martyniuk, W. and Noijons, J., 2007. Executive summary of results of a survey on: The use of the CEFR at national level in the Council of Europe Member States. Strasbourg: Council of Europe.

McArthur, T., 1992. Models of English. *English Today*, 8(4), pp.12-21.

McCarthy, M., 1990. *Vocabulary*. Oxford: Oxford University Press.

McCarthy, M., 1998. *Spoken language and applied linguistics*. Cambridge: Cambridge University Press.

McCarthy, M., 1999. What constitutes a basic vocabulary for spoken communication? *Studies in English Language and Literature*, 1, pp.233-249.

McCarthy, M., 2006. *Explorations in corpus linguistics*. Cambridge: Cambridge University Press. Available at: <<http://www.cambridge.org/elt/teacher-support/pdf/McCarthy-Corpus-Linguistics.pdf>> [Accessed 14 June 2016].

McCarthy, M., 2010. Spoken fluency revisited. *English Profile Journal*, 1(1), pp.1-15.

McCarthy, M., 2016. Discussion and roundup. In: English Profile, 10th anniversary *English Profile seminar*. Cambridge, Cambridge University Press. 5th February 2016.

McCarthy, M. and Carter, R., 2001. *Ten criteria for a spoken grammar*. Available at: <<http://www.cambridge.org/elt/touchstone/images/pdf/Ten%20criteria%20for%20a%20spoken%20grammar.pdf>> [Accessed 16 June 2016].

McEnery, T. and Hardie, A., 2012. *Corpus linguistics*. Cambridge: Cambridge University Press.

McEnery, T. and Wilson, W., 2001. *Corpus linguistics*, 2nd ed. Edinburgh: Edinburgh University Press.

McEnery, T. and Xiao, R., 2010. What corpora can offer in language teaching and learning. In: E. Hinkel, ed. 2010. *Handbook of research in second language teaching and learning*, vol.2. London: Routledge. pp.364-380.

- McEnery, T., Xiao, R. and Tono, Y., 2006. *Corpus based language studies: An advanced resource book*. Abingdon: Routledge.
- McNamara, T.F., 1995. Modelling performance: Opening Pandora's box. *Applied Linguistics*, 16(2), pp.159-179.
- McNamara, T.F., 2000. *Language testing*. Oxford: Oxford University Press.
- McNamara, T.F., 2004. Language testing. In: A. Davies and C Elder, eds. 2004. *The handbook of applied linguistics*. Malden: Blackwell.
- McNamara, T.F., 2006. *Language testing: The social dimension*. Oxford: Blackwell.
- Meara, P., 1996. The dimensions of lexical competence. In: G. Brown, K. Malmkjær and J. Williams, eds. 1996. *Performance & competence in second language acquisition*. Cambridge: Cambridge University Press. pp.70-88.
- Medgyes, P., 1992. Native or non-native: Who's worth more? *ELT Journal*, 46(4), pp.340-349.
- Meyer, C., 2002. *English corpus linguistics: An introduction*. Cambridge: Cambridge University Press.
- Milambiling, J., 2000. Comments on Vivian Cook's going beyond the native speaker in language teaching: How nonnative speakers as teachers fit into the equation. *TESOL Quarterly*, 34(2), pp.324-328.
- Miles, M.B., Huberman, A.M. and Saldaña, J. 2014. *Qualitative data analysis: A methods sourcebook*. London: Sage.
- Modiano, M., 1999. International English in the global village. *English Today*, 15(2), pp.22-28.
- Moon, R., 2010. What can a lexis tell us about lexis? In: A. O'Keeffe and M. McCarthy, eds. 2010. *The Routledge handbook of corpus linguistics*. Abingdon: Routledge, pp.197-211.
- Naiman, N., Frolich, M., Stern, H.H. and Todesco, A., 1978. *The good language learner*. Clevedon: Multilingual Matters.
- Nation, I.S.P., 2001. *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I.S.P., 2001. *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, P. and Chung, T., 2009. Teaching and testing vocabulary. In: M.H. Long and C.J. Doughty, eds. 2009. *The handbook of language teaching*, Oxford: John Wiley & Sons, pp.543-559

- Nation, P. and Waring, R., 1997. Vocabulary size, text coverage and word lists. *Vocabulary: Description, acquisition and pedagogy*. 14, pp.6-19.
- Nattinger, J.R. and DeCarrico, J.S., 1992. *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- North, B., 2007. The CEFR illustrative descriptor scales. *Modern Language Journal*, 91(4), pp.656-659.
- O’Keeffe, A. and Mark, G. (forthcoming). The English Grammar Profile – describing a criteria-based empirical methodology for calibrating learner grammar competencies using the Cambridge Learner Corpus. *International Journal of Corpus Linguistics*
- O’Keeffe, A., McCarthy, M. and Carter, R., 2007. *From corpus to classroom: Language use and language teaching*. Cambridge: Cambridge University Press.
- O’Malley, J. and Chamot, A.U., 1990. *Learning strategies in second language acquisition*. Cambridge: Cambridge University Press.
- Oakes, M.P., 1998. *Statistics for corpus linguistics*. Edinburgh: Edinburgh University Press.
- Oxford University Press, (OUP). 2016. *Oxford dictionaries: Language matters*. [online] Available at <<http://www.oxforddictionaries.com/definition/english/success>> [Accessed 14 March 2016].
- Qian, D.D., 2002. Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*. 52(3), pp. 513–536.
- Paulston, C.B., 1992. *Linguistic and communicative competence*. Avon: Multilingual Matters Ltd.
- Pawley, A. and Syder, F.H., 1983. Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In: J.C. Richards and R.W. Schmidt, eds. 1983. *Language and communication*. New York: Longman. pp. 191-227.
- Phillipson, R., 1996. Linguistic imperialism: African perspectives. *ELT Journal*, 50(2), pp.160-167.
- Piller, I., 2002. Passing for a native speaker: Identity and success in second language learning. *Journal of Sociolinguistics*, 6(2), pp.179-208.
- Preston, D.R., 1981. The Ethnography of TESOL. *TESOL Quarterly*, 15(2), pp.105-116.
- Prodromou, L., 1997. Correspondence. *ELT Journal*, 63(4), pp.439-440.
- Prodromou, L., 2008. *English as a lingua franca: A corpus-based analysis*. London: Continuum.

- QSR, 2012. NVivo qualitative data analysis (software version 10). [computer software] QSR International Pty Ltd.
- Rampton, M.B.H., 1990. Displacing the ‘native speaker’: Expertise, affiliation, and inheritance. *ELT Journal*, 44(2), pp.97-101.
- Riggenbach, H., 1998. Evaluating learner interactional skills. In: R. Young and A.W. He, eds. 1998. *Talking and testing: Discourse approaches to the assessment of oral proficiency*. Amsterdam: John Benjamins Publishing. pp.53-68.
- Rock, F., 2001. Policy and Practice in the Anonymisation of Linguistic Data. *International Journal of Corpus Linguistics*. 6(1), pp.1-26.
- Salkind, N.J., 2014. *Statistics for people who (think they) hate statistics*. 5th ed. Thousand Oaks: Sage Publications.
- Saniei, A., 2011. Who is an ideal native speaker? *International Proceedings of Economics Development & Research*, 26, pp.74-78.
- Saville, N. and Hargreaves, P., 1999. Assessing speaking in the FCE. *ELT Journal*, 52(1), pp.42-51.
- Saville-Troike, M., 2012. *Introducing second language acquisition*. Cambridge: Cambridge University Press.
- Schachter, J., 1996. Learning and triggering in adult L2 acquisition. In: G. Brown, K. Malmkjær and J. Williams, eds. 1996. *Performance & competence in second language acquisition*. Cambridge: Cambridge University Press. pp.70-88.
- Schmitt, N., 2000. Key concepts in ELT. *ELT Journal*, 54(4), pp.400-401.
- Schmitt, N., 2008. Instructed second language vocabulary learning. *Language Teaching research*, 12(3), pp.329-363.
- Schmitt, N., 2010. *Researching vocabulary: A vocabulary research manual*. Basingstoke: Palgrave Macmillan.
- Scott, M., 2015. *WordSmith Tools.v.6.0.0.252*. [computer programme manual]. Lexical Analysis Software Ltd. Available at: <<http://lexically.net/wordsmith/>> [Accessed 29 December 2015].
- Scott, M. and Tribble, C., 2006. *Textual patterns: Keywords and corpus analysis in language education*. Amsterdam: John Benjamins Publishing.
- Seidlhofer, B., 2001. Closing a conceptual gap: The case for a description of English as a lingua franca. *International Journal of Applied Linguistics*, (11)2, pp.133-158.
- Seidlhofer, B., 2005. English as a lingua franca. *ELT Journal*, 59(4), p.339.
- Selinker, L., 1972. Interlanguage. *International Review of Applied Linguistics*, 10, pp.209-231.

- Selinker, L., Swain, M., and Dumas, G., 1975. The interlanguage hypothesis extended to children. *Language Learning*, 25(1), pp.139-152.
- Shohamy, E., 1996. Competence and performance in language testing. In: G. Brown, K. Malkjmaer and J. Williams, eds. 1996. *Performance and competence in second language acquisition*. Cambridge: Cambridge University Press, pp.138-151.
- Sinclair, J., 2004. Corpus and text: Basic principles. In: M. Wynne, ed. 2004. *Developing Linguistic Corpora: A guide to good practice*. Oxford: Oxbow Books. pp.1-16.
- Smith, V.L. and Clark, H.H., 1993. On the course of answering questions. *Journal of Memory and Language*, 32, pp.25-38.
- Spolsky, B., 1968. Language testing: The problem of validation. *Tesol Quarterly*, 2(2), pp.88-94.
- Spolsky, B., 1989. Communicative Competence, Language Proficiency, and Beyond. *Applied Linguistics*, 10(2), pp.139-156.
- Stæhr, L.S., 2008. Vocabulary size and the skills of listening, reading and writing. *The Language Learning Journal*, 36(2), pp.139-152.
- Stern, H.H., 1983. *Fundamental Concepts of Language Teaching: Historical and Interdisciplinary Perspectives on Applied Linguistic Research*. Oxford: Oxford University Press.
- Stubbs, M., 2004. Language corpora. In: A. Davies and C. Elder, eds. 2004. *The handbook of applied linguistics*. Oxford: Blackwell Publishing, pp.106-132.
- Taylor, D.S., 1988. The meaning and use of the term 'competence' in linguistics and applied linguistics. *Applied Linguistics*. 9(2), pp.148-168.
- Taylor, L., 2011. *Examining speaking: Research and practice in assessing second language speaking*. Cambridge: Cambridge University Press.
- Thompson, P., 2004. Spoken language corpora. In: M. Wynne, ed. 2004. *Developing linguistic corpora: A guide to good practice*. PLACE: Oxbow books, pp.
- Thornbury, S. and Slade, D. 2006. *Conversation: From description to pedagogy*. Cambridge: Cambridge University Press.
- Timmis, I., 2002. Native-speaker norms and International English: A classroom view. *ELT Journal*, 56(3), pp.240-249.
- Timmis, I., 2005. Towards a framework for teaching spoken grammar. *ELT Journal*, 59(2), pp.117-125.
- Timmis, I., 2015. *Corpus linguistics for ELT*. Abingdon: Routledge.

- Tognini-Bonelli, E., 2001. *Corpus linguistics at work*. Amsterdam: John Benjamins Publishing.
- Tono, Y., 2003. Learner corpora: Design, development and applications. *Proceedings of the 2003 Corpus Linguistics Conference*. Pp.800-809.
- UCREL, (n.d.). Constituent Likelihood Automatic Word-tagging System (CLAWS). [online] Available at: <<http://ucrel.lancs.ac.uk/claws/trial.html>> [Accessed 31/05/2016].
- Underhill, N., 1987. *Testing spoken language: A handbook of oral testing techniques*. Cambridge: Cambridge University Press.
- University of Cambridge ESOL Examinations (CamESOL), 2011. *Using the CEFR: Principles of good practice*. [online] Available at: <<http://www.cambridgeenglish.org/images/126011-using-cefr-principles-of-good-practice.pdf>> [Accessed 12 August 2015].
- University of Central Lancashire (UCLan), n.d.-a *Ethics: e-Ethics System*. [online] Available at: <<https://www.uclan.ac.uk/students/research/ethics.php>> [Accessed 18 May 2015].
- University of Central Lancashire (UCLan), 2012a. *Ethical principles for teaching, research, consultancy, knowledge transfer and related activities*. [online] Available at: <https://www.uclan.ac.uk/students/research/files/Research_ethical_principles_Oct2012.pdf> [Accessed 18 May 2015].
- University of Central Lancashire Examination Board (UCLanEB), n.d.-a *12 Quality assurance of assessment*. [booklet] no date. s.n.
- University of Central Lancashire Examination Board (UCLanEB), n.d.-b *14 Examination handbook: Certificates in ESOL*. [booklet] no date. s.n.
- University of Helsinki, n.d. *CEFTTrain Project: CEFTTrain (Common European Reference for Languages in teacher training)*. [online] Available at: [Accessed 24 March 2016].
- Van Zeeland, H. and Schmitt, N. 2013. Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, 34(4), pp.457-479.
- WebLingua, 2014. *Text Inspector*. [online] Available at: <<http://www.textinspector.com/workflow>> [Accessed 30 April 2014].
- Weir, C.J., 2005a. *Language testing and validation: An evidence-based approach*. Basingstoke: Palgrave Macmillan.
- Weir, C.J., 2005b. Limitations of the Common European Framework for developing comparable examinations and sets. *Language Testing*, 22(3), pp.281-300.
- Widdowson, H.G., 1989. Knowledge of Language and Ability for Use. *Applied Linguistics* 10(2), pp.128-137.

- Widdowson, H. G., 1994. The ownership of English. *TESOL quarterly*, 28(2), pp.377-389.
- Wray, A., 2000. Formulaic sequences in second language teaching: Principle and practice. *Applied linguistics*, 21(4), pp.463-489.
- Wray, A., 2002. *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Young, R. and He, A. W., 1998. *Talking and testing: Discourse approaches to the assessment of oral proficiency*. Amsterdam: John Benjamins Publishing.
- Zhang, X. and Lu, X., 2014. A longitudinal study of receptive vocabulary breadth knowledge growth and vocabulary fluency development. *Applied Linguistics*, 35(3), pp.283-304.

9. APPENDICES

Contents:

1. UCLanESB exam structure – B1, B2 and C1
2. UCLanESB Marking criteria
3. UCLanESB Marking scheme
4. Number of UCLanESB exams taken at Preston campus
5. Nationality figures from international office
6. Advance notice email
7. Presentation for obtaining consent from pre-sessional learners
8. UCLanESB Consent form
9. Vocabulary profiles B1, B2 and C1
10. Text Inspector analysis – B1, B2 and C1
11. LINDSEI 20 most frequent words
12. Final can-do statements - B1, B2 and C1
13. Can-do statement verification exercise
14. Interaction can-do statements (seeking opinions) - B1, B2 and C1

Appendix 1: Exam structure

Table 64: UCLanESB speaking exam structure (UCLanEB, n.d.-b: 23, 26, 29)

B1 Test Format				
Exam section	Task description	Task purpose	Task duration	
			Two candidates	Three candidates
Part A	Interlocutor elicits personal information from candidates individually.	Eliciting of personal information, short questions and answers.	2 minutes	3 minutes
Part B	Interlocutor gives each candidate a written prompt to talk about on their own. Partner asked to comment.	Long turn for description, range of lexis and grammar.	2 x 2 minutes	3 x 2 minutes
Part C	Candidates are given another prompt based on the topic introduced in Part B and discuss it together. Interlocutor follows up with one short question to each candidate.	Expression of opinions, range of lexis, range of simple tenses, effective communication, accuracy.	4 minutes Total: 10 minutes	6 minutes Total: 15 minutes

B2 Test Format				
Exam section	Task description	Task purpose	Task duration	
			Two candidates	Three candidates
Part A	Interlocutor elicits personal information from candidates individually.	Eliciting of personal information, short/longer questions and answers on familiar topics.	3 minutes	4 minutes
Part B	Interlocutor gives candidates written prompts on a designated topic to discuss together. Interlocutor does not take part in the discussion.	Discursive section to engage candidates in a range of effective communication strategies. Accurate production of range of language.	4 minutes	6 minutes
Part C	Candidates respond to questions put by interlocutor and react to their partner's comments if invited to do so.	Expression of opinions, range of lexis and tenses, effective communication, accuracy.	5 minutes Total: 12 minutes	7 minutes 30 seconds Total: 17 min. 30 secs.

C1 Test Format				
Exam section	Task description	Task purpose	Task duration	
			Two candidates	Three candidates
Part A	Interlocutor asks candidates individual questions.	Eliciting personal information.	4 minutes	5 minutes
Part B	Interlocutor gives candidates a written prompt which they discuss together.	Discursive section to engage candidates in a range of effective communication strategies. Accurate production of range of language.	5 minutes	7 minutes
Part C	Interlocutor engages candidates in a question and answer session based on the same topic as in Part B.	Expression of opinions, range of lexis, range of tenses, effective communication, accuracy.	5 minutes Total: 14 minutes	7 minutes Total: 19 minutes

Appendix 2: UCLanESB marking criteria

Figure 56: B1 Marking criteria (UCLanEB, n.d.-b: 24)

Mark	Grammar	Vocabulary	Pronunciation	Discourse Management	Interactive Ability
B1 5	Structures mostly accurate for the level with only occasional minor slips.	Consistently demonstrates appropriate and extensive range of lexis for this level.	Use of stress and intonation puts very little strain on listener and individual sounds are articulated clearly. Utterances are consistently understandable.	Consistently makes extensive, coherent and relevant contributions to the achievement of the task.	Sustained interaction in both initiating and responding which facilitates fluent communication. Very sensitive to turn-taking.
4.5 4	More features of band 4 than band 5. Generally structurally accurate for the level but some non-impeding errors present.	Evidence of an extensive and appropriate range of lexis with occasional lapses.	Stress and intonation patterns may cause occasional strain on listener. Individual sounds are generally articulated clearly.	Contributions are generally relevant, coherent and of an appropriate length.	Meaningful communication is largely achieved through initiating and responding effectively. Hesitation is minimal and the norms of turn-taking are generally applied.
3.5 3	More features of band 4 than band 3. Reasonable level of structural accuracy but some impeding errors are acceptable.	Lexis is mostly effective and appropriate although range and accuracy are restricted at times.	Use of stress and intonation is sufficiently adequate for most utterances to be comprehensible. Some intrusive L1 sounds may cause difficulties for the listener.	Contributions are normally relevant, coherent and of an appropriate length but there may be occasional irrelevancies and lack of coherence.	Sufficient and appropriate initiation and response generally maintained throughout the discourse although there may be some undue hesitation. Turn-taking norms may not always be observed.
2.5 2	More features of band 3 than band 2. Frequent basic errors and a limited command of structure leading to misunderstandings.	Lexis is limited in terms of range and accuracy and may be inappropriate for the task.	Inadequacies in all areas of pronunciation put considerable strain on the listener.	Discourse is not developed adequately and may be incoherent and irrelevant at times.	Contributions limited and the patience of the listener may be strained by frequent hesitations. The norms of turn-taking are rarely observed.
1.5 1	More features of band 2 than band 1. Serious structural inaccuracy and lack of control which obscure intended meaning.	Insufficient or inappropriate lexis to deal with the task adequately.	Limited competence in all areas of pronunciation severely impedes comprehension.	Monosyllabic responses. Performance lacks relevance and coherence throughout.	Fails to initiate and/or respond. The interaction breaks down as a result of persistent hesitation. The norms of turn-taking are not observed.
0.5 0	More features of band 1 than 0. Too little speech to assess effectively.	Too little speech to assess effectively.	Too little speech to assess effectively.	Too little speech to assess effectively.	Too little speech to assess effectively.

Figure 57: B2 Marking criteria (UCLanEB, n.d.-b: 27)

Mark B2	Grammar	Vocabulary	Pronunciation	Discourse Management	Interactive Ability
5	Structures mostly accurate for the level with only occasional minor slips.	Consistently demonstrates appropriate and extensive range of lexis for this level.	Use of stress and intonation puts very little strain on listener and individual sounds are articulated clearly. Utterances are consistently understandable.	Consistently makes extensive, coherent and relevant contributions to the achievement of the task.	Sustained interaction in both initiating and responding which facilitates fluent communication. Very sensitive to turn-taking.
4.5	More features of band 4 than band 5.				
4	Generally structurally accurate for the level but some non-impeding errors present.	Evidence of an extensive and appropriate range of lexis with occasional lapses.	Stress and intonation patterns may cause occasional strain on listener. Individual sounds are generally articulated clearly.	Contributions are generally relevant, coherent and of an appropriate length.	Meaningful communication is largely achieved through initiating and responding effectively. Hesitation is minimal and the norms of turn-taking are generally applied.
3.5	More features for band 4 than band 3.				
3	Reasonable level of structural accuracy but some impeding errors are acceptable.	Lexis is mostly effective and appropriate although range and accuracy are restricted at times.	Use of stress and intonation is sufficiently adequate for most utterances to be comprehensible. Some intrusive L1 sounds may cause difficulties for the listener.	Contributions are normally relevant, coherent and of an appropriate length but there may be occasional irrelevancies and lack of coherence.	Sufficient and appropriate initiation and response generally maintained throughout the discourse although there may be some undue hesitation. Turn-taking norms may not always be observed.
2.5	More features of band 3 than band 2.				
2	More features of band 3 than band 2. Frequent basic errors and a limited command of structure leading to misunderstandings.	Lexis is limited in terms of range and accuracy and may be inappropriate for the task.	Inadequacies in all areas of pronunciation put considerable strain on the listener.	Discourse is not developed adequately and may be incoherent and irrelevant at times.	Contributions limited and the patience of the listener may be strained by frequent hesitations. The norms of turn-taking are rarely observed.
1.5	More features of band 2 than band 1.				
1	Serious structural inaccuracy and lack of control which obscure intended meaning.	Insufficient or inappropriate lexis to deal with the task adequately.	Limited competence in all areas of pronunciation severely impedes comprehension.	Monosyllabic responses. Performance lacks relevance and coherence throughout.	Fails to initiate and/or respond. The interaction breaks down as a result of persistent hesitation. The norms of turn-taking are not observed.
0.5	More features for band 1 than 0.				
0	Too little speech to assess effectively.	Too little speech to assess effectively.	Too little speech to assess effectively.	Too little speech to assess effectively.	Too little speech to assess effectively.

Figure 58: C1 Marking criteria (UCLanEB, n.d.-b: 30)

Mark	Grammar	Vocabulary	Pronunciation	Discourse Management	Interactive Ability
C1					
5	Comprehensive range of structures used accurately, appropriately and fluently. More or less error-free.	Consistently demonstrates a very extensive and comprehensive range of lexical competence.	Use of stress and intonation puts no strain on listener and individual sounds are articulated clearly. Utterances are consistently understandable.	Consistently makes extensive, coherent and relevant contributions to facilitate the achievement of the task.	Sustained interaction in both initiating and responding which facilitates natural, fluent communication. Very sensitive to turn-taking.
4.5	More features of band 4 than band 5.				
4	Wide range of structures used accurately, appropriately and fluently but minor non-impeding errors present.	Evidence of an extensive and appropriate range of lexis with only occasional lapses.	Stress and intonation patterns may cause minimal strain on listener. Individual sounds are generally articulated clearly.	Contributions are generally relevant, coherent and of an appropriate length. The task is dealt with effectively.	Meaningful communication is largely achieved through initiating and responding effectively. Hesitation is minimal and the norms of turn-taking are generally applied.
3.5	More features of band 4 than band 3.				
3	Reasonable range of structures used with generally consistent accuracy and fluency, but some non-impeding errors are evident.	Lexis is mostly effective and appropriate although range and accuracy are restricted at times.	Use of stress and intonation is sufficiently adequate for most utterances to be comprehensible. Some intrusive L1 sounds may cause difficulties for the listener.	Contributions are normally relevant, coherent and of an appropriate length but there may be occasional irrelevancies and incoherence. The task is dealt with satisfactorily.	Sufficient and appropriate initiation and response generally maintained throughout the discourse although there may be some undue hesitation. Turn-taking norms may not always be observed.
2.5	More features of band 3 than band 2.				
2	A rather limited range of structures. Consistent errors especially when attempting more ambitious grammatical forms.	Lexis is limited in terms of range and accuracy and lacking in precision.	Inadequacies in all areas of pronunciation put considerable strain on the listener.	Discourse is not developed adequately and may be incoherent and irrelevant at times. Task achievement is only partially realised.	Contributions limited and the patience of the listener may be strained by frequent hesitations. The norms of turn-taking are rarely observed.
1.5	More features of band 2 than band 1.				
1	Very limited range with serious structural inaccuracy and lack of flexibility.	Insufficient and inappropriate range of lexis to deal with the task adequately.	Limited competence in all areas of pronunciation severely impedes comprehension.	Monosyllabic responses. Performance lacks relevance and coherence throughout. Task achievement is very limited.	Fails to initiate and/or respond. The interaction breaks down as a result of persistent hesitation. The norms of turn-taking are not observed.
0.5	More features of band 1 than band 0.				
0	Too little speech to assess effectively.	Too little speech to assess effectively.	Too little speech to assess effectively.	Too little speech to assess effectively.	Too little speech to assess effectively.

Appendix 3: UCLanESB marking scheme

Figure 59: Marking scheme for B1, B2 and C1 (UCLanEB, n.d.-b: 31)

Mark	Global Scale
5	Demonstrated levels of linguistic competence and interactive skills indicate that tasks have been completed thoroughly and the message fully conveyed to the listener.
4.5	More features of band 4 than band 5.
4	Some characteristics of band 5 and band 3 in equal proportion.
3.5	More features for band 4 than band 3.
3	Demonstrated levels of linguistic competence and interactive skills indicate that tasks have been completed adequately and the message has been satisfactorily conveyed to the listener.
2.5	More features of band 3 than band 2.
2	Some characteristics of band 3 and band 1 in equal proportion.
1.5	More features for band 2 than 1.
1	Demonstrated levels of linguistic competence and interactive skills were insufficient to complete the task and the message was confused and very difficult for the listener to grasp.

Appendix 4: Number of UCLanESB exams taken at Preston campus

Table 65: Number of exams taken at Preston campus

Year	Date	Preston			
		B1	B2	C1	C2
2013	March	0	51	35	8
	May	2	75	42	0
	August	0	30	22	0
	December	15	29	21	0
	Total	17	185	120	8
2014	March	0	79	25	9
	June	3	72	27	0
	July	0	27	6	0
	September	0	8	0	0
	December	10	61	11	0
	Total	13	247	69	9
2015	March	6	27	29	0
	May	0	59	14	0
	August	0	28	2	0
	December	0	9	0	0
	Total	6	123	45	0
Total		36	555	234	17
Percentage		4.28	65.91	27.79	2.02

**Appendix 5: Nationality figures for students in the School of language, literature
and international studies**

Table 66: Nationality of students from the School of Languages, Literature and International Studies

Country	Number of Students	%
China	348	53.95
Oman	53	8.22
Germany	45	6.98
Saudi Arabia	23	3.57
Nigeria	21	3.26
Spain (includes Ceuta, Melilla)	20	3.10
Greece	13	2.02
France (includes Corsica)	11	1.71
India	8	1.24
Lithuania	7	1.09
Hong Kong (Special Administrative Region of China)	6	0.93
Italy (Includes Sardinia, Sicily)	6	0.93
Qatar	6	0.93
Cyprus (European Union)	5	0.78
Ireland	5	0.78
Poland	5	0.78
Others*	63	9.77
Total	645	100.00

Appendix 6: Advance notice email

Figure 60: Advanced notice email

Dear candidate,

In order to develop and improve examinations, a study is being carried out to obtain spoken test data that will be used for research and publication. It is hoped that the study will result in the development of better test materials and will help teachers and students know what is expected to achieve a pass in the B1, B2 and C1 level exams. This will have no effect on your exams or studies but it will hopefully help future students and their experience of sitting exams. We guarantee that **all** test data obtained for this purpose will:

- a) Be used anonymously so that you cannot be identified
- b) Not impact on your test result
- c) And will remain confidential

On the day of your speaking test, you will be asked to complete a consent form to show if you are happy for us to use your data or if you prefer not to be included in the study. If you do **not** want to take part in the research, please indicate on the **consent form** that you wish not to be included.

Please be aware that if you consent to your data being used but then change your mind, you **can** withdraw from the study. In order to withdraw, contact Shelley by email (sbyrne@uclan.ac.uk) **within 7 days** of you completing your speaking test. Your test data will be taken out and will not be used. Due to the nature of the research, it will not be possible to remove data after the 7 day period.

Thank you for reading this email. If you have any further questions, please send Shelley an email and she will respond as quickly as possible.

Appendix 7: Presentation for obtaining consent from pre-sessional learners

Figure 61: Presentation for pre-sessional students

SPEAKING EXAM RESEARCH
Shelley Byrne

With this research, I hope to:

- Find out what language students use in speaking exams
- Examine whether the language changes as the level increases
- Help future students and teachers know what language learners need to know to pass exams

Possible concerns

I will be embarrassed if I don't pass or get a low mark
Only successful exams will be listened to and used so don't worry about your mark.

If I do/do not take part in the research, it will affect my final speaking mark
Absolutely not. The research has no effect on your exam mark.

People will be able to identify me from the results/data
NO names will be used when the data is reported

Anybody will be able to listen to the recording
No. **Only Shelley**, and people helping with the research, will listen to the exams. Nobody else will have access to them.

All exams will be anonymous – no names will be used

<\$0> First I'm going to ask you some questions about yourselves okay? Okay erm <\$5F> what is the most popular free time activity in your country?
<\$5F> In my country? Erm I think it's computer game for most boy and shopping for most girls yeah.
<\$0> Mhm erm <\$6F> what's the most popular free time activity in your country?
<\$6F> Er I think er maybe erm sometimes it's most er the same as her but er er I sometimes I like stay at home yeah.
<\$0> Okay erm how do you like to spend your weekends <\$5F>?
<\$5F> Erm you know I have a dog yes I very like pets so I will spend all my weekend to erm play with my dog yeah.
<\$0> And <\$6F> how do you spend your weekends?
<\$6F> Er er maybe I will travel with my friends er you know? We always travel er city erm kinda.

Giving consent

- You will receive a permission form.
- Complete the consent form with all your details.
- Shelley will be the only person who sees these. Make sure you tick (✓) to give/decline consent

Exam format

- There are 3 parts:
 - Part 1 = 2 minutes
 - Part 2 = 4 minutes
 - Part 3 = 4 minutes
- 10 minutes in total

YOUR COOPERATION WILL BE GREATLY APPRECIATED. THANK YOU!

Appendix 8: UCLanESB consent form

Figure 62: UCLanESB consent form

<u>Research Permission Form</u>	
Dear candidate,	
So that we can continue to develop and improve the examinations, we would like to use samples from written and spoken tests in our research including possible publication. We can guarantee that when we use test data for these purposes that:	
<ul style="list-style-type: none"> (a) Samples of test data will be used anonymously (names will be taken out) so that you cannot be identified; (b) Use of any test data will not affect your test result. 	
Please support the development of our examinations by ticking 'Yes' below and providing us with the necessary information.	
I give permission for data from my tests to be used for the research purposes outlined above.	Yes, I give permission: ___ No, I do not give permission: ___
Signature:	
Surname:	
First Name(s):	
Date of Birth:	Day/Month/Year
Gender:	Please tick Male: ___ Female: ___
Nationality:	
First language(s):	
How many years have you studied English in total? (both here and in other countries?)	_____ years
How long have you spent in English-speaking countries	_____ years _____ months
If we wish to release video data as part of a publication, we will contact you via e-mail to request this. Please could you provide an e-mail address for us.	
E-mail :	

Appendix 9: Vocabulary profile: B1

Table 67: B1 vocabulary profile

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token %
K-1 Words :	515 (63.50)	715 (66.88)	13751 (<u>93.33</u>)	93.33
K-2 Words :	161 (19.85)	189 (17.68)	483 (<u>3.28</u>)	96.61
K-3 Words :	55 (6.78)	60 (5.61)	238 (<u>1.62</u>)	98.23
K-4 Words :	26 (3.21)	27 (2.53)	40 (<u>0.27</u>)	98.50
K-5 Words :	19 (2.34)	20 (1.87)	43 (<u>0.29</u>)	98.79
K-6 Words :	10 (1.23)	10 (0.94)	21 (<u>0.14</u>)	98.93
K-7 Words :	7 (0.86)	8 (0.75)	11 (<u>0.07</u>)	99.00
K-8 Words :	8 (0.99)	8 (0.75)	11 (<u>0.07</u>)	99.07
K-9 Words :	1 (0.12)	1 (0.09)	1 (<u>0.01</u>)	99.08
K-10 Words :	3 (0.37)	3 (0.28)	5 (<u>0.03</u>)	99.11
K-11 Words :	2 (0.25)	3 (0.28)	9 (<u>0.06</u>)	99.17
K-12 Words :	2 (0.25)	2 (0.19)	3 (<u>0.02</u>)	99.19
K-13 Words :	1 (0.12)	1 (0.09)	1 (<u>0.01</u>)	99.20
K-14 Words :				
K-15 Words :				
K-16 Words :				
K-17 Words :	1 (0.12)	1 (0.09)	1 (<u>0.01</u>)	99.21
K-18 Words :				
K-19 Words :				
K-20 Words :				
Off-List:	??	59 (5.52)	115 (<u>0.78</u>)	99.99
Total (unrounded)	811+?	1069 (100)	14733 (100)	100.00

RELATED RATIOS & INDICES

Pertaining to whole text

Words in text (tokens):	14733
Different words (types):	1069
Type-token ratio:	0.07
Tokens per type:	13.78

Pertaining to onlist only

Tokens:	14618
Types:	1010
Families:	811
Tokens per family:	18.02
Types per family:	1.25

Table 68: B2 vocabulary profile

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token %
K-1 Words :	595 (62.57)	844 (67.09)	17710 (<u>93.51</u>)	93.51
K-2 Words :	205 (21.56)	235 (18.68)	748 (<u>3.95</u>)	97.46
K-3 Words :	66 (6.94)	75 (5.96)	133 (<u>0.70</u>)	98.16
K-4 Words :	30 (3.15)	33 (2.62)	62 (<u>0.33</u>)	98.49
K-5 Words :	18 (1.89)	20 (1.59)	46 (<u>0.24</u>)	98.73
K-6 Words :	10 (1.05)	10 (0.79)	25 (<u>0.13</u>)	98.86
K-7 Words :	5 (0.53)	5 (0.40)	12 (<u>0.06</u>)	98.92
K-8 Words :	3 (0.32)	3 (0.24)	4 (<u>0.02</u>)	98.94
K-9 Words :	3 (0.32)	3 (0.24)	4 (<u>0.02</u>)	98.96
K-10 Words :	5 (0.53)	5 (0.40)	7 (<u>0.04</u>)	99.00
K-11 Words :	3 (0.32)	3 (0.24)	26 (<u>0.14</u>)	99.14
K-12 Words :	3 (0.32)	3 (0.24)	4 (<u>0.02</u>)	99.16
K-13 Words :	3 (0.32)	4 (0.32)	11 (<u>0.06</u>)	99.22
K-14 Words :				
K-15 Words :	1 (0.11)	1 (0.08)	1 (<u>0.01</u>)	99.23
K-16 Words :				
K-17 Words :				
K-18 Words :	1 (0.11)	1 (0.08)	2 (<u>0.01</u>)	99.24
K-19 Words :				
K-20 Words :				
Off-List:	??	53 (4.21)	144 (<u>0.76</u>)	100.00
Total (unrounded)	951+?	1258 (100)	18939 (100)	100.00

RELATED RATIOS & INDICES

Pertaining to whole text

Words in text (tokens):	18939
Different words (types):	1258
Type-token ratio:	0.07
Tokens per type:	15.05

Pertaining to onlist only

Tokens:	18795
Types:	1205
Families:	951
Tokens per family:	19.76
Types per family:	1.27

Table 69: C1 vocabulary profile

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token %
K-1 Words :	607 (58.48)	927 (63.71)	19307 (92.25)	92.25
K-2 Words :	236 (22.74)	285 (19.59)	976 (4.66)	96.91
K-3 Words :	78 (7.51)	90 (6.19)	243 (1.16)	98.07
K-4 Words :	44 (4.24)	50 (3.44)	96 (0.46)	98.53
K-5 Words :	28 (2.70)	28 (1.92)	47 (0.22)	98.75
K-6 Words :	14 (1.35)	14 (0.96)	31 (0.15)	98.90
K-7 Words :	8 (0.77)	8 (0.55)	13 (0.06)	98.96
K-8 Words :	9 (0.87)	10 (0.69)	12 (0.06)	99.02
K-9 Words :	3 (0.29)	4 (0.27)	4 (0.02)	99.04
K-10 Words :	4 (0.39)	4 (0.27)	6 (0.03)	99.07
K-11 Words :	1 (0.10)	1 (0.07)	5 (0.02)	99.09
K-12 Words :	4 (0.39)	4 (0.27)	6 (0.03)	99.12
K-13 Words :				
K-14 Words :				
K-15 Words :				
K-16 Words :				
K-17 Words :				
K-18 Words :				
K-19 Words :	2 (0.19)	2 (0.14)	2 (0.01)	99.13
K-20 Words :				
Off-List:	??	73 (5.02)	181 (0.86)	99.99
Total (unrounded)	1038+?	1455 (100)	20929 (100)	100.00

RELATED RATIOS & INDICES

Pertaining to whole text

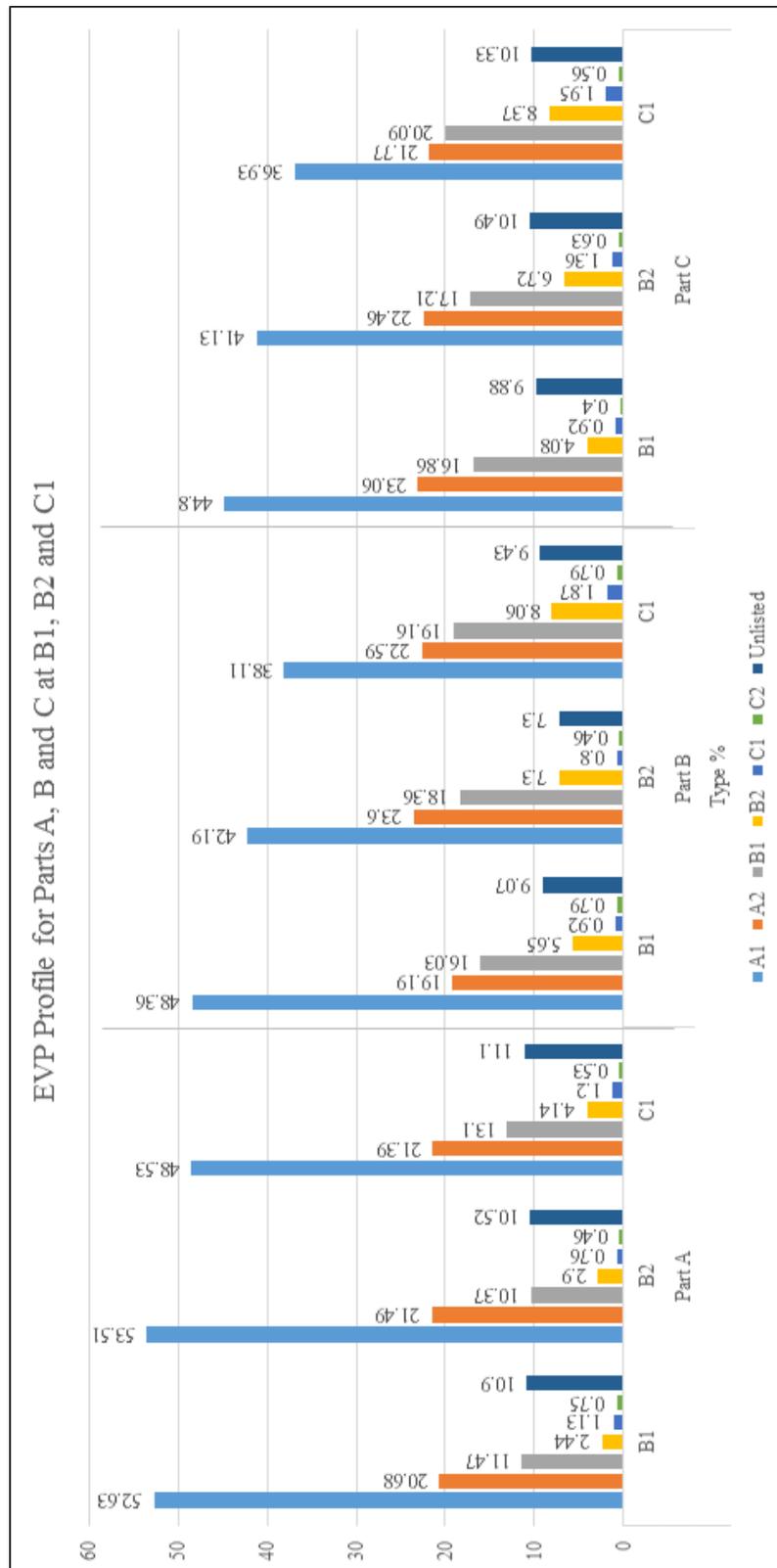
Words in text (tokens):	20929
Different words (types):	1455
Type-token ratio:	0.07
Tokens per type:	14.38

Pertaining to onlist only

Tokens:	20748
Types:	1382
Families:	1038
Tokens per Family:	19.99
Types per Family:	1.33

Appendix 10: *Text Inspector* analysis of B1, B2 and C1 speech across exam parts

Figure 63: Learner speech at B1, B2 and C1 according to CEFR levels



Appendix 11: LINDSEI 20 most frequent words

Table 70: LINSEI 20 most frequent words

N	Word	Freq.	%	Texts	%
1	I	37,060	4.67	1	100
2	THE	30,978	3.91	1	100
3	AND	30,149	3.8	1	100
4	ER	23,925	3.02	1	100
5	TO	21,608	2.72	1	100
6	A	17,189	2.17	1	100
7	IN	13,130	1.66	1	100
8	IT	12,833	1.62	1	100
9	EH	12,369	1.56	1	100
10	OF	10,688	1.35	1	100
11	ERM	10,354	1.31	1	100
12	WAS	10,278	1.3	1	100
13	THAT	10,257	1.29	1	100
14	SO	10,075	1.27	1	100
15	YOU	9,842	1.24	1	100
16	BUT	9,058	1.14	1	100
17	IS	8,944	1.13	1	100
18	IT'S	7,776	0.98	1	100
19	YEAH	7,498	0.95	1	100
20	THEY	7,335	0.92	1	100

Appendix 12: Final can-do statements

Table 71: B1 can-do statements

B1	
CEFR category	Can-do statement
Production	Can briefly give reasons and explanations for opinions, plans and actions.
	Can describe dreams, hopes and ambitions.
	Can relate the plot of a book or film and describe his/her reactions.
	Can give detailed accounts of experiences, describing feelings and reactions.
	Can relate a straightforward narrative or description as a linear sequence of points.
Interaction	Can express agreement and disagreement politely.
	Can seek personal views and opinions in discussing topics of interest.
	Can give brief comments on the views of others.
	Can express his/her thoughts about abstract or cultural topics such as music, films. Can explain why something is a problem.
	Can express and respond to feelings such as surprise, happiness, sadness, interest and indifference.
	Can maintain a conversation or discussion but may sometimes be difficult to follow when trying to say exactly what he/she would like to.
	Can express personal opinions and exchange information on topics that are familiar, of personal interest or pertinent to everyday life (e.g. families, hobbies, work, travel and current events).
Strategies (production)	Can start again using a different tactic when communication breaks down.
	Can correct mix-ups with tenses or expressions that lead to misunderstandings provided the interlocutor indicates there is a problem.
	Can define the features of something concrete for which he/she can't remember the word. Can convey meaning by qualifying a word meaning something similar (e.g. a truck for people = bus).
Strategies (interaction)	Can summarise the point reached in a discussion and so help focus the talk.
	Can invite others into the discussion.
	Can ask someone to clarify or elaborate what they have just said.
	Can repeat back part of what someone has said to confirm mutual understanding and help keep the development of ideas on course. Can invite others into the discussion.
	Can exploit a basic repertoire of language and strategies to help keep a conversation or discussion going.
	Can initiate, maintain and close simple, face-to-face conversation on topics that are familiar or of personal interest.
	Can intervene in a discussion on a familiar topic, using a suitable phrase to get the floor.

Table 72: B2 can-do statements

B2	
CEFR category	Can-do statement
Production	Can give clear, detailed descriptions and presentations on a wide range of subjects related to his/her field of interest, presenting it as a linear sequence of points.
	Can develop an argument systematically with appropriate highlighting of significant points, and relevant supporting detail.
	Can explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
Interaction	Can convey degrees of emotion and highlight the personal significance of events and experiences.
	Can account for and sustain his/her opinions in discussion by providing relevant explanations, arguments and comments.
	Can outline an issue or a problem clearly, speculating about causes or consequences
Strategies (production)	Can use circumlocution and paraphrase to cover gaps in vocabulary and structure.
	Can correct slips and errors if he/she becomes conscious of them or if they have led to misunderstandings.
	Can make a note of 'favourite mistakes' and consciously monitor speech for it/them.
Strategies (interaction)	Can intervene appropriately in discussion, exploiting appropriate language to do so.
	Can initiate discourse, take his/her turn when appropriate, and end conversations when he/she needs to, though he/she may not always do this elegantly.
	Can use stock phrases (e.g. 'That's a difficult question to answer') to gain time and keep the turn whilst formulating what to say.
	Can help the discussion along on familiar ground, inviting others in, etc.
	Can ask follow-up questions to check that he/has understood what a speaker intended to say, and get clarification of ambiguous points.

Table 73: C1 can-do statements

C1	
CEFR category	Can-do statement
Production	Can give clear, detailed descriptions of complex subjects.
	Can integrate sub-themes, develop particular points and round off with an appropriate conclusion.
Interaction	<i>(As B2)</i> Can outline an issue or a problem clearly, speculating about causes or consequences, and weighing advantages and disadvantages of different approaches.
	<i>(As B2)</i> Can help along the progress of the work by inviting others to join in, say what they think, etc.
Strategies (production)	Can backtrack when he/she encounters a difficulty and reformulate what he/she wants to say without fully interrupting the flow of speech.
	<i>(As B2+)</i> Can use circumlocution and paraphrase to cover gaps in vocabulary and structure
Strategies (interaction)	<i>(As B2)</i> Can ask follow-up questions to check that he/she has understood what a speaker intended to say, and get clarification of ambiguous points.
	Can relate own contribution skillfully to those of other speakers.
	Can select a suitable phrase from a readily available range of discourse functions to preface his/her remarks appropriately in order to get the floor, or to gain time and keep the floor whilst thinking.

Appendix 13: Can-do statement verification exercise

Figure 64: Can-do statement verification exercise

Instructions			
<p>You will receive one exam script (at B1, B2 or C1) and 8 can-do statements from the CEFR for the corresponding level. Please read the exam and highlight any candidate language that you think satisfies each of the statements. Highlighting the language can be done by hand or on the computer but please remember:</p>			
<ol style="list-style-type: none">1. to highlight all of the words which satisfy the statement. Highlighting one word will not signal where the selected language starts or ends2. to distinguish between different statements. Using the same colour/highlight for each statement will make it difficult to see which statement the language matches.3. that some language can satisfy more than one statement. If this does happen in your exam script, please make it clear i) that more than one statement is indicated, and ii) which statements the language relates to.			
<p>I have provided an example below in which statements have been highlighted in different colours as I found printing the exam and using highlighter pens was easiest. Do use any technique you wish though.</p>			
Example			
B1 Production can-do statements:			
<table border="1"><tbody><tr><td style="background-color: #00FFFF;">Can describe dreams, hopes and ambitions.</td></tr><tr><td style="background-color: #00FF00;">Can briefly give reasons and explanations for opinions, plans and actions.</td></tr></tbody></table>	Can describe dreams, hopes and ambitions.	Can briefly give reasons and explanations for opinions, plans and actions.	
Can describe dreams, hopes and ambitions.			
Can briefly give reasons and explanations for opinions, plans and actions.			
Language from exam script:			
<p><\$IM> Yeah I would like to start my own er business because I can I can er plan it by myself.</p>			
<p>Thanks once again for agreeing to help with this. I really do appreciate it.</p> <p>Kind regards, Shelley</p>			

Appendix 14: Interaction can-do statements: Seeking opinions

Table 74: B1 seeking opinion questions

Question type	Freq.	%	Example
Do/Did you...?	9	25.00	Do you like sports that's easy or hard to learn like bodybuilding or going to walking?
			Do you think meeting er new people is er good or erm er interesting when you are when you meet er a good new people?
			Do you agree with me?
			Do you have?
			Do you think it?
			Are you diary every day?
			Do you see the Did you watch the...?
			Do do have you recommend do you recommed a best movie film?
What do you think about...?	3	8.33	And what do you think about easy to learn?
			What do you think about this?
			What do you think about the food?
Which of these reasons/choice/ of the following...?	3	8.33	Which of the reasons you will to watch a film?
			Erm erm which choice you want to start a new sports er easy to learn?
			Which of the following would make you watch a film?
What's your favourite...?	3	8.33	mhm what's your favourite film star in the film?
			Oh what's your erm favourite film you have ever seen?
			For you what's your favourite film?
What/Which kind of...?	2	5.56	Which kind of cartoon you like?
			What kind of person make a good friends?
How about...?	2	5.56	Okay how about you?
			How about you?
How do you think...?	2	5.56	How do you think er when a special events happens we we take photo to remember it?
			Yeah er another how do you think the advantages of keeping a diary? Are you diary every day?
What is the best thing	1	2.78	What is the best thing about travelling abroad?
Have you...?	1	2.78	Have you read some film er summary book erm where when they have done some films?
You know?	1	2.78	You know?
...yeah?	1	2.78	Without the friend you will to see because you love love the star and you feel already read the book which is based on yeah?
Who is...?	1	2.78	Okay who's the star?
Really?	1	2.78	Really?
What about...?	1	2.78	What about?
What else?	1	2.78	What else?
You don't like...?	1	2.78	You don't like any film star?
Other	3	8.33	Yeah so when you say some advertisement in television or magazine you er maybe take some attractive to watching the movie.
			So I can I can teacher you badminton how to play?
			Any questions?

TOTAL 36

Table 75: B2 seeking opinion questions

Question type	Freq.	%	Example
Do/Did you...?	15	27.78	Er do you like er hot weather or really cold weather?
			Did you know?
			Do you agree?
			Do you agree with me?
			Do you agree with it?
			Er do you think do you think do sports outside is healthier than in indoors sport centre?
			And do you think there are still some dangers there are still some dangers sports to er for young people?
			Do you recommend some indoor sports for me?
			So do you like yoga or playing table tennis?
			Mm do you think of some gyms or like bad gyms?
			Do you like gyms?
			Do you enjoy your life now?
			Er mm er do you think er do you think listening to music or playing musical instruments can help you to reduce stress?
Er do you think erm music is not useful subject for studying at school?			
Mm do you have something something you like to listen?			
What do you think about/of...?	14	25.93	Erm do what do you think about the weather?
			So what do you think of lifestyle? Healthy lifestyle?
			but er what do you think of er these sports?
			What do you think this is a very good thing for you?
			What do you think about learning about learn about different culture or and learning like new language?
			so what what do you think er er as I mean erm is it good for you?Is it learn something new?
			What is the different between your culture and er er what what another culture?
			Okay what do you think about the first topic?
			What do you think about it?
			So what do you think about it if we just refer to these two words?
			So what do you think?
			What do you think about this?
			and er what do you think about it?
Yeah so what do you think about the er young people they feel bored when they stay at home?			
Okay er what do you think about the children?			
How about...?	9	16.67	And er mm how about this one?
			So how about the food?
			How about you?
			How about you? What do you think of it?
			How about I found the advanced technology it changed our life in a positive way?
How...?	4	7.41	How will you know the job if you don't have qualifications?
			How we can communicate
			How you will know about the culture and how he will know?
What about...?	3	5.56	What about you?
			What about you?
			Yes what about you?
Why?	2	3.70	So why?
Why?	2	3.70	Why?
...yeah?	2	3.70	Yeah?
So you can learn from culture yeah?	2	3.70	So you can learn from culture yeah?
Have you...?	1	1.85	So you think now you have enjoyed the present er but er have you have you worried about your future?
What else...?	1	1.85	So erm what else can you think can just improve the teamwork?
What...would you...?	1	1.85	Can you give me an answer what what advice would you give er to a foreign student who is suffering from culture shock?
Can you...?	1	1.85	Er and a new ideas is very good for the business but er can you think er from any other side?
Other	1	1.85	They prefer the old or modern?

TOTAL: 54

Table 76: C1 seeking opinion questions

Question type	Freq.	%	Example
What do you think about/of...?	17	23.61	And what do you think about a reasonable price?
			What do you think about the reviews of the hotel?
			I used to I used to see the rating of it like I don't go through the wording itself I just see erm like if going to state it out of ten what do you think about it er so yeah so.
			Okay what do you think?
			So what do you think about these unhealthy aspects?
			Okay mm what do you think that of these options will be best will be suitable for improve the road safety?
			What do you think?
			What do you think?
			What do what do you think about that?
			Yeah and this might affect the hotel rating what do you think about the hotel rating also?
			So tourism in my country is not really good. What about?
			And erm erm what do you think the energy efficient application in homes?
			What do you think?
			All right then how about the location of the hotel?
			And so others what about mm reviews of the hotel?
What do you think what do you what do you think?			
We can study life language from the film what do you think?			
Do/Did you...?	7	9.72	Like do you spend time going through it see what other erm other people stayed in the hotel thinks about the hotel itself or?
			Do you think if we visit the places where it happened er it will still be the same?
			And do you try to be like him?
			Do you agree with me?
			Do you like sea?
How about...?	6	8.33	Do you know pet? Pet overpopulation? Can I talk about pet population?
			Do you have any city in mind?
			Hello. How about the first one?
			How about the second point?
			And how about number three? Sports centres and events?
Which .../of these reasons/choice/of the following...?	5	6.94	And how about the cheap travel around the country?
			Yeah of course er this one is important but er erm how about the comfort of the hotel room?
			How about you?
			Which three we use for the choose as being the most important for travelling?
			Which one is the?
How do you think...?	4	5.56	Which one do you think is suitable?
			Er which part?
			Okay in this er which two do you think is very important?
			Sites and historical interests. How do you think this?
What do you...?	3	4.17	And then is the location of the hotel. How do you think so?
			How do you think about this?
			And how do you think?
Why?	3	4.17	and what do you feel movies change?
			Erm what do you admire most about him?
			Er erm how what do you reckon about this one?
...yeah?	3	4.17	Why?
			Why?
			Why is he famous?
Have you...?	2	2.78	Yeah?
			Yeah?
So...?	2	2.78	Yeah?
			Have you ever been on a tour to a museum?
How...it is...?	2	2.78	That's out of nature and er er have you been to the forest?
			So you are not the type of person who travels a lot?
What about...?	2	2.78	So you learned a lot?
			Erm alright fine er how's it important that the location of the hotel how important is it to you the location of the hotel?
How much...?	1	1.39	Yeah the place where you would spend the night over yeah how important do you think it is?
			What about finding information on the internet?
Right?	1	1.39	What about interviewing people?
			Like if there was a hotel in London how much are you going are you looking for it to be for a four star hotel in London?
Who...?	1	1.39	Alright and how do you feel about the hotel amenities how is it important to you?
			Yeah I do the same and most importantly first is the hotel room right?
Can we...?	1	1.39	and er who do you think is famous citizen in Dubai or Emirates?
			What else do you think about these?
No?	1	1.39	Can we change the countryside?
			No?
Other	8	11.11	Yes yeah and er shall we focus on the other element?
			And where do you read about him or how do you learn different facts about him?
			Yeah you know my country um where are you from?
			What are effect about this?
			I heard about that if you don't if you are not Muslim you can't go there is that a lie?
			Hot?
Is there a sea there?			
Is it the kind of Dubai?			
There have been any more?			

TOTAL: 72