Exploring Arab learners' use of lexical collocations of English: A corpus-based study

Reem Roghailan M. Alshammari

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

School of Languages, Cultures and Societies

30th May 2023

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

Publications

- **Alshammari, R. 2021. Adverb-adjective collocation use by Arab EFLLs and British English native speakers: a comparative corpus-based study. *The Language Scholar Journal*, *9*, pp.24-104.**

Poster and conference presentations:

- **"A Corpus-Based Comparison of Lexical Collocation Use amongst Native English Students and Arab Learners of English" presented**
- **a poster presentation in the University of Leeds PGR Conference 2020.**

- **Attended the *17th Corpus Linguistics in the South conference* which took place at the University of Rohampton, London on 24th November 2018.**

- **Presented a paper presentation at the *1st Symposium on Language Education and Research* that took place within the premises of the University of Greenwich, London on 18th and 19th February 2023. The conference paper entitled "*A comparison of lexical collocation use amongst native and Arab learners of English – A corpus-based study".***

Acknowledgements

I would like to express my sincere gratitude to my supervisors Dr. James Wilson, Prof. Michael Ingleby, and Prof. James Dickins for their invaluable guidance, encouragement, and support throughout my work. Their insightful feedback, constructive criticism, and unwavering dedication have been instrumental in the completion of this PhD. I am particularly grateful for their willingness to share their expertise, time, and resources to help me navigate through the challenges I encountered during this journey. Their patience, understanding, and encouragement have been a constant source of motivation and inspiration. I feel incredibly fortunate to have had the privilege of working with such amazing supervisors, and I will always be grateful for their contribution to my success. Thank you for everything.

The data utilized in this study was collected from two corpora, the Arab Learner English Corpus (ALEC) and the BUiD Arab Learner Corpus, and I would like to acknowledge the creators and maintainers of these resources for making them publicly available.

During my MA degree, I was introduced to the exciting field of corpus linguistics, a relatively new area of study at the time. The course was challenging, but I found it fascinating, and to my surprise, I received an unexpected grade in the Corpus Linguistics Module. One of the most influential figures in my education during this time was Professor Paul Baker, who taught me with such passion and insight. Professor Baker is a well-respected authority in the field of Corpus Linguistics, and his influence inspired me to pursue a PhD in this area, particularly because it was not being taught in Saudi Arabia at that time.

I would like to express my deepest gratitude to my father, Dr Roghailan and my mother for their unwavering support, encouragement, and love throughout my academic journey. Without their sacrifices and constant guidance, I would not have been able to achieve my goals and complete this thesis. Their belief in me and their unwavering encouragement has been a constant source of strength and

motivation. I am truly blessed to have such amazing parents, and I am forever indebted to them. Thank you for everything.

I would like to express my sincere gratitude to my brother Eng. Abdulaziz, who has been living with me in Leeds for the past four years, for his unwavering support, encouragement, and companionship during my PhD journey. I will not forget all the support you have shown to me and to Lana.

I would like to extend my sincere thanks to my brothers (Sultan, Mobarak, Mohammed and Omar) and sisters (Hifa and Raneem), for their unwavering support and encouragement during the journey of completing this PhD.

I would like to express my heartfelt gratitude to my husband Abdulelah and two children Lana and Abdullah for their unwavering love, support, and patience throughout my academic journey. Their constant encouragement and understanding have been a constant source of motivation for me. Their love has given me the strength to persevere and pursue my academic goals, even when the going got tough. I am truly blessed to have such a loving and supportive family. Thank you for always being there for me, and for your endless patience and understanding. I could not have done this without you.

I cannot deny the deep sadness and longing I feel for my son, Abdullah. Although I have not been able to be with him every day, he is always in my thoughts and my heart. I miss his laughter, his hugs, and his presence in my life. Thank you to my mother for taking care of Abdullah. Thank you for my father for keeping me connected with my son by sharing photos and videos to be part of his life in every way possible.

Thank you …

Transliteration system used in this thesis

Arabic letter	Trans-liter-ation	Phonetic description
ا	ʾ *or* ā	voiced glottal stop/fricative *or* open unrounded long vowel
ب	b	voiced bilabial stop
ت	t	voiceless apico-dental stop
ث	t̲	voiceless interdental fricative
ج	j	voiced dorso-prepalatal stop
ح	ḥ	voiceless pharyngeal fricative
خ	x	voiceless post-dorso-post-velar fricative
د	d	voiced apico-dental stop
ذ	d̲	voiced interdental fricative
ر	r	apico-alveolar trill/tap
ز	z	voiced apico-dental fricative
س	s	voiceless apico-dental fricative
ش	š	voiceless dorso-prepalatal fricative
ص	ṣ	voiceless emphatic apico-alveolar fricative
ض	ḍ	voiced emphatic apico-alveolar stop
ط	ṭ	voiceless emphatic apico-alveolar stop
ظ	d̲̣	voiced emphatic interdental fricative
ع	ʕ	voiced pharyngeal fricative
غ	ġ	voiced post-dorso-post-velar fricative
ف	f	voiceless labio-dental fricative
ق	q	voiceless uvular plosive
ك	k	voiceless post-dorso-velar stop
ل	l	apico-alveolar lateral
م	m	bilabial nasal
ن	n	apico-alveolar nasal

ه	h	voiceless glottal fricative
و	w *or* ū	bilabial glide *or* back close rounded long vowel
ي	y *or* ī	dorso-palatal glide *or* front close unrounded long vowel

The transliteration system used in this work is adapted from the DIN system (for which, see http://en.wikipedia.org/wiki/DIN_31635). The system is almost the same as the DIN (Deutsches Institut für Normung / German Institute for Standardisation Registered Association) system (cf. https://en.wikipedia.org/wiki/Romanization_of_Arabic; https://en.wikipedia.org/wiki/DIN_31635). Regarding the consonants, it differs from the DIN system in the following respects: 'j' is used instead of the DIN 'ǧ', 'x' is used instead of the DIN 'ḫ' for خ, and 'ḏ' is used instead of the DIN 'ẓ' for ظ. The use of 'j' and 'x' is more standard for transliterations in English-language texts than the 'ǧ' and 'ḫ' of the DIN system. 'ḏ' is used for the emphatic interdental fricative, rather than the DIN 'ẓ', because the subscript line under 'ḏ' parallels the subscript lines under the other two Arabic interdentals: 't̲' for ث, and d̲ for ذ. 'ẓ' might also suggest an emphatic apico-dental fricative (paralleling non-emphatic 'z') rather than an emphatic interdental fricative.

The vowels are: *a* for ـَ (open unrounded short vowel); *i* for ـِ (front close unrounded short vowel); *u* for ـُ (back close rounded short vowel); *ā* for ـَا (open unrounded long vowel); *ī* for ـِي (front close unrounded long vowel); *ū* for ـُو (back close rounded long vowel); *ay* for the diphthong ـَيْ (open unrounded short vowel, followed by dorso-palatal glide); and *aw* for the diphthong ـَوْ (open unrounded short vowel, followed by bilabial glide).

The definite article الـ is constantly written as *al-* in all instances, regardless of whether it assimilates to the following consonant, or whether the initial *a* disappears after a previous vowel. A (ّ) *šaddah* results in a geminate (consonant written twice). A (ة) (*tā' marbūṭa)* is transcribed as word-final *-a* or *-at*. A (ى) *ʾalif maqṣūra* appears as *ā*, rendering it indistinguishable from *alif*. Nunation (*tanwīn*)

is ignored in transliteration except where it would unavoidably be pronounced in speech. I have adopted Dickins' "*sakkin taslam* approach" which he translates as "'leave off the (normally) vocalic endings, and you'll be okay" (2020, p.31). This is because the case and mood system of Standard Arabic is complex and can be a hazard for even educated Arabs as they might sometimes get the nunation wrong.

Transliterations of other authors' names are kept in the same forms given by these authors. In the case of some well-known names of people and places that have a standard English transliteration-type form, this form has been retained.

Abstract

This PhD thesis focuses on investigating the use of four different lexical collocations by Arab English foreign language learners (Arab EFLLs) from Kuwait and Dubai, and native British English students (NBESs) from the UK. The four types of collocation are the Adverb-Adjective, Verb-Noun, Verb-Adverb, and Adjective-Noun collocations. The main objective of this research is to identify issues with the use of lexical collocations by Arab EFLLs, who often struggle to produce natural collocations due to the influence of their L1. The study uses a corpus-based frequency approach to analyze data collected from two comparable corpora consisting of academic essays from both native speakers of British English and Arab EFLLs. The corpus data for Arab EFLLs is derived from two corpora, the Arabic Learner English Corpus (ALEC) and the BUiD Arab Learner Corpus (BALC), while the corpus data for native speakers is collected from the British Academic Written English Corpus (BAWE). The log-likelihood statistical test was used to identify significantly different collocations between Arab EFLLs and NBESs. The research also employed a contrastive error analysis approach to understand the difficulties that Arab EFLLs face when using lexical collocations. This study provides valuable insights into the use of lexical collocations by Arab EFLLs and highlights the impact of Arabic on their use of collocations in academic writing. The findings of this research can be used to develop targeted interventions that can help Arab EFLLs improve their use of lexical collocations in their academic essays.

Keywords: corpus linguistics, lexical collocations, word frequency, EFL, first language (L1) interference, non-parametric tests

Table of Contents

Publications ..... **i**
Acknowledgements ..... **i**
Transliteration system used in this thesis ..... **iii**
Abstract ..... **vi**
Table of Contents ..... **vii**
List of Figures ..... **xii**
List of Tables ..... **xiv**
Abbreviations ..... **xviii**
Foreword ..... **xix**
Chapter 1 Introduction **1**
Introduction **1**
1.1 Interest in this thesis 2
1.2 Aims and Objectives 3
1.3 Research Statement 4
1.4 Importance of the Study 5
1.5 Research Gap and Contribution 8
1.6 Research Questions and Research Hypothesis 10
1.7 Ethical Approval 11
1.8 Thesis Structure 12
Chapter 2 An Overview of the Literature on Lexical Collocation **13**
Introduction **13**
2.1 Definition of Concepts and Terms 13
2.1.1 'Corpus Linguistics' and the use of 'Corpora' 13
2.1.2 The benefits of Corpus Linguistics for EFL 15
2.1.3 Learner corpora 16
2.1.4 Reference/general corpora 17
2.1.5 Raw frequency (RF) and normalised frequency (NF) 18
2.1.6 Token and Type 19
2.1.7 Concordances 19
2.1.8 Node word and collocate 20
2.1.9 Lemma and word-form 21
2.1.10 First language, target language and foreign language 22
2.1.11 Defining the construct of 'collocations.' 22
2.1.11.1 Collocations from the perspective of Semantics and the concept of 'naturalness' in collocations 27

2.1.11.2 How 'collocation' is understood in this thesis.....29
2.1.11.3 Word frequencies and collocation identification criteria in corpus-based studies.....32
2.2 Theoretical Framework.....34
2.2.1 L1 Interference theory.....34
2.2.2 Error analysis (EA) theory.....36
2.2.2.1 Accuracy and grammatical complexity of collocation.....37
2.3 A Review of Collocation.....39
2.3.1 Adverb-Adjective collocation.....42
2.3.2 Verb-Noun Collocation.....44
2.3.3 Verb-Adverb Collocation.....46
2.3.4 Adjective-Noun Collocation.....47
2.4 Previous Studies on the use of collocation by learners of English.....51
2.4.1.1 Corpus-based error analysis studies of EFLLs' use of collocations.....53
2.4.1.2 Corpus-based error analysis studies of the use of collocation by Arab EFLLs.....56
2.5 Conclusion.....58
Chapter 3 Methodology.....60
Introduction.....60
3.1 Methodological Approaches.....61
3.1.1 Benefits of a Two level, Mixed-methods Approach.....62
3.1.2 The Two-level Mixed-methods Approach.....63
3.1.2.1 Level 1 - The corpus-based frequency approach.....63
3.1.2.2 Level 2 - The contrastive error analysis approach.....64
3.1.2.3 Level 2 extended – The descriptive-contextual approach.....65
3.2 Sample and Sampling Design.....69
3.2.1 Corpus data.....69
3.2.2 Access to existing corpora.....72
3.3 Tools.....75
3.3.1 General research tools.....75
3.3.1.1 Microsoft Excel.....75
3.3.1.2 SPSS.....75
3.3.1.3 Jisc online survey tool.....75
3.3.2 Corpus-based tools.....76
3.3.2.1 The Keyword-in-Context (KWIC) concordance.....76

3.3.2.2 Frequency lists and the compare frequency function 78
3.3.2.3 The collocation function 80
3.4 Statistical tests used in this thesis 82
3.4.1 Ordinal Statistics and choices made for the available collocation association measures 82
3.4.1.1 Justification for LL as the primary collocation association measure 84
3.4.2 Descriptive Statistics 85
3.4.2.1 Maximum and Minimum value 86
3.4.3 Inferential Statistics 86
3.4.3.1 Data distribution 86
3.4.3.2 Mann-Whitney U test 87
3.4.3.3 Kruskal-Wallis H test 89
3.5 Quantitative and Qualitative Procedures 90
3.5.1 The procedure for the quantitative analysis 90
3.5.1.1 Collocation identification and extraction in this thesis 91
3.5.1.2 Statistical analysis of the collocations 92
3.5.2 The procedure for the qualitative analysis 92
3.6 Exploratory Study: Determining corpora size effect on the collocation patterns in the Arab EFLLs and the NBES corpora 98
3.6.1 Method for determining corpus size effect 99
3.6.2 Results for exploratory study of corpus size effect 100
3.6.2.1 Descriptive statistics results (Frequency approach) 100
3.6.2.2 Inferential Statistics Results 101
3.6.3 Concluding remarks on the experimental study 104
3.7 Conclusion 105
Chapter 4 Results 106
4.1 A Corpus-based frequency analysis comparing collocation use in British and American English 106
Introduction 106
- Null hypothesis (H 107
4.1.1 Previous literature on the differences between BE and AE in the use of collocations 108
4.1.2 Methods 109
4.1.2.1 Procedure 110
4.1.2.2 Research corpora used 110
4.1.2.3 Research tools 110
4.1.3 Results 111

4.1.3.1 Statistical analysis....113
4.1.4 Discussion....116
4.1.5 Limitations and conclusion....117
4.2 The results of the comparative corpus-based frequency analysis of the use of lexical collocations by Arab EFLLs and NBESs....118
Introduction....118
4.2.1 Quantitative results: a Corpus-based Frequency analysis..118
4.2.1.1 Verb-Noun collocation results....120
4.2.1.1.1 Statistical Analysis Results for Verb-Noun collocations Results....122
4.2.1.1.2 Statistical Analysis Results for Verb-(Intermediate)-Noun collocations....127
4.2.1.1.3 Statistical Analysis Results for to + Verb-Noun Collocations....130
4.2.1.1.4 Statistical Analysis Results for to + Verb-(intermediate)-Noun collocations....132
4.2.1.2 Verb-Adverb Collocation Results....135
4.2.1.3 Adjective-Noun Collocation Results....140
4.2.2 Measuring word choice complexity in collocation use through the CEFR....147
4.2.2.1 CEFR System for Collocations (CEFR-Coll)....152
4.2.3 Qualitative results: Contrastive Error-analysis (CEA)....156
4.2.3.1 Native-speaker Rater Judgment of the Collocations...156
4.2.3.2 The Contrastive-Error analysis of the Lexical Collocations....159
4.2.3.2.1 CEA Results for the Adverb-Adjective Collocations....160
4.2.3.2.2 CEA Results for the Verb-Noun Collocations....162
4.2.3.2.3 CEA Results for the Verb-Adverb Collocations....175
4.2.3.2.4 CEA Results for the Adjective-Noun Collocations 179
4.3 Conclusion....183
Chapter 5 Discussion....184
Introduction....184
5.1 Answers to the research questions....184
5.1.1 First and Second Research Questions (RQ 1)....184
5.1.2 Second Research Questions (RQ 2)....188
5.1.3 Third Research Question (RQ 3)....189
5.2 Conclusion....197

Chapter 6 Conclusion 200
Introduction 200
6.1 Contribution 200
6.1.1 Contribution to the pedagogy 200
6.1.2 Contribution to EFL Curriculum Designers 202
6.2 Limitations 204
6.3 Future Research 205
References 209
Appendix A 235
A.1 Frequency analysis tables for the comparison of collocations in British and American English 235
A.2 Appendix for the results of the corpus-based frequency analysis of lexical collocations produced by Arab EFLLs and NBESs 249
A.3 The native raters' results: 334
A.4 Native Rater's Data for Verb-Noun Collocation 336

List of Figures

Figure 2-1: Screenshot of concordance lines for the node word *make* 20

Figure 2-2: A screenshot of the possible noun collocates for the verb *make* 21

Figure 3-1: A screenshot of the *compare frequency* function in IntelliText.....78

Figure 3-2: A screenshot of the output of the compare frequency function in IntelliText. 79

Figure 3-3: A screenshot for the collocation query function taken from IntelliText.....80

Figure 3-4: Screenshot of the collocation query tool output in IntelliText.....81

Figure 3-5: The combined Contrastive Error Analysis framework used in this thesis (adapted from Gass and Selinker, (2008)).....93

Figure 4-1: Histogram of *WORK-Noun* collocations 114

Figure 4-2: Histograms for the Verb-Noun collocation frequencies in Arab EFLL and NBES corpora 125

Figure 4-3: Histograms for the Verb-Noun collocation frequencies allowing intermediates in Arab EFLL and NBES corpora 128

Figure 4-4: Mann-Whitney *U* test for the use of the Verb-Noun collocation when allowing intermediates in Arab EFLLs and NBES corpora. 129

Figure 4-5: Histograms for the frequencies of the use of *to* + Verb - Noun collocation in Arab EFLL and NBES corpora 131

Figure 4-6: Mann-Whitney *U* test for the *to* + Verb-Noun Collocation for Arab EFLL and NBES corpora 132

Figure 4-7: Histograms for the frequencies of the use of *to* + Verb-(intermediates)-Noun collocations in the Arab EFLL and NBES corpora 133

Figure 4-8: Mann-Whitney *U* test for the use of *to* + Verb-(intermediates)-Noun collocation in Arab EFLL and NBES corpora 134

Figure 4-9: Histogram showing the distribution of Verb-Adverb collocations for Arab EFLL and NBES corpora.....138

Figure 4-10: Mann-Whitney *U* test for the use of Verb-Adverb collocation for Arab EFLLs and NBESs 140

Figure 4-11: Histograms showing the distribution of the Adjective-Noun collocation for Arab EFLLs and NBESs 143

Figure 4-12 CEFR levels.....147

Figure 4-13: Suggested Methodological Framework for CEFR-Coll....155

Figure 4-14: A screenshot is taken from IntelliText web interface for *'BE + gum'* in the AEFLL Corpus. 170

Figure A.1-1: Histograms showing the distribution of the noun collocates for the verb *MAKE* in BE06, NBES and Arab EFLLs...240

Figure A.1-2: Histograms showing the distribution of the noun collocates for the verb *HAVE* in BE06, NBES and Arab EFLLs. ..240

Figure A.1-3: Histograms showing the distribution of the noun collocates for the verb *DO* in BE06, NBES and Arab EFLLs.....240

Figure A.1-4: Histograms showing the distribution of the noun collocates for the verb *GET* in BE06, NBES and Arab EFLLs.....241

Figure A.1-5: Histograms showing the distribution of the noun collocates for the verb *TAKE* in BE06, NBES and Arab EFLLs. ..241

Figure A.2-1: Concordance lines for to+Verb-Noun collocations.....262

Figure A.3-1: Overview of the native raters'responses on the Verb-Noun collocations 334

Figure A.3-2: The native-speaker acceptability results for the Verb-Adverb Collocations 334

Figure A.3-3: The native-speaker acceptability results for the Adjective-Noun Collocations.....335

List of Tables

Table 1-1: Possible equivalents for the Arabic word ضمان *damān* in English (Izwaini, 2016, 320). 7

Table 2-1: Possible combinations of parts of speech in lexical collocations in English (Combinations examined for this research are marked by ♦) 31

Table 2-2: Semantic accuracy and Lexical complexity measures 39

Table 2-3: Some Arabic examples (taken from Brashi (2005, pp.37-39))41

Table 2-4: Summary of the use of the four lexical collocations in Arabic and English 49

Table 3-1 Summary of information about the corpora selected for this PhD research project 71

Table 3-2 The learner corpora sizes 72

Table 3-3: Learners' levels in the Arab EFLLs Corpus 74

Table 3-4. The relative frequency (per million words) for the five verbs within the three corpora 100

Table 3-5: One-way ANOVA for the five collocation sets 102

Table 3-6: Statistical significance testing for the five verbs (*MAKE, HAVE, DO, GET, and TAKE*) in the BE06, BAWE and Arab EFLLs corpora 103

Table 4-1: Normalised Frequency Comparison Results: BE vs. AE Usage in the Selected Corpora 107

Table 4-2: Most frequent verbs in BNC and I-EN....111

Table 4-3: Most frequent nouns in BNC and I-EN....112

Table 4-4: Most frequent adjectives in BNC and I-EN 113

Table 4-5: Normal distribution in the collocation sets investigated for the comparison between BE and AE....114

Table 4-6: Percentage of Verbs of with threshold noun *collocates* in Verb-Noun Collocations 122

Table 4-7: Total number of Verb-Noun collocations within Arab EFLL and NBES corpora....123

Table 4-8 Verb-Noun collocations in the Arab EFLL and NBES corpora 124

Table 4-9: Descriptive statistics for the use of Verb-Noun collocations 125

Table 4-10: Levene's Test of Homogeneity of Variances for Verb-Noun collocations 126

Table 4-11: Independent Sample t-Test of the Verb-Noun collocation127

Table 4-12: Descriptive statistics for the use of Verb-(Intermediate)-Noun collocations 128

Table 4-13: Levene's Test of Homogeneity of Variances for Verb-(Intermediates)-Noun collocations 129

Table 4-14: Descriptive statistics for the use of *to* + Verb-Noun collocations 130

Table 4-15: Levene's Test of Homogeneity of Variance for *to* + Verb-Noun collocations 131

Table 4-16 Descriptive statistics for the use of *to* + Verb-(Intermediates)-Noun collocations 133

Table 4-17: Levene's Test of Homogeneity of Variances for *to* + Verb-(Intermediate)-Noun collocations 134

Table 4-18 Verb-Adverb collocations in the Arab EFLL and NBES corpora 136

Table 4-19: Overall descriptive statistical results for the Verb-Adverb collocations 137

Table 4-20: Percentage of Verbs of with threshold adverb *collocates* in Verb-Adverb Collocations 137

Table 4-21: Levene's Test of Homogeneity of Variances for Verb-Adverb collocations 139

Table 4-22: Adjective-Noun collocations in the Arab EFLL and NBES corpora 141

Table 4-23 Overall descriptive statistical results for the Adjective-Noun collocations by Arab EFLLs and NBESs 142

Table 4-24: Percentage of Adjectives with threshold *noun collocates* in Adjective-Noun collocations 142

Table 4-25 Levene's Test of Homogeneity of Variances for Adjective-Noun collocations 144

Table 4-26: The Independent Sample t-Test of the Adjective-Noun Collocation 144

Table 4-27: Overview of the statistical significance testing for the collocations. 145

Table 4-28: CEFR categorization of the Adverb-Adjective collocation 148

Table 4-29: CEFR categorization the Verb-Noun Collocations 149

Table 4-30 CEFR categorization for the Verb-Adverb collocations 150

Table 4-31: CEFR categorization for the Adjective-Noun collocations 151

Table 4-32: Overview of the native-speaker judges' response to the collocations investigated in this study 157

Table 4-33: Contrastive Error analysis types in the AEFLL Corpus 159

Table 4-34: The LL ratios of Verb-Noun collocations allowing intermediates 163

Table 4-35: List of the *to* + Verb-Noun Collocation with LL scores for Arab EFLLs and NBESs 164

Table 4-36: List of the *to* + Verb-(intermediates)-Noun Collocation with their LL ratio for Arab EFLLs and NBES 165

Table 4-37: Examples of Noun collocates in the Verb-Noun collocation in both corpora 169

Table 4-38: Verbs collocating with 'plastic surgery' in the BNC..... 173

Table 4-39 Examples of delexical verbs used in the Verb-Noun collocation. 175

Table 4-40: The total frequencies of the use of *break fast vs. breakfast in the English native speaker corpora 177

Table 4-41: Normalised frequencies of the adverb 'very' 179

Table A-1: The noun collocates for the verb lemma *MAKE* in BE06 and NBES (BAWE corpus) 235

Table A-2 The noun collocates for the verb lemma *HAVE* in BE06 and NBES (BAWE corpus) 236

Table A-3 The noun collocates for the verb lemma *DO* in BE06 and NBES (BAWE corpus) 237

Table A-4 The noun collocates for the verb lemma *GET* in BE06 and NBES (BAWE corpus) 238

Table A-5 The noun collocates for the verb lemma *TAKE in* BE06 and NBES *(BAWE corpus)* 239

Table A-6: Verb-Noun collocations in the BNC and I-EN..... 242

Table A-7:Noun-Noun Collocations in the BNC and I-EN 244

Table A-8: Adjective-Noun collocations in the BNC and I-EN..... 246

Table A-9: Mann-Whitney U Test for Verb-Noun collocations in BE and AE 248

Table A-10:Mann-Whitney U Test for Noun-Noun collocation in BE and AE 248

Table A-11: Mann-Whitney U Test for Adjective-Noun collocations in BE and AE 248

Table A-12: The normalised frequencies of the selected 200 *Verb* lemmas in the Arab EFLLs corpus 249

Table A-13: The normalised frequencies of the selected 200 *Verb* lemmas in the NBESs Corpus 250

Table A-14: Extracted Verb-Noun Collocations with LL scores for the Arab EFLLs Corpus 252

Table A-15: Extracted Verb-Noun collocations with LL scores for the NBESs Corpus. 253

Table A-16: Extracted Verb-Adverb Collocations with LL scores for the Arab EFLLs corpus 263

Table A-17: Extracted Verb-Adverb Collocations with LL scores for the NBESs Corpus.....264

Table A-18: The normalised frequencies of *Adjectives* in the Arab EFLLs corpus 272

Table A-19: The normalised frequencies of *Adjectives* in the NBESs Corpus.....273

Table A-20: Extracted Adjective-Noun Collocations with LL scores for the Arab EFLLs corpus.....275

Table A-21: Extracted Adjective-Noun Collocations with LL scores for the NBESs Corpus 284

Abbreviations

AEFLL	Arab English as a Foreign Language Learner
AME06	American English corpus from 2006
BE06	British English corpus from 2006
BNC	British National Corpus
CEA	Contrastive Error Analysis
CEFR	The Common European Framework of Reference for Languages
CEFR-Coll	CEFR Modification for Collocations
EFFLs	English as a Foreign Language Learners
I-EN	Internet texts in English
IPMW	Instances per million words
L1	First language
L2	Second language (English)
NBESs	Native British English-speaking Students

Foreword

The Arabic language is perceived as one of the major languages used in the world alongside Greek and Latin due to its historical and cultural importance, its central role in a world religion[1], its number of speakers and geographical spread (Chejne, 1969, 3). There is an increasing demand among Arabs for learning English as a foreign language to communicate with other people from around the world. It is thus important, interesting, and useful to investigate the influence of Arabic on how Arab learners of English learn and use lexical collocations. The thrust of this thesis is that lexical collocation is a key element to mastering a language, and that it requires a complex level of control of the language and a high level of proficiency. Most of the studies of EFLLs have shown that using collocations poses a difficulty, as these central elements within a language require knowledge of its vocabulary, particularly the ability to select the appropriate words for a given context. Their correct use indicates a level of language proficiency close to that of a near-native user of that language--the more learners produce correct collocations, the more proficient they appear to be in that language. The use of lexical collocation has received a lot of research attention as it is viewed as troublesome not only for language learners but also for translators (Nesselhauf, 2003; Laufer and Girsai, 2008).

---

[1] The importance of Arabic goes back to its being the language of Islam and Muslims from ancient times, neatly summed up in the title of Jaspal, R. and Coyle, A. 2010. “Arabic is the language of the Muslims–that's how it was supposed to be”: exploring language and religious identity through reflective accounts from young British-born South Asians. *Mental Health, Religion and Culture.* **13**(1), pp.17-36. In Islam, the Holy Quran, written in Arabic, which Muslims refer to as the “Preserved Tablet [Āllwḥ ĀlmāHfwḍ]” ‘اللوح المحفوظ’(Nasr, S.H. 2015. *The Study Quran. A new Translation and Commentary.* New York: HarperOne, ibid.) is the miraculously preserved revelation from Allah (God) which has remained constant throughout its history. The Holy Quran had, and still has, a huge influence on defining Arabic Grammar (see Gholitabar, M. and Kamali, A.D. 2012. The Quran and the development of Arabic linguistics. *2012 International Conference on Language, Medias and Culture.* **33**, pp.26-30, ibid.). The central role of the Quran explains Arabs’ familiarity with Standard Arabic whatever contemporary dialect of Arabic they speak in whatever region of the world they may live in.

Chapter 1 Introduction

Introduction

This study aims to shed light on Arab learners' use of English collocations by comparing the usage of lexical collocations by Arab learners of English as a foreign language (henceforth Arab EFLLs) from Kuwait and Dubai with the usage of lexical collocations by native British English-speaking students (henceforth NBESs). To gather data for the Arab EFLLs, the study uses a combination of two corpora, namely the Arabic Learner English Corpus (ALEC) and the BUiD Arab Learner Corpus (BALC). The corpus data for NBESs is obtained from the British Academic Written English Corpus (BAWE). The three corpora consist of undergraduate and school student essays composed for academic English purposes. The study focuses on four types of lexical collocation in learners' academic essays and considers the following collocations:

(i) **Adverb-Adjective collocations** (e.g. *closely related*)
(ii) **Verb-Noun collocations** (e.g. *pay attention*) including the following:
- **Verb-(intermediate)-Noun collocations** (e.g. *book an appointment*)
- **(to) + Verb-Noun collocations** (e.g. *to pay attention*)
- **(to) + Verb-(intermediate)-Noun collocations** (e.g. *to book an appointment*)

(iii) **Verb-Adverb collocations** (e.g. *appreciate sincerely*)
(iv) **Adjective-Noun collocations** (e.g. *best friends*)

The collocations will be extracted and then investigated using statistical significance tests to determine whether there are any significantly differences in frequency between the Arab EFLLs and NBESs (§4.2.1). Statistical significance tests will be conducted after consulting two English reference corpora which are the BNC and INTERNET-EN to determine the occurrence rates of the collocations extracted (§2.1.4 for more information about the used reference corpora). This step will assist in determining the level of appropriateness in a given context and help in identifying and analyzing discrepancies between the collocations' statistical 'overuse/underuse' and their actual acceptability. The

Arab EFLLs' choice of lexical items were subjected to a further check: two English reference corpora were used to obtain sample usages of each collocation. Native English-speaking raters will be recruited to judge a sample of the selected collocations used by the Arab EFLLs to determine their acceptability (§4.2.3.1). Finally, the level of difficulty of the collocations will be assessed by applying the CEFR classification of the words involved in the collocation.

This process will involve multiple steps to ensure the accuracy and appropriateness of the extracted collocations. Classifying the lexical items used in the collocations, and ranking them for difficulty according to the CEFR levels (§4.2.2). This CEFR difficulty score provides a standardized measure that will help in the creation of teaching materials. Furthermore, a contrastive error analysis will be performed to investigate the causes of the faulty collocations produced by Arab EFLLs (§4.2.3.2). Overall, I hope that this approach will be effective in informing the design of teaching materials and pedagogical strategies to help Arab EFLLs choose appropriate and effective lexical items for use in their English language learning.

1.1 Interest in this thesis

My decision to focus on written essays for my research stems from my experience teaching the Advanced Writing Module at Princess Norah Bint Abdulrahman University. As a teacher, I noticed that many of my students struggled with writing, and I recognized the need to improve our teaching practices and develop a more effective Advanced Writing curriculum. I believe that this research can contribute to the advancement of teaching practices for writing in higher education, a vital language skill that all students must master.

My journey towards obtaining my PhD has allowed me to deepen my knowledge so that I will be able to apply my research findings to practical teaching methods. I am grateful for the twin opportunity to contribute to the field of Corpus Linguistics and to make a positive impact on the education system in Saudi Arabia.

1.2 Aims and Objectives

Based on my experience of studying and teaching English in Saudi Arabia, I have noticed a lack of emphasis on the teaching of collocations. Instead, vocabulary is often taught as stand-alone words without considering their possible collocates. This gap in language instruction is what initially motivated me to undertake this thesis. The aim of the current research is to address this instructional gap by fulfilling the following objectives:

- To investigate the differences between Arab EFLLs and NBESs in their usage and their frequencies of usage of the four collocation types listed above (p.1).

- To test the validity of the hypotheses formulated for the four collocation types investigated in this study.

- To clarify whether there is a direct link between the use of collocations and language proficiency. Such a link may originate in learners' attempts to avoid using specific words during writing, or, conversely, their overuse of them. This will be assessed based on the categorisation of word difficulty laid out in the Common European Framework of Reference for Languages (CEFR).

- To determine the acceptability of Arab EFLLs' collocations through the use of native English-speaking raters/judges, and two English reference corpora.

- To identify the commonly occurring collocational errors of Arab EFLLs, in order to suggest some improvements to existing teaching methods and materials.

- To identify whether and how interference from the Arab EFLLs' L1 affects their use of lexical collocations.

1.3 Research Statement

A corpus-based study comparing collocations used by native English speakers and Arabic learners of English can be potentially beneficial for understanding the influence of their first language on learners resulting from structural and lexical differences between the learners' L1 and L2. Relatively little is understood about the sources of Arab EFLLs' "difficulties" in producing accurate collocations (Mohamed and Eltayeb, 2016). Learners' difficulties in producing collocations are clearly visible in the products of speaking and writing (Brashi, 2009, p.22; El-Dakhs, 2015, p.61; Khoja, 2019, p.2). Taking this into account, the study aims to reveal the sources of interference between Arabic as an L1 and English as an L2. To explain this, let us consider the example of 'commit a crime' as a Verb-Noun collocation. I have noticed that most of my Arab compatriots use the verb *do* or *make* to explain this criminal act. The justification behind this choice of word is L1 interference, as 'did a crime' is said as ارتكب جريمة (*artakab jarīma*) in Arabic (Al-Shormani and Al-Sohbani, 2012; Sabah, 2022).

One of the primary challenges for learners of English when it comes to collocations is the vast number of possible combinations. Research has shown that the number of English collocations far exceeds the number of lexical items (Shin and Nation, 2008, p.339; Habtoor and Al-Swaidan, 2019, p.725). For example, the BBI Combinatory Dictionary of English includes over 70,000 combinations and phrases under 14,000 entries. Similarly, Collins COBUILD English Words in Use presents roughly 100,000 examples of collocation arranged around 5,000 headwords derived from the core vocabulary of contemporary English (Bahns, 1993). These figures highlight the significant challenge learners face in mastering English collocations and underscore the importance of developing effective teaching methods to support their acquisition.

Fluency is often used in EFL teaching to describe learners' performance in a given language (Chambers, 1997, p.535), but its precise definition remains elusive. Brumfit argues that fluency, as natural language use, is "The maximally effective operation of the language system so far acquired by the students" (1984, p.57). While fluency is typically associated with spoken language, Hughes

defines it more generally as "the ability to express oneself in a reasonable, intelligible, and accurate manner without hesitation. Otherwise, communication will break down because the listeners will lose interest" (2002, p.14). In language learning, the value of collocations is viewed as key to achieving fluency (Wang and Good, 2007, p.1; Uchihara et al., 2022). Correct usage of collocations by learners demonstrates that a more advanced level of language.

1.4 Importance of the Study

A small number of studies have been conducted on the use of collocation by Arab EFL students (Mahmoud, 2005; Shehata, 2008; Farooqui, 2016; Alzi'abi, 2017). Collocation is believed to impose a certain burden on language acquisition as many studies have found that even advanced language learners tend to face difficulties when using collocations (Laufer and Waldman, 2011, p.654; Chang, 2018, p.9); yet this burden is not restricted to language learners. Translators also view collocations as a 'trouble spot' during the process of translation particularly when collocation types vary in both languages (Bahumaid, 2006, p.152; Izwaini, 2016, p.307). This view regarding collocation competence has been supported by several recent studies that confirm that collocation is a problematic issue (Izwaini, 2016; Alqaed, 2017; Demir, 2017) not restricted to Arab learners of English as it is also evident for learners of the following language backgrounds: Indonesian, German, Hebrew, Chinese, Indian, Turkish and Korean, respectively (Moehkardi, 2002; Nesselhauf, 2005; Laufer and Waldman, 2011; Ke, 2013; Agah and Soori, 2015; Demir, 2017; Chang, 2018). One interesting finding of Bahumaid's study is that translators have a particular problem translating culturally-specific collocations (2006, p.149). All the translators participating in that study failed to correctly produce a translation for 'duly authorized' in English, and الخلفاء الراشدون (āl-xulafāʾ al-rrāšdūn) was translated correctly as 'orthodox caliphs' by only one of the translators. The current study aims to add to the literature on the use of collocations by Arab EFLLs, and explore the reasons for the difficulties they encounter, including possible interference from Arabic. A previous study by Mahmoud (2005) showed that Arabic speakers of English tend to incorrectly use lexical collocations because of negative transfer from Arabic to English. Some of the lexical collocations in Mahmoud's study, 224 out of 269

collocations were found to be incorrect, the error rate being relatively high at 83.27%. Errors were perceived within two types of lexical combinations, *verb + noun* and *adjective + noun* collocations, in which either the whole combination of words was faulty, or one word was in an incorrect position, such as 'repair his mistake' and 'large thinking', instead of 'correct his mistake' and 'thinking big'. In Arabic, the collocation 'correct his mistake' is either صحح خطأه or صلح خطأه (*ṣaḥaḥa xaṭā'hu* or *ṣalḥa xaṭā'hu*); as the word 'correct' is understood in Arabic as either 'to correct' or 'to repair'. Therefore, this could be a result of negative transfer from the learners' L1. In the second example, 'large thinking', the confusion could be a result of the Arabic superlative and comparative construction in the form of افعل (*āfʕl*), the word 'large' having many equivalents in Arabic such as 'big' and 'huge'. Regarding the other types of collocation errors that Mahmoud found were in the use of incorrect prepositions in the following examples '*affect in health' instead of 'affects health' and '*addicted by' instead of 'addicted to' proving that the use of English prepositions is a problematic and challenging element for Arab learners of English. This is in line with Baheej (2015), Faris and Sahu (2013), Sabah (2015) who found that English prepositions are a source of difficulty for Arabic learners of English. The preposition difficulty for Arab EFLLs is due to the limited number of prepositions in Arabic amounting to 20 prepositions compared to English which has approximately 150 prepositions (Essberger, 2000).

Arabic has a large and diverse number of lexical collocations that need special attention during the process of translation into English, taking into consideration each collocation's "pragmatic, semantic and stylistic profiles" (Izwaini, 2016, p.326). According to many translation studies, the best process for collocation translation is to start with translating the node word and then translating the word it collocates with (Izwaini, 2016). This explains the difficulty that Arab learners encounter with collocations in that there are often many equivalents for a single Arabic word in the English language. Hence, this might lead them to produce unnatural collocations. Izwaini (2016, p.320) exemplifies this problem with the example of the many possible English equivalents for the word ضمان (*damān)* such as 'guarantee', 'insurance', 'security' and 'bond', shown in table 1-1

Table 1-1: Possible equivalents for the Arabic word ضمان *damān* in English (Izwaini, 2016, 320).

Term/version	*Arabic*	*English*
Bond	□ **ضمان** العطاء ***(ḍamān** al-ʿaṭā)ʾ* □ **ضمان** الأداء ***(ḍamān** al-ʾadā)*	□ Bid ***bond*** □ performance ***bond***
Security	□ **ضمان** السندات ***(ḍamān** al-sanadāt)* □ **ضمان** اجتماعي ***(ḍamān** iğtimāʿiy)*	□ bond ***security*** □ social ***security***
Guarantee	□ **ضمان** الحقوق (***ḍamān** al-ḥuqūq)* □ **ضمان** مالي ***(ḍamān** māliy)*	□ the ***guarantee*** of rights □ financial ***guarantee***
Other equivalents	□ **ضمان** صحي ***(ḍamān** ṣiḥiy)* □ **ضمان** الجودة ***(ḍamān** al-ğawdaẗ)*	□ health ***insurance*** □ quality ***warranty***

The diverse equivalents above for the Arabic word ضمان *ḍamān* explain some of the difficulties Arab EFLLs encounter in terms of choosing the right word for a particular context in English.

A major aspect of the difficulty is that the grammatical structures of Arabic lexical collocations differ from those in English. This can be seen in the case of Verb-Noun collocations which are often translated into a Verb-Preposition-Noun collocation in Arabic (Faris and Sahu, 2013, p.54). The translation of this structure poses some difficulties as it differs and is "quite unpredictable" in terms of how it is arranged in the second language (Faris and Sahu, 2013, p.56). Some examples of this structure given by Faris and Sahu (2013) are the following:

- 'pay a visit' is translated to يقومُ بِالزِيَارةِ (*yaqūmu bi al-zziyārati*) '*pay in a visit'
- 'shake hands' is translated to one word in Arabic as يُصافح (*yuṣāfḥ*) 'to shake'.

Other scholars, such as Aziz (1982), have observed that there could be equivalents to English collocations in Arabic but that differ in grammatical structure. For example, the Adjective-Noun collocation is said in reverse order in Arabic (i.e., Noun-Adjective), so that 'net weight' is said as in Arabic as الوزن الصافي(*al-wazn al-ṣṣāfī)* 'weight net' (Aziz, 1982; Faris and Sahu, 2013, p.53)

The above differences are just a sample. The importance of this study lies in the increased understanding it will bring of the differences between English and Arabic.

1.5 Research Gap and Contribution

First of all, to the best of my knowledge, all of the authors who have investigated lexical collocations have done so through the use of other types of methods including manual compilation of collocations in learners' essays (El-Gazzar, 2006), or administering questionnaires where learners were asked to fill in gaps with the correct collocation (El-Dakhs, 2015), or using multiple-choice tests (Brashi, 2009; Mohamed and Eltayeb, 2016). Some studies, such as (Diab, 1997), were not clear about how they identified collocations in learners' essays.

I am not the first scholar to have used a corpus-based approach of Arab use of collocations in EFL. Dukali (2018) used AntConc 3.2.1w to extract collocations, with a specific focus on only 12 node words for both verbs and adjectives in two types of collocation: Verb-Noun and Adjective-Noun. Additionally, Farooqui (2016) did a corpus-based study of the use of collocations but in learners from many different language backgrounds. She used theses written by UK postgraduate university students majoring in Computer Science, including some Arab and other EFLLs of other nationalities. Hitherto, though, there have been no studies that have used statistical tests to measure collocation frequencies between Arab EFLLs and NBESs. The present study adds to and expands on these previous studies and also tries to identify the relationship between the collocations used and the influence of Arabic as an L1. According to Basal (2017, p.342), there needs to be more literature addressing the best ways to teach collocations. It is hoped that this study will make significant contributions to pedagogy in that area, and to both linguistics and translation studies in a number of domains, as outlined below:

(i) This study contributes to the existing literature because of its original application of statistical methods in a comparative investigation of lexical collocation use by Arab EFLLs and NBESs. To my knowledge, hypothesis

testing through frequency analysis has not been performed by other scholars on Arab EFLLs' collocation use. Such an approach is a gap in the study of Arab learners' use of collocation types.

(ii) To the best of my knowledge, none of the previous studies has proposed utilising collocation word lists or dictionaries to improve the instruction of lexical collocation in English writing for Arab EFLLs. Moreover, no other research has been done to extend the CEFR lists to include collocations (the CEFR Modification for Collocations, hereafter referred to as the CEFR-Coll, that is one of the outcomes of this study (§4.2.2.1)). Such a list will provide detailed information about learners' language proficiency when using collocations. The CEFR-Coll is still under development, and can be further improved in follow-up research.

(iii) The vast majority of previous work in this area has used limited or pre-selected sets of lexical collocations. This study, however, is entirely data-driven, and thus covers a wider and more comprehensive range of the possible English lexical collocations used by Arab foreign language learners, as all the realisations of the four collocation types studied are extracted from the corpora used. Thus, this study contributes to the existing literature because of the wider scope and bigger database, and studies collocation and in much greater depth then the somewhat piecemeal efforts of the past.

Moreover, this study contributes to our understanding of the importance of statistics in data analysis. The use of statistical tests has been mostly associated with parametric tests. However, this study contributes to our understanding that parametric tests are not suitable to the field of Applied Linguistics. Therefore, the use of robust non-parametric tests are the most suitable types of tests when considering linguistic data and this can be justified by several reasons. First, linguistic data usually do not follow a normal distribution because linguistic data can take the form of categorical or ordinal data which do not assume continuous data. In the case of this study, the collocation data are ordinal and categorized by an association measure. Second, sample size is another factor that plays a role in favouring non-parametric tests for linguistic data. Non-parametric tests

would be more appropriate for a small sample size where the assumptions of parametric tests may not hold. Having non-parametric tests for linguistic data is more robust against violations of assumptions. Performing non-parametric tests requires a thorough investigation of the data type to conduct the right statistical significance tests.

1.6 Research Questions and Research Hypothesis

Having introduced the four lexical collocation types selected for study above (p.1), in this section (§1.6), I present the research questions and highlight the hypothesis formulated for the statistical significance testing that guided this doctoral study. The four selected lexical collocations are given below:

(i) **Adverb-Adjective collocations** (e.g. *closely related*)
(ii) **Verb-Noun collocations** (e.g. *pay attention*) including the following:
- **Verb-(intermediate)-Noun collocations** (e.g. *book an appointment*)
- **(to) + Verb-Noun collocations** (e.g. *to pay attention*)
- **(to) + Verb-(intermediate)-Noun collocations** (e.g. *to book an appointment*)

(iii) **Verb-Adverb collocations** (e.g. *appreciate sincerely*)
(iv) **Adjective-Noun collocations** (e.g. *best friends*)

The following are the research questions that will be investigated in this thesis:

The first research question is:

- ***RQ1:*** *Can hypothesis testing of relevant corpus data help us to identify the significant differences between Arab EFLLs and NBESs in their use of the four lexical collocation types chosen for this study?*

The second research question is:

- ***RQ2:*** *Using the CEFR classification of words into typical user proficiency levels, how can analysing word choices within language learners' collocations assist in understanding their language proficiency?*

The third research question is:

- ***RQ3:*** *What parts of the lexical collocation errors made by Arab EFLLs can be explained by the influence of Arabic as an L1 (interlingual errors)?*

As this study follows an experimental design, it is essential to formulate some hypotheses based on the data collected from the available corpora. The hypotheses are essentially predictions or untested statements about the data, and the testing of hypotheses, by inferential statistics, is a method that helps in formulating predictions about additional data beyond those collected according to the experimental design (Scott, 2014; Gries, 2017, p.316).

The research hypothesis considered throughout this study relate to the above research questions:

- **The null hypothesis ($H1_0$):** There are no statistically significant differences in frequency between the lexical collocations used by Arab EFLLs and NBES.

- **Alternative to the null hypothesis ($H1_A$):** There is a statistically significant difference between the lexical collocation use of Arab EFLLs and NBESs.

1.7 Ethical Approval

In this PhD thesis, I used three corpora of language learners' English, all available for public use. I also conducted surveys where I recruited two native English speakers. There was no need for ethical approval because my study did not involve any personal information that would generate ethical concerns. This was stated by the University of Leeds' ethical team after I consulted them regarding the data types used in this study.

1.8 Thesis Structure

This thesis has six chapters, including this one. In this chapter, I have provided an introductory discussion of this thesis' aims, objectives, research questions and related hypothesis.

In Chapter 2, I explain the possible definitions of collocations and approaches to identifying collocations through a discussion of some relevant concepts. I also present the theoretical framework of this thesis. In addition, I provide a review of previous studies of collocation use.

In Chapter 3, I describe the methods used when investigating the research questions listed above (§1.6). Information about the corpora used and the sample information are also given. I demonstrate an investigation into the corpus size effect between the two data sets for Arab EFLLs and NBESs.

In Chapter 4, prior to reporting the results of the main comparative study between Arab EFLLs and NBESs, I also present a feasibility study carried out to evaluate the research design and methodology, identify potential problems, refine research questions and hypotheses, and determine the comparability of the AE and BE data. I then discuss the findings and results of the statistical hypothesis testing and the contrastive error analysis for the four lexical collocation types in the main study.

In Chapter 5, I present and discuss the answers to the research questions in the light of the previous literature.

In Chapter 6, I summarise the conclusions and contributions of this PhD research, discuss the limitations and provide suggestions for future research.

Chapter 2 An Overview of the Literature on Lexical Collocation

Introduction

In this chapter, I will first define and explain some of the concepts and terms used in the study, and then set up the theoretical framework. I will focus, in particular, on the existing conversations around the written lexical collocations produced by Arab EFLLs, and how they compare with those of NBESs (§2.2).

2.1 Definition of Concepts and Terms.

This section reviews some concepts used in this study; to avoid potential confusion consistent terminology should be adopted. The following sub-sections briefly explain and put into context the terms referred to throughout the study, such as *corpus linguistics and the use of corpora*, *the benefits of corpus linguistics for EFL*, *learner corpora*, *reference/general corpora, raw* and *normalised frequency*, *token* and *type,* and other relevant terms.

2.1.1 'Corpus Linguistics' and the use of 'Corpora'

The term *corpus linguistics* refers to an approach to linguistics that aims to investigate linguistic aspects through the analysis of compiled textual data using a corpus. Leech views Corpus Linguistics as a methodology "that defines not just a newly emerging methodology for studying language, but a new research enterprise, and in fact a new philosophical approach to the subject" (1992b, p.106). Taylor (2008, p.180) argues that the term corpus linguistics can be explained in many possible ways including viewing it as a tool, method, methodological or theoretical approach, discipline, or a combination of all these listed. Many authors define Corpus Linguistics as a methodological approach to studying language use (McEnery and Wilson, 1996; Meyer, 2002). The purpose

of compiling linguistic data is to study a language intensively to identify language patterns in all kinds of written/spoken language. Corpus linguistics is generally perceived as a research method for language analysis (McEnery et al. 2006, p.7). Thus, though the term corpus linguistics refers to a field of linguistics that analyses a language by computational means, it is also referred to as a research methodology that linguists apply to language analysis. It is nowadays used in the service of various branches of linguistics when seeking evidence of language use (Leech, 1992a; Matthews, 2014).

McEnery and Hardie observe that, in its development, Corpus Linguistics "has spawned, or at least facilitated the exploration of, new theories of language – theories which draw their inspiration from attested language use and the findings drawn from it" (2011, p.1). According to Hunston (2002), corpus linguistics can allow for a better understanding of linguistic phenomena such as collocations. Moreover, corpus linguistics has a great correlation and compatibility with the field of second language acquisition, as linguists have started to address more issues in the field of linguistics and language learning. Brezina defines corpus linguistic as a 'scientific tool' used for language analysis to 'provide empirical evidence' from the corpora used (2018a, p.2). I agree with McEnery and Hardie (2011) and Brezina (2018) who view corpus linguistics as a quantitative methodology because corpus linguistics is usually the study of how linguistic patterns can be understood through numbers (i.e., frequencies and association measures).

In linguistics, a corpus can be defined as a collection of naturally occurring written texts or recordings of spoken language that is machine-readable and can be accessed and analysed through specialist software packages for the purpose of linguistic analysis (Leech, 1992b; Adolphs, 2006). Many current definitions emphasise the fact that corpora are accessed on a computer, but a corpus is just a collection (a body) of texts. The exact definition of the word corpus varies 'in terms of its form and its purpose' (Hunston, 2002, p.2). There are many types of corpora, such as specialised corpora, general/reference corpora, comparable corpora, parallel corpora and finally, learner corpora, such as the ones this study is based on. As the name suggests, a learner corpus is a corpus that contains samples of learners' language production: that is, either spoken or written data,

for the purpose of investigating linguistic phenomena. The two types of corpora used in this thesis, learner and general/reference corpora have been explained below (§2.1.3 and 2.1.4).

2.1.2 The benefits of Corpus Linguistics for EFL

Regarding the use of a corpus and a concordance tool in the improvement of existing teaching methods, and the crafting of new pedagogical methods for linguistic acquisition, one study, by Yoon & Hirvela (2004), aimed to understand learners' attitudes to using a corpus in their second language writing. The participants were either intermediate or advanced level ESL learners at a large university in America. The authors used quantitative and qualitative methods to gain deeper insights into how the learners viewed using a corpus in their second language writing and found that he use of a computer-based program, such as a corpus tool, was rather beneficial in their learning process, improving their development in second language writing and confidence in using their second language. By consistently utilizing corpora in their course assignments, the learners gradually became more accustomed to using them, resulting in enhanced writing skills and fewer instances of grammar mistakes being pointed out by their instructor. Consequently, these learners ended up advocating for the inclusion of a corpus component in all ESL writing courses.

In the same vein, Charles (2014) examined EAP learners' habits in using a personal corpus over a period of 12 months. She found that the students tended to refer to the corpus to check their grammar usage and lexis; also, 93% of the learners perceived an improvement in their academic writing due to corpus use. Charles aimed to understand the participants' reasons behind their use and non-use of the corpus over the duration of the study. She found that the learners kept on referring back to the corpus as it helped them to correctly choose the right way to use collocations in English sentences. The participants faced difficulties when using the corpus at the beginning but then started to 'have fun' with it. One participant stated that he relied more on what the concordance showed, and he clearly stated that if he had only three hits for a certain search, the answer would

not be ‘Yes, it is right to use it’. However, a second participant, a non-user of the corpus, was excited at the beginning but later said that he had no time, and that the use of a corpus was rather too much work. Despite the negative assessment of this second informant, this case study supports the findings found in Yoon and Hirvela (2004), in which corpus use in language teaching was found to have a generally positive impact on improving learners’ second language writing.

2.1.3 Learner corpora

A learner corpus is a collection of written texts or spoken recordings produced by learners of a particular language (Granger et al., 2002; Hunston, 2002). This type of corpus is used in identifying learner errors according to a number of parameters such as language proficiency level, L1, and how the errors differ from the usage patterns of native users. There is a growing interest by researchers in second/foreign language acquisition, language teaching, and language assessment in the use of learner corpora for pedagogical purposes (Flowerdew, 2014). A well-known learner corpus is the International Corpus of Learner English (ICLE[1]). The learner corpora used in the examination of lexical collocations produced by Arab EFLLs were the Arab Learner English Corpus (ALEC[2]) and the BUiD Learner Corpus (BALC[3]), and, for the NBESs, the British Academic Written English Corpus (BAWE[4]) (§these three corpora are discussed in the following section). Although material for the BAWE corpus was collected in an English-speaking University, it must be noted that it is not entirly L1 production.

---

[1] For more information about the ICLE corpus: https://uclouvain.be/en/research-institutes/ilc/cecl/icle.html

[2] For more information about the ALEC corpus: https://dspace.auk.edu.kw/handle/11675/1757

[3] For more information about the BALC corpus: https://ucrel.lancs.ac.uk/publications/cl2009/54_FullPaper.doc

[4] For more information about the BAWE corpus:
Nesi, Hilary; Gardner, Sheena; Thompson, Paul; et al., 2008, British Academic Written English Corpus, Oxford Text Archive, http://hdl.handle.net/20.500.12024/2539.

2.1.4 Reference/general corpora

A reference/general corpus is a large corpus containing different types of texts (e.g. journalistic, literary, or technical) representing a particular language. Hunston (2002, pp.22-23) explains that the terms 'general' and 'reference' are used to describe this kind of corpus because it is large in size and shows the general uses of a particular language, making it suitable as reference material for pedagogical purposes. Leech explains a reference corpus as one that "provides comprehensive information about the language" (2002, p.1). The reference corpora used in this study are the BNC, BE06 and INTERNET-EN[5].

BNC stands for the British National Corpus, which contains 100 million words of spoken and written British English. Trinant and Kijpoonphol (2021) and Hunston (2002) illustrate that the BNC is used as a reference corpus for comparative purposes to check native British English-speaker collocations. One reason for referring to the BNC is that it represents the same regional variety of English as the BAWE corpus, which was collected from native British English-speaking students. This therefore allows a fairer comparison. In addition, the BNC represents a standard formal type of English, which can be beneficial for this study for several reasons. First, the BNC contains data collected from various registers and genres (Brezina and McEnery, 2020, p.16). Another reason for choosing the BNC is that it is one of the largest balanced English reference corpora, as the big Internet corpora are not balanced, and is considered to be representative of general English comprising a variety of genres (i.e., the BNC features academic writing, newspapers, Wikipedia[6] articles and fiction) (Lee, 2001, p.50). Therefore, the BNC will be particularly useful when comparing the Arab EFLLs' collocations to those of the NBESs because the register in BAWE is limited to academic English. Leech (1992a) argues that the use of the BNC corpus will assist in the study of any uses of the English language and its linguistic aspects.

---

[5] These three corpora were accessed via Leeds IntelliText: http://corpus.leeds.ac.uk/itweb/htdocs/Query.html

[6] Note that the BNC predates Wikipedia. However, it does include a fair amount of reference information texts, such as factsheets.

The second reference corpus is the BE06 corpus. This is a written British English reference corpus comprising one million words (Baker, 2009).The corpus was collected over a two-year period (i.e., from 2005 to 2007), and the median point of this period, 2006, was taken to name the corpus as BE06. The BE06 corpus was modelled on the Brown corpus[7] and is not only the same size but also contains similar types of written text. Because the BE06 is also relatively similar in size to the Arab EFLLs corpus it was used in some parts of the current study, especially when investigating the corpus size effect when comparing the corpora of AEFLL and NBES essays (§3.6).

The third reference corpus is the INTERNET-EN, henceforth referred to as I-EN, which is a corpus of 150 million words of texts collected from online sources. In this study, the I-EN corpus is used as a reference corpus in the error analysis of the Arab EFLLs' use of lexical collocations. Sharoff (2006) has claimed that Internet corpora have the potential to be even more valuable than representative corpora that were manually generated because Internet corpora can provide a sample of contemporary language as opposed to traditionally constructed corpora such as the BNC from 2000 or the BE06 from 2006. It must be noted that Internet corpora are not as well-balanced as those compiled more traditionally, such as the BNC (Sharoff, 2006, p.436), and there is no guarantee that all the texts are contributed by native speakers. Although web corpora are large and contemporary, Rainer claims that there could be an over-representation of some text types and not much representation of others (2003, p.138).

2.1.5 Raw frequency (RF) and normalised frequency (NF)

The term *frequency* refers to the number of instances of a particular item occurring within a data set. Two frequencies will be reported: raw frequency and relative/normalised frequency. For example, in the BNC, the word *take* has a raw frequency of 173,646 and a normalised frequency of 1,545.64. *Raw frequency* is

[7] The Brown Corpus is the first collection of written texts in American English, compiled and annotated for linguistic research. For more information about the Brown corpus: https://www.sketchengine.eu/brown-corpus/

commonly referred to as an absolute frequency, representing the actual number of occurrences of an instance, commonly reported when considering a single corpus. Raw frequency is simply the number of times a word/token is observed in a corpus.

The second type, *relative/normalised frequency,* is usually reported when comparing two or more corpora of different sizes. Normalised frequency is usually presented as the number of instances per thousand words for smaller corpora, or per million words (*ipm*) for larger corpora. In this research, normalised frequency means frequency per million words; normalisation is used to allow for a fair comparison of corpora that are different in size.

2.1.6 Token and Type

Token and *type* are two terms that must be clarified when discussing word frequencies in a corpus. Both terms represent some aspect about the words in a corpus (Hunston, 2002; McEnery and Hardie, 2012). The total number of tokens in a corpus refers to the total number of words while the total number of types refers to the number of distinct words, ignoring the repetitions of each type. Brezina (2018a, p.39) gave the following example: '*took them 26 years to win the title. During that time we won it 3 times despite being in the second division for half that time'.* has 26 tokens but only 23 types because the words 'the', 'that' and 'time' said twice, therefore each is considered as one type. (§Table 3-2 for detailed information about the number of tokens and types for the learner corpora used in this thesis).

2.1.7 Concordances

A *concordance* is a list of occurrences of a word searched for in a text. A set of concordance lines is the collection of occurrences which is the outcome of searching for an item in a compiled corpus (see Figure 2-1 for a concordance line view in a corpus). Sinclair explains a concordance as simply being "an index" in which each word is indexed and referenced to indicate its place of occurrence in a text (1991, p.32). Concordance lines are useful in showing how words behave

in context and in establishing the precise meanings of a node word (Millon, 2011, p.131). More features of concordance lines will be explained in Chapter 3 which describes how concordance lines were used in the analysis.

2.1.8 Node word and collocate

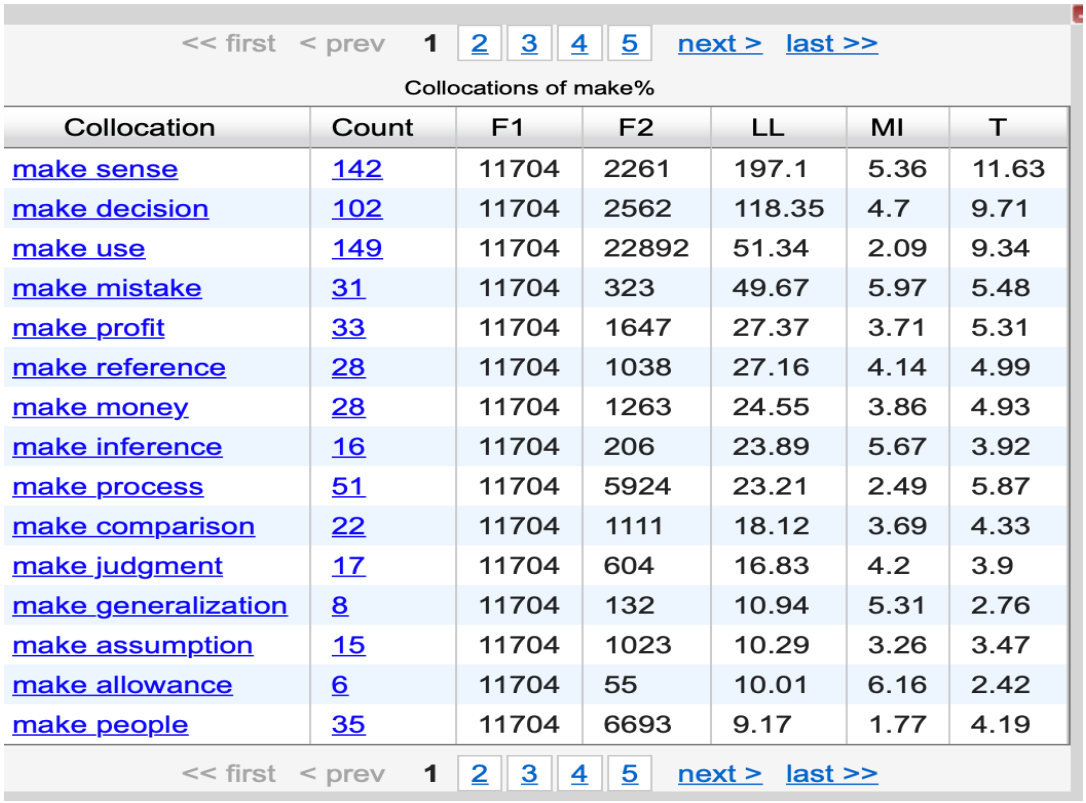
A *node word* is the word searched for in a corpus, which in my study, is the first part of a collocation. A *collocate* is the word that co-occurs with a node word. In other words, a node word can be explained as the specified lexical item searched for within a corpus data-set, and its collocates are the various words that appear in the context of the node word (Sinclair, 1991; Pirhayati, 2021). For example, in a Verb-Noun collocation, the node word would be a verb drawn from a word list of selected verbs. The collocates for these verbs are all the possible nouns that can collocate with that verb (§I explain this in detail Chapter 3).

Concordances for make

leid	left	match	right
001	political parties and the links they	make	between non-white immigrants and
001	should have children as it will	make	them happy, and see it as natural
001	). However, although this does	make	the young more vulnerable, the
001	. These developments therefore	make	men more vulnerable to unemployment
002	society, paving the way for women to	make	their own voices heard in the

Figure 2-1: Screenshot of concordance lines for the node word *make*

Figure 2-1 shows a screenshot from IntelliText of how *make* appears as the node word in concordance lines (§Figure 2-1 for more information about Leeds IntelliText). The node word is the central word appearing in the concordance lines showing the searched-for item within a corpus data-set, though the position of the node can vary based on the researchers' choice of search output.



<< first < prev 1 2 3 4 5 next > last >>

Collocations of make%

Collocation	Count	F1	F2	LL	MI	T
make sense	142	11704	2261	197.1	5.36	11.63
make decision	102	11704	2562	118.35	4.7	9.71
make use	149	11704	22892	51.34	2.09	9.34
make mistake	31	11704	323	49.67	5.97	5.48
make profit	33	11704	1647	27.37	3.71	5.31
make reference	28	11704	1038	27.16	4.14	4.99
make money	28	11704	1263	24.55	3.86	4.93
make inference	16	11704	206	23.89	5.67	3.92
make process	51	11704	5924	23.21	2.49	5.87
make comparison	22	11704	1111	18.12	3.69	4.33
make judgment	17	11704	604	16.83	4.2	3.9
make generalization	8	11704	132	10.94	5.31	2.76
make assumption	15	11704	1023	10.29	3.26	3.47
make allowance	6	11704	55	10.01	6.16	2.42
make people	35	11704	6693	9.17	1.77	4.19

<< first < prev 1 2 3 4 5 next > last >>

Figure 2-2: A screenshot of the possible noun collocates for the verb *make*

Figure 2-2 shows a screenshot of the IntelliText web interface for the collocation search output. The screenshot shows the possible noun collocates for the node *make*.

2.1.9 Lemma and word-form

A *lexeme* is an abstract word group containing all the forms of a word that are related by inflection (Crystal and Alan, 2023, p.266). For example, *adapt, adapts, adapting, adapted* are all realizations of a single, abstract verbal lexeme**,** which, because it is abstract and has no visual form, is conventionally represented by the *lemma* ADAPT**.** The *lemma* is therefore simply a conventionally chosen word form assigned to represent a lexeme and all its different forms, and is equivalent to a dictionary headword, which is the word used in dictionaries for introducing an entry (Knowles and Don, 2004, p.69; Crystal and Alan, 2023, p.265). A lemma is a shorthand used to refer to all the possible inflections of a lexeme for a particular part of speech. For example, PLAY is the verbal lemma that covers all

the inflected forms of the verb, such as *play*, *plays*, *played*, and *playing* (Hunston, 2002; Matthews, 2014). The lemma function in a concordance can be viewed as a query function for searching a group of inflectionally related words. For example, searching the verbal lemma PLAY will bring up all its possible forms: *play*, *plays*, *played*, and *playing*. This lemma function was used for collocation extraction in the current study in order to quickly obtain all the related examples. Then, I would examine each word form separately (e.g. *play, plays, played, playing)* because each can have its own unique collocates. These two steps were taken to increase the number of the extracted collocations from the Arab EFLLs corpus. Therefore, a collocation node word PLAY may appear in its base form in my collocation list (e.g. *play football*), but the actual collocation patterns could include any of the possible forms of the verb that derive from the base form *play*, such as *play, plays, playing, played football.*

2.1.10 First language, target language and foreign language

The concept of a 'first' language (L1), as explained by Hammarberg (2010, p.97), refers to the first language acquired by a speaker of a language "during infancy" (i.e., their native language). In this thesis, the first language for the Arab EFLLs Corpus is Arabic; thus, L1 for them refers to Arabic. The target language refers to the language being learnt (Sakaria and Priyana, 2018, p.175), which, in this thesis, is the English language. A foreign language is a language that is not officially spoken or used in a specific country (Wolff, 2009, p.12). For instance, English is a foreign language in the Arabian Gulf countries. Therefore, the target language for the Arab learners in the two corpora is the foreign language English, which explains why the joint corpus is entitled the Arab English as a Foreign Language Learner (Arab EFLLs) corpus.

2.1.11 Defining the construct of 'collocations.'

Several definitions of collocations are proposed in the literature (Mel'čuk, 1998; Manning et al., 1999). John Rupert Firth (1957) was the first to pursue the term *collocation* after it was introduced by Palmer (1933). Since then, Firth has been

recognised as the father of collocation studies (Williams and Millon, 2011; Talakoob and Koosha, 2017). Firth defined collocations as ‘words in habitual company’, stating that ‘you shall know a word by the company it keeps’ (1957, p.183). Firth’s definition explains that collocations are words that co-occur habitually, thus creating a particular meaning in language production.

The frequency-based approach relies on statistics to define collocations in terms of how frequently they occur in a specified context. This approach is also known as the Firthian approach. Most descriptions of frequency-based approaches go back to Firth’s definition, but were further developed later on (e.g. Sinclair, 2003, and Halliday, 1966). Frequencies, in the simplest terms, can be explained as numbers that numbers that allow the calculation of the probabilities of co-occurrences. Another aspect that must be considered is relative frequency, which is the extent to which, given the frequency of the first word, the second is conditional on the first.

Collocation studies that are concerned with frequency or those that deal with statistics in their approach mainly consider the span that exists between the words that make up the collocation (Walker, 2011). This span is usually manifested as the number of words to the right and left of the ‘node’. Nesselhauf stated that a collocation is regarded as a co-occurrence when the combination is “more frequent than could be expected if words combined randomly in a language” (2005, pp.11-12). To further discuss this frequency-based approach, it is worth mentioning Sinclair, who is mainly associated with this type of work, who viewed collocations in terms of probability, and the strength of a particular collocation was determined by how frequently it appeared in a large body of representative data. Moreover, according to Halliday, native speakers of a language will react to a collocation as “being more or less acceptable on a scale of acceptability using it at a normal frequency” (1966, p.159). More generally, most of the authors in the literature agree that collocation refers to the clustering of two or more words from different parts of speech in a lexical unit (Baker, 1992; Hunston, 2002; Williams, 2002).

The term *collocation* has also been approached from a more phraseological point of view. Nesselhauf follows Firth's view in deeming that collocation is a concept at the 'syntagmatic level' (1957, p.196), and calls it the "syntagmatic relations of words" (Nesselhauf, 2005, p.11). This word *syntagmatic* refers to the relationship that occurs between at least two linguistic units arranged sequentially to create a well-formed structure; e.g. there is a syntagmatic relation between the words *night* and *dark* as there is a habitual association creating the collocation *dark night*. The term collocation is generally understood to mean a group of words which usually goes together to create a specific comprehensible meaning. According to the compilers of the *BBI Combinatory Dictionary of English*, "In English as in any other language, there are many fixed, identifiable, non-idiomatic phrases and constructions. Such groups of words are called recurrent combinations, fixed combinations or collocations." (Benson et al., 1997, p.IX).

The phraselogical approach is mainly associated with the work of Cowie (1998). To identify collocations from a phraseological perspective, Cowie and Wood (2019) relies on native speaker intuition or Laufer and Waldman (2011) suggested to use examples drawn from collocation dictionaries. Hunston stated that the phraseological approach emphasizes the importance of analysing words not only "in isolation" but also in the context of the larger phrases or chunks in which they occur (2010, p.14). Collocation in phraseology is not random pairing of words (Evert, 2005, p.138), yet it is rather systematic and meaningful linguistic unit. For example, 'break the ice' is phrase used to do something to relieve tension used in conversational activities.

A frequency-based approach is more related to statistics, while a phraseological one concerns the type of words that usually go together. Several terms have been used in the literature to explain this contrast, such as quantitative vs qualitative (Bartsch, 2004, p.80), and empirical vs theoretical, the terms used by Evert (2008, p.3) respectively for the two approaches.

Both the frequency-based approach and the phraseological approach have been used in previous studies about collocation use for EFLLs (Nesselhauf, 2005; Farooqui, 2016; Deveci, 2018; Alangari, 2019). Though Farooqui's (2016) and Alangari (2019) study might sound similar to my proposed research, there are many differences. First, Farooqui's data were dissertations selected only from highly proficient postgraduate students studying in a UK university. Therefore, this mean that they are more proficient English users. Second, the students may also have used proof-readers before the actual submission (Mumin, 2022, p.52). In comparison, the data chosen for my research project were essays written by Arab EFL students with varying levels of proficiency. Third, the scope of this study includes a wider range of collocation patterns, while Farooqui focused on Noun-preposition-Noun and Noun-adjective-Noun, and Alangari focused on the use of Verb-Noun collocations.

Though the development of corpora has provided a means of studying previously unnoticed issues in a language in ways the human eye cannot easily perceive (Kennedy, 2014, p.208), human judgment is still required in some cases, which is considered a drawback that complicates the statistical processes relating to corpora (Hunston, 2002, p.43). Many researchers view that judging the acceptability of collocations in corpus-based studies is difficult because the judgment is typically based on the intuitive phraseological knowledge of the researcher (Williams and Millon, 2011, p.28). For example, there might be a high frequency of certain phenomena, such as the verb being followed by an infinitive or *-ing* form (Scott et al., 1974). The choice between an infinitive or an *-ing* participle following a verb poses some difficulty for Arab learners of English as they tend to confuse them in the initial stages of their learning and might formulate sentences such '*I like <u>to going</u> to the gym'. However, using corpora gives us a complement to intuition as certain language phenomena can now be tested on a collection of authentic data (McEnery et al., 2006, p.6).

Before going into further discussion about collocation, it is fundamental to address the existing terms for collocations available in the literature. One term is ‘lexical bundles’, which is used to refer to multi-word expressions, usually a group of three or more words that normally go together (Biber et al., 1999; Biber et al., 2004). Other terms used to express collocations are ‘prefabricated chunks’ (Orenha-Ottaiano, 2012, p.242) and ‘semantic sequences’ (Hunston, 2008, p.8).

From a statistical perspective, collocations refer to the co-occurrence of two words within a short distance: for example, *appreciate sincerely,* which is a *Verb-Adverb* collocation, and *have lunch,* is a *Verb-Noun* collocation. Hunston proposes a further statistical view of collocations: “the statistical tendency of words to co-occur” (2002, p.12), stating, that collocation can be explained as the tendency of attraction of one word to another or ‘the tendency of words to be biased in the way they co-occur’ (2002, p.68). This suggests that there is a logical explanation that motivates words to co-occur. From a statistical perspective, collocation demonstrates the tendency of certain words to occur close to, or to associate themselves closely with, certain other words. For this study, collocation is quantitively and more narrowly defined as:

> *Two elements drawn from the English content-word classes, A and B, occurring in close proximity to each other in a corpus, where B is conditional on A.*

This definition highlights that the second part will be selected based on the particular selection criteria for the extraction of collocations; so, for example, for Adverb-Adjective collocations, the *adjective* collocates are conditioned on the previously chosen adverbs. A collocation has many other alternative definitions with different purposes, such as grammatical collocation or lexical collocation. The former refers to the clustering of different content word classes with a grammatical or function word such as a preposition: for example, ‘at night’ or ‘in advance’ or involving an infinitive such as ‘to be afraid of’. The second type, lexical collocation, which this study is based on, does not include any prepositions or particles. While this study does not investigate grammatical collocations, it is worth explaining the difference between grammatical and lexical collocations, as

the latter is the focus of this research. Although grammatical collocations are an important aspect of language learning, I had to narrow down my research focus in order to conduct a more in-depth analysis of lexical collocations. Therefore, I did not include grammatical collocations in my research. Abedi and Mobaraki define grammatical collocations as those that include either "a noun, or an adjective or a verb, plus a particle (a preposition, an adverb or a grammatical structure such as an infinitive, a gerund, or clause)" (2014, p.632). These types of collocations are reliant on the syntax of a language and are frequently employed to convey precise meanings during communication. An example of a grammatical collocation is 'afraid of' which comprises the essential term 'afraid' and the preposition 'of'. If used separately, these words would convey a different meaning than when used together in this specific collocation. Lexical collocation includes two or more-word combinations from different parts of speech: for example, *book an appointment* is a Verb-article-Noun lexical collocation; I must clarify here that this collocation is treated as a Verb-Noun collocation in this study. *Lexical collocation* is the term used in this study and the term *word combination* is also used interchangeably based on its transparent, literal sense.

2.1.11.1 Collocations from the perspective of Semantics and the concept of 'naturalness' in collocations

Semantics is the field of linguistics that studies meaning. In his 1981 book on Semantics, Geoffrey Leech introduced seven types of meaning: *conceptual meaning*, *connotative meaning*, *social meaning*, *affective meaning*, *reflected meaning*, *collocative meaning*, and *thematic meaning*. The most relevant kind of meaning for this research is *collocative meaning*. Leech stated that collocative meaning "consists of the association a word acquires on account of the meanings of words which tend to occur in its environment" (1981, 17). Collocative meaning is clearly relevant to collocation formation and entails a relationship at the lexical level of language. For example, Leech explains that using the adjectives *handsome* and *pretty* requires a kind of association with the nouns that would co-occur with these adjectives. Thus, there are some instances of 'handsome woman' as well as 'pretty woman', as both examples are correct. Yet, 'pretty woman' sounds more natural at the lexical level based on its collocative meaning.

He also gives the example of how ‘to wander’ and ‘to stroll’ collocate differently, in that “cows may wander, but may not stroll” (1981, p.19).

Collocative meaning cannot always explain the differences in some co-occurrences because collocative meaning is mainly associated with the ‘idiosyncratic property of individual words’ (1981, p.17). The association between the two words in the collocations will be examined to determine whether they are appropriate to be used next to each other. For example, the noun ‘crime’ usually collocates with the verb ‘commit’ more than ‘do’ or ‘make’.

In this thesis, I use the term naturalness to explain collocation use by Arab EFLLs. The term naturalness explains whether the collocation is accurate and natural from a native English speaker’s perspective. Collocation use is better explained as natural or unnatural because there is no concept of true or false for collocation in linguistics. In my study, as I explained earlier, collocations are not pre-set phrases or idioms. Therefore, the better way to explain the Arab EFLLs’ use of collocation is from a natural perspective. Naturalness is used in the literature in both error analysis and translation studies (Obeidat and Sepora, 2019).

Native speaker intuition judgment is important in the definition of a collocation because collocations are established based on natural combinations of words in a language. Native speaker intuition judgments are a valuable reference point for identifying and defining collocations. Through exposure and familiarity with the language, native speakers have an innate sense of how and which words may sound natural and appropriate (Bley-Vroman and Yoshinaga, 1992, p.192). However, native speakers may have variation in their judgment of word combinations. This can be influenced by several factors such as variation between national varieties of English (e.g. British vs. American English), individual language use, or changes of language over time. Regardless of these variations, linguistic research often seeks to capture a consensus or prevailing usage among native speakers rather than relying solely on individual judgments.

The term “collocation” is tackled from different approaches in this thesis. First, a collocation, quantitatively can be defined as the combination of two words in close proximity to each other. The main essence of these collocations through the quantitative approach is that the collocation is defined and represented as numerical data (i.e., through a number either about the frequency of co-occurrence or by an association measure). Second, a collocation from the phraseological approach can be explained as how units beyond individual words, such as multi-word expressions, are used and understood (Greaves and Warren, 2010). The clustering of the word combinations is “not necessarily meaningfully associated”; therefore, the quantitative approach, phraseology, and intuition are linked in the analysis as each approach complement the other (Cheng et al., 2009, p.241). Intuition is needed in all of the three in defining collocations because the quantitative approach can help in extracting collocations use but when refining the outcome, native intuition is needed to interpret and assess the outcome.

For example, “bring a boy” and “give birth to a boy” are grammatically correct. However, the Arab EFLLs would use “bring a boy” when they literally mean a lady who gave birth to a new-born. Both meanings are correct and natural in the three approaches. However, when providing the context, judges might change their view because the natural form would be the second. Therefore, native speaker intuition is vital for understanding and defining collocations, as it reflects the intuitive sense of how words typically co-occur in a language. The terms “phraseological,” “qualitative,” and “intuition” are interconnected in this thesis about collocations as they collectively contribute to the study of language patterns and usage.

2.1.11.2 How ‘collocation’ is understood in this thesis

The collocation selection criterion in this thesis is based on syntactic relations (e.g. verb-object), such as *make a decision*. Only the lexical collocations will be studied, so any collocations that include grammatical components such as a preposition or a clause are not considered in my thesis. Although I am not focusing on grammatical collocations as such, I do focus on the underlying

grammar of lexical collocations such as the grammatical agreement in the collocations including their word order, verb tenses and other grammatical issues that might arise from interference from Arabic. And most importantly, from the perspective of this thesis, collocation can be defined as two content words that co-occur together naturally, with/without intermediates. Occurring naturally means that collocation meanings are not governed by its collocability. For example, the collocation '*bring a boy' does not sound like a natural collocation to a native speaker but 'give birth' does. In addition to naturalness, collocations should be recognised as a frequent combination by native speakers. For example, there is nothing unnatural in the collocation in *bring a boy* in 'the girl will bring a boy to the party', but it is not and recognised as a frequent combination by native speakers. It is mainly the clustering of two words without being governed by other rules that limit the creation of collocations. Collocation, from the perspective of my study, can be defined as the clustering of two words where the second word (the collocate) is conditional on the first (base word). For example, if the base was the verb *make,* the collocates would be all the possible nouns that co-occur after that verb in the corpus. Therefore, the second part of the collocation is conditional on the choice of vocabulary for the first part, the base (such as the verb 'make' in the example), and the base word in my study is always a topmost frequent word. My definition of collocations follows the typical lexical collocation categorisation by Benson et al., (1997), Kilgarriff (2006) and Martin (2008), who categorise collocation as structural-type combinations (i.e., including two-word combinations of lexical items, either verb, noun, adverb or adjective). Benson et al.'s view of collocation goes hand in hand with Martin and Kilgariff's definitions as a two-word combination of syntactic types.

The collocation searches in this study were not only based on the four types of word combination listed above; I aimed to consider the different possible sequencing of the lexical collocations. Thus, the possible combinations in English are those shown in Table 2-1. The main impetus behind this goes back to the existing differences between English and Arabic in terms of word order. Also, this study allowed a span of two intermediary words between each of the two parts of a collocation. These words could be any part of speech that could fall within these

pairs of lexical collocations. Some of these are articles (*a/an, the*) or other words that act as modifiers placed either following or after the collocates.

Table 2-1: Possible combinations of parts of speech in lexical collocations in English (Combinations examined for this research are marked by ♦)

Head word part of speech	Verb	Noun	Adjective	Adverb
Verb	Yes (e.g. *help find*)	Yes ♦ (e.g. *take cover*)	Yes (e.g. *stand tall*)	
Noun	Yes (e.g. *the dog barks*)	Yes (e.g. *head teacher*)		
Adjective		Yes ♦ (e.g. *bad habit*)		
Adverb	Yes ♦ (e.g. *drink greedily*)		Yes ♦ (e.g. *highly recommended*)	Yes (e.g. *quite often*)

The possible combinations of parts of speech in lexical collocations in English given in Table 2-1 are in total nine. The possible collocations that are not investigated in this thesis are Adverb-Adverb collocation (e.g. 'quite often'), Noun-Noun collocation (e.g. 'head teacher'), Verb-Verb collocations (e.g. 'help find'*),* Noun-Verb collocations (e.g. 'the dog barks'), and Verb-Adjective (e.g. 'stand tall'). The main reason for not including these collocations is because there a limited output for them in the Arab corpora. Also, the collocations did not meet the assigned extraction criteria for this thesis. The other four were included and examined in this thesis (marked by the diamond symbol ♦ in the table) (§see 2.3 on the four collocation).

In this thesis, I use the term *combination* interchangeably to refer to collocations because I investigate two different combinations for the same collocation type. For example, the Verb-Noun collocation is regarded as one collocation type, but it includes cases where there is an intermediate word within the collocation (i.e., Verb-(Intermediate)-Noun collocation).

2.1.11.3 Word frequencies and collocation identification criteria in corpus-based studies

There is a relationship between word frequency and language proficiency. The English proficiency level of learners can be identified by how much they rely on a basic set of words that have a high frequency (i.e., beginners tend to use only the simpler, most frequent words of English). Also, word frequency lists are used in language teaching, with different vocabulary lists for each level of learning. With the introduction of corpora has emerged the notion of a word frequency list, with words arranged according to their frequency of occurrence within a given set of data (e.g. spoken language, written language, academic texts, or newspaper articles).

The collocation identification criteria used in previous studies tend to share similar aspects in terms of having a defined set of words for the first part of the collocation (Alanazi, 2022; Alhassan et al., 2022). For example, Alanazi (2022) investigated the Verb-Adverb and Verb-Noun collocations of two verbs; *affect* vs *impact*. Alanazi aimed to investigate their semantic and grammatical preferences through a corpus-based study of the BNC using the top 30 collocates with a minimum LogDice score of 0 (another measure of collocational strength). The logDice score is a measure of strength for collocations extraction. Collocates within the top 30 that did not meet the inclusion criteria, such as function words, were removed and more collocates from further down the list were added until there were 30 lexical examples. Alanazi used LogDice because it provides a standardized measure that can be compared across corpora of different sizes, and is preferred for use with large corpora. Also, he claims that LogDice has the ability to provide accurate scores for large corpora without skewing the results. Therefore, the writer chose to use LogDice as it is a more reliable and efficient measure for identifying strong collocations in large corpora.

Similarly, Alhassan et al., (2022), identified Verb-Noun collocations through raw frequency counts by generating a wordlist for the most frequent verbs in their corpus. The frequency threshold for the node word was set to 34. Then the

frequency counts were ranked to analyse the collocations. Finally, they searched manually for the possible noun collocates.

Astia and Yunianti (2020) investigated the most frequent adjectives used in discourse about COVID-19. Their method for extracting them was to start with the most frequent adjectives in the corpus, using a wordlist from Sketch Engine. Then, these adjectives were investigated through concordance analysis to describe the type of adjectives used during the COVID-19 pandemic by focusing on the context surrounding the adjectives.

Fawi and Delmonte (2015) investigated multi-word terms in an Italian-Arabic parallel corpus. They used the log-likelihood ratio measure as their criterion for extraction from the corpus. Obukadeta (2019) used frequency analysis to extract Verb-Noun and Adjective-Noun collocations. He checked the corpus for the node words of both collocation sets (i.e., the first part of each collocation set which were verbs and adjectives respectively) and assigned a frequency threshold of $f \geq 5$. The statistical analysis was mainly descriptive, analysing the results based on the normalised frequency of the collocations.

Therefore, the criteria for collocation extraction are highly dependent on the type of data, corpus size, and research objectives. Despite these variations, many studies share similar aspects, mainly in the first step of collocations extraction. For instance, frequency analysis may not be the most effective approach for identifying rare or low-frequency collocations, and other measures such as Log-Likelihood or t-test may be more appropriate. Subsequent steps for collocation identification can vary significantly, with different studies using different measures such as LogDice or manually selected collocates. It is essential to evaluate and compare the effectiveness of different methods to ensure reliable and valid results. Most of the studies described above used ordinal frequency analysis (§3.4.1 for more discussion about ordinal statistics) as the initial step in their corpus-based investigation. The wordlist of the top most frequent words of each word class was the main choice for the node words in all their collocation investigations.

2.2 Theoretical Framework

In Section 2.1, I set out the conceptual framework of this research and provided definitions for some of the terms used in this study. In this section (§2.2), I explain the theoretical framework for this PhD thesis. This thesis investigates the concept of L1 interference based on EA theory underpinned by a corpus-based study exploring the use of lexical collocation by Arab EFLLs and NBESs. Error analysis (EA) is an approach to analysing learners' use of language by focusing on the types of repeated error in their language production. In this study, L1 interference was measured through an error analysis of Arab EFLLs' collocation use. This will, I hope, explain how these learners' L1 interferes when using lexical collocations. The main impetus for taking this approach was to clarify the issues that might affect the accuracy of English lexical collocations used by Arab EFLLs. Further discussion of each theory being implemented will be given in the following sub-sections.

2.2.1 L1 Interference theory

In this thesis, *L1 interference theory* can be explained as an approach to explaining the influence of the mother tongue/ first language on production in the second/foreign language. *L1 transfer* and *crosslinguistic influence* are some terms similar to *L1 interference*, which is the grammatical and structural influence of the L1 on the intended meaning in the target language which could cause interlingual errors (Hashim, 2017, pp.60-61). Weinreich (1953) initially defined transfer as a deviation from the norms of either language because of a familiarity with other languages. Some more recent scholars view transfer or interference as the effect of the mother/first language on the L2. Odlin (1989) and Ellis (1997) both view it as influence resulting from the similarities and differences between the two languages. Therefore, based on the literature, the concepts of L1 interference or transfer have something in common, which is the effect of the L1 on L2 language use. Previous studies of L1 interference have shown two outcomes of this process, known as positive and negative transfer, the latter leading to interlingual or intralingual errors (Al-Khresheh, 2010; Khansir, 2012; Hashim, 2017). Positive transfer leads to a well-formed, successful utterance in

the target language. It typically happens because of linguistic similarities between two languages when learners are still dependent on their L1 background (Derakhshan and Karimi, 2015, p.2113). Negative transfer results from learners' attempts to rely on their L1 linguistic background when using a second language, which leads to L1 transfer errors known as interlingual or intralingual errors. For example, Sabah gives the following example of negative transfer: '*Prices have increased a gradual increase' instead of 'Prices have increased gradually' (2015, p.275). Negative transfer is also seen in the use of plural adjectives because of the influence of Arabic as an L1 (e.g. '*These are girls beautifuls'). A transfer is perceived as positive when the impact on the use of the target language is positive; otherwise it is negative (Selinker, 1983; Bardovi-Harlig and Sprouse, 2018, p.2). For instance, the collocation 'two hours' in 'I waited two hours for the bus' is placed incorrectly by Arab learners as '*I waited the bus two hours' as the result of negative transfer from Arabic. Alghamdi example of a positive transfer is in the use of definite article 'the' and 'ال' (2019, p.22). Arab learners will positively transfer the definite article as it exists in their L1 in expressions such as الولدُ (*al-wld*) 'the boy'.

Interlingual errors are those errors attributed to the learner's first language. For example, Mahmoud (2005) found some interlingual errors in the choice of one word within collocations such as **rest a full rest* instead of *take a full rest*. Intralingual errors are those errors that can be attributed to an imperfect knowledge of the language being taught. Intralingual errors are caused by more general learning processes and pathways (e.g. overgeneralisation, incomplete application of language rules or ignorance of language rules). For example, Mahmoud (2005) found some errors (e.g. **took my attention* and **produce a decision*) that he classified as intralingual, because a transfer from Arabic would not cause these errors. There are other terms also used in the literature relating to both interlingual and intralingual errors, such as 'developmental errors' for errors caused by the L2, 'ambiguous errors' for errors caused by the rules of the L2, and 'interference errors' or 'unique errors', which include errors that are cannot be classified under ambiguous or developmental error (Heydari and Bagheri, 2012, p.1854; Richards, 2015, p.4; Tajareh et al., 2015, p.8). In this

thesis, unique errors are considered to be ‘overlap errors’. An overlap error is a type of unique error that is difficult to classify under one specific category. This type of error occurs when a learner produces an utterance that contains features of both the target language and the learner's first language, making it difficult to categorize as an interlingual or intralingual error. This type of error is referred to as an ‘overlap error’ in this thesis to distinguish it from other types of errors and to emphasize its distinct nature (§4.2.3.2 explains the analysis for the encountered errors by Arab EFLLs).

2.2.2 Error analysis (EA) theory

In this section, I explain the second theory relevant to this PhD thesis which is *Error Analysis Theory*. Stephen Corder first introduced the *Theory of Error Analysis* in the 1960s as a reaction to *Contrastive Analysis* (CA), the then dominant approach in the field of *Foreign Language Pedagogy* (FLP). Error analysis aims to provide a sufficient linguistic rationalisation for the types of error in any given learning situation (Pym, 1992). The main difference between contrastive and error analysis is in the source of the difficulties that cause the errors. Contrastive Analysis accounts for the errors in terms of the learners’ L1, while Error Analysis considers the effect of the L2 on the learning process (Heydari and Bagheri, 2012). CA had been criticised for neglecting intralingual factors and considering L1 interference as the only source of learners’ errors. (Dulay and Burt, 1972; Brown, 2000; Heydari and Bagheri, 2012, p.1583; Keshavarz, 2012, p.11).

Error analysis is considered beneficial in two respects; first, it is equally important for pedagogical and theoretical purposes in terms of enhancing the understanding of language use in learning English as a second, (ESL), or as a foreign language, (EFL) (Corder, 1975). Pedagogical understanding can be investigated through qualitative and quantitative means in the process of error classification. Qualitative means providing a linguistic explanation of the errors. According to Lambert and Lambert (2012, p.256), the purpose of qualitative descriptive research is to characterise phenomena directly. Describing errors

from a qualitative perspective allows for the categorical analysis of the error types emerging from investigated data.

In contrast, quantitative means depending on the frequency of errors in determining whether the overall frequency of errors is evidence of absence or existence of difficulty. Quantitative error analysis should be thought of as the second step in applying error analysis theory because initially conducting a qualitative error analysis allows for typifying the errors first, which are then quantified. Although the role of L1 influence has been debated since the decline of Contrastive Analysis in the 1970s, L1 influence remains an important cause of some of the difficulties in the use of collocations (Wang, 2016, p.202; Pinto et al., 2021, p.5).

2.2.2.1 Accuracy and grammatical complexity of collocation

Grammatical rules pose particular difficulty for English Language learners. The difficulty of grammatical rules leads to grammatical errors due to the influence of the learners' L1 (Alsulayyi, 2015, p.34), and learners often lack grammatical understanding and collocational knowledge (Al-Jarf, 2022, p.66).

Foreign language learning researchers are interested in measuring learners' language proficiency, and terms used in the literature to investigate this include 'syntactic complexity' (Kyle and Crossley, 2018, p.333), 'second language complexity' (Bulté and Housen, 2012, p.21), and 'phraseological complexity' (Paquot, 2019, p.124). The use of collocations is one aspect of language complexity. When L2 learners can combine words next to each other or in close proximity to each other and create the same collocational meaning as native speakers, it shows language complexity and proficiency. The meaning of a two-part collocation is governed by both the lexical elements within the collocation (i.e., the node word and the collocate), and there are meaning restrictions that govern the use of collocations that can be used to assess whether the meaning of a collocation is conveyed accurately (Labov, 1973). For example, the

collocation ‘give an example’ is said literally in Arabic as ضرب مثلا (*ḍarb maṯal)* ‘hit an example’. Therefore, it is an achievement when L2 learners can use the right collocations to convey the intended meaning.

The third research question depends on two measures for comparing the lexical collocations use of the Arab EFLLs and NBESs. These two measures are accuracy and lexical/grammatical complexity. The first measure, accuracy, can be defined as the correct use of collocations without errors that deviate from typical linguistic use (Lahuerta, 2018, p.77). For the purpose of this research, Lahuerta (2018) definition of accuracy was selected because I agree with her position that accuracy in collocations implies an absence of any errors in the formation of the collocation. For Lahuerta (2018), accuracy is measured by calculating the total number of error-free collocations used over the total number of extracted collocations. Semantic accuracy is closely related to Izwaini’s (2015) concept of collocational meaning. Therefore, semantic accuracy is a subtype of accuracy that measures whether collocations show an accurate use of the collocation to deliver the intended meaning.

The second measure is lexical/grammatical complexity, which measures errors in word order, wrong use/absence of intermediates and word choice errors within collocations. The first two errors are caused by grammatical differences between the language structures of Arabic and English (e.g. تشرب لانا القهوة (*tašrabu lānā al-qahwati*) ‘Lana drinks coffee’ said in Arabic in the following order as ‘drinks Lana coffee’). In this example, the learner has produced a VSO (Verb-Subject-Object) word order, which matches the word order in Arabic. In this thesis, an investigation of word order issues that might occur within the formation of collocations based on L1 influence will be used as a measure to determine the grammatical complexity of collocation use.

Table 2-2: Semantic accuracy and Lexical complexity measures

Accuracy Form	Measures
Semantic accuracy	**Error-free collocation ratio** Total number of error-free collocation uses/Total number of collocations
	Accurate use of meaning Total number of collocations accurate in meaning/Total number of collocations
Lexical and grammatical complexity	**Error-free word order ratio** Total number of collocations with error-free word order/Total number of collocations
	Collocation Vocabulary level **A score calculated using a newly devised system: *CEFR-Coll***

Collocation vocabulary level is a new measure that I have devised for this study. It calculates a score based on the word choices within the collocations, using a classification of vocabulary that maps onto the Common European Framework of Reference for Languages (CEFR). There are six CEFR[8] levels, A1, A2, B1, B2, C1, C2, which can be grouped into three higher-level pairs (A, B, C) that are given the proficiency labels “Basic User”, “Independent User” and “Proficient User”. IntelliText and Language Profile (https://www.englishprofile.org) were used to assess the word choices in the collocations because these tools automatically tag each word with a CEFR level, and these vocabulary difficulty levels can then be converted into ordinal ranks that can be fed into an equation to derive what I call a CEFR-Coll score. (§refer to 4.2.2.1 for the details on this score).

2.3 A Review of Collocation

In this section, I discuss previous studies about the use of collocations in English and Arabic. This section starts by explaining how terms roughly equivalent to ‘collocation’ are used in Arabic to help us understand the nature of some of the collocational errors in English that can be attributed to speakers of Arabic as an

[8] This website explains what each CEFR level means (https://www.britishcouncil.pt/en/our-levels-and-cefr)

L1. The notion of *collocation* was first referred to by Al-Jaḥiḍ in his book *Al-bayān wa al-ttabayīn* 2nd/3rd century AH (8th/9th century CE). The first mention of the modern term 'collocation' in Arabic was مُصاحبات (*muṣāḥabāt*) introduced by the scholar Abu Al-farj in 1966. Ghazala (2007) used المتلازمات اللفظية (*al-mutlāzmāt al-lafḍhīa)* '*verbal co-occurrences'* to express collocations in Arabic. The Arabic language is rich in collocations, but they have historically been overlooked by Arab linguists (Mustafa, 2010; Igaab and Abdulhasan, 2018). This seems to be because only now, with the development of computer-based corpus linguistics, can linguists tackle this issue easily.

'Collocation'*,* as a term, has many alternative renderings in Arabic including الترادف اللفظي (*al-tarāduf al-lafḍī*) 'lexical sequence'[9], التضام (*al-taḍāmm*) 'inclusion of words together' (Igaab and Abdulhasan, 2018, p.90), المصاحبة اللغوية (*al-muṣāḥabah al-luġawīah*) 'associative relationship [between lexical items]' (Brashi, 2005, p.33; Abd Ai-Qadir, 2015). The second term, التضام (*al-taḍāmm*), which could also be translated as 'inclusion of words together', indicates the correlational relationship of one word with another that together constitute a collocation. Husamaddin (1985) utilises the term المصاحبة اللغوية (*al-muṣāḥabah al-luġawīah*) to refer to collocations in Arabic, denoting the linguistic companionship between words that form idiomatic expressions. Husamaddin explains this term as المصاحبة الاعتيادية لكلمة ما في اللغة بكلمات أخرى معينة (*al-muṣāḥabah al-ʾʕtiyādyiah Likalimaẗ mā fi āl-luġh bikalimāt āuxrā muʕyanh*) 'the normal co-occurrence of a word within a language with other specific words' [Author's translation] (1985, p.257). Husamaddin illustrates this using collocations based on several classifications related to meaning, including representative sounds such as زئير الأسد (*Z'ayir Al-'asad*) 'lion's roar' or with reference to the names of places where animals live, in this case اسطبل الخيل (*Āisṭablu Āl-xyl*) 'horse stable'. Husamaddin's definition seems to allude to a sense of collocability within collocations, as in the example just referred to, of how زئير (*z'ayir*) 'roar' collocates with اسد (*asad*) 'lion'. The expression زئير الأسد (*Z'ayir Al-'asad*) 'lion's roar' is frequently used metaphorically to describe a powerful or intimidating sound or action. The term can only be used in conjunction with the noun 'lion'. The statement implies that زئير (*z'ayir*) 'roar' and اسد (*asad*) 'lion' have

[9] Author's translation of the term الترادف اللفظي (*al-tarāduf al-lafḍī*).

a strong collocational relationship in Arabic, as evidenced that a lion's ferocious roar is only described when the زئير (*z'ayir*) 'roar' is used. Another Arabic linguist, Ghazala (1993), uses المتلازمات اللفظية (*al-mutlāzmāt al-lafḍhīa*) 'verbal collocations' [Author translation] to denote a linguistic correlation between words. There are other possible terms for collocation in Arabic based on lexical and contextual approaches, such as التلازم (*al-talāzum*) refers to 'correlations between words', الأقتران (*al-iqtirān*) 'conjugation',الرصف (*Al-rāṣf*), or التوارد (*al-tāwārud*) are all possible synonyms for 'collocation' in Arabic (Igaab and Abdulhasan, 2018, p.90).

Ghazala classifies collocations into three types: grammatical patterns, verbal patterns and stylistic patterns. Ghazala's first type is of concern because he further classifies this collocation type into twenty grammatical patterns, three of which are classified according to the semantic relationship involved. Ghazala's grammatical patterns which are related to this study include Verb-Noun collocations, Verb-Adverb collocations and Adjective-Noun collocations. Interestingly, Ghazala's patterns do not include Adverb-Adjective collocations, or the reverse order of this collocation (Adjective-Adverb).

Table 2-3: Some Arabic examples[10] (taken from Brashi (2005, pp.37-39))

Verb-Noun collocation	سَمِعَ صوتاً(sami'a sawtan)	to hear a sound (non-figurative meaning)
	لَقَّنَ درساً (laqna darsan)	teach a lesson (figurative meaning)
Verb-Adverb collocation	بات شبعاناً (baata shab3anan)	to sleep with full stomach
Adjective-Noun collocation	قولٌ سديد (qawlun sadeed)	a true saying

These three collocational types appear in Ghazala's (1993) classification, although due to Arabic syntax, the Adjective-Noun collocation appears in reverse order in Arabic as Noun-Adjective.

[10] The transliteration of these examples is presented as they appear in Brashi (2005) without adaptation.

In the following sections (§2.3.1-2.3.4), I explain the four collocation types investigated in this study and illustrate how these collocation types are structured in Arabic and English to clarify some issues that may arise when the learners' L1 is Arabic. The following subsections will assist in understanding the background to Arab EFLLs' use of English lexical collocations.

2.3.1 Adverb-Adjective collocation

The Adverb-Adjective collocation type is used to express purpose or to add a feeling or a special quality to the adjective (Nofal, 2012, p.82; Siddig et al., 2022, p.4203). Ghazala's classification suggests that the Adverb-Adjective collocation type is not found in Arabic. However, Brashi's (2005, p.46) classification of Arabic collocations includes the Adjective-Adverbial phrase. The adverbial part of the phrase consists of a preposition plus noun (e.g. بشدة (*bišiddah*), literally 'with strength', in مُسْتَنْكر بشدة (*musttnkir bišiddah),* translated into English as 'strongly condemns'. The problem with Brashi's classification is that when his example is translated into English, it creates a Verb-Adverb collocation, which supports the idea that there is no Adverb-Adjective collocation in Arabic. In the example given, مُسْتَنْكر (*musttnkir*) 'condemns*'* is an active participle in Arabic, functioning as a verb and having the form اِسْتَفْعَلَ- يَسْتَفْعِلُ (*āistafʕalā - yāsttfʕil*) (Schulz, 2004, p.29). Based on Rabeh classifications to collocations, Adverb-Adjective collocations may not exist in Arabic (2010, p.24). However, some of the Adverb-Adjective collocation are said in the reversed order, such as the collocation كبير جدا (*kabīr jidā*) 'very big' are Adverb-Adjective collocations (where جدا is an adverb, and كبير an adjective). Similarly, Grimm states that there is no equivalent category in Arabic to the English Adverb-Adjective collocation, as concepts expressed as adjectives in English are expressed as nouns in Arabic (2009, p.38). This is because the grammatical category of noun (اسم) in Arabic, covers (subsumes) both 'noun' and 'adjective' in English; i.e. what is an 'adjective' in English grammatical terminology is a type of noun in Arabic grammatical terminology. Grimm claims that there are no corresponding combination of Adverb-Adjective as a collocation in Arabic, due to the intensification of an adjective been expressible by other means. For example, Grimm explains that an adjective can be expressed by a single lexeme (e.g., the collocation جَمِيلٌ جداً (*jamīl jiddan)* 'very beautiful' can be translated as a single word as وسيمٌ (*wasīm*) *'*handsome*'*);

similarly صافيٌ جداً (*sāfy jiddan*) and واضِحٌ جداً (wādḥ *jiddan)* can be translated comparatively (e.g. اصفى من الماء (*aṣfā man almāʾ*) 'clearer than water'); or as a simile (e.g. واضح وضوح الشمس (*wāḍḥ wḍūḥ alššams*) 'clear as the sun'). Based on Grimm's argument, it seems because of the fact that the Adverb-Adjective collocation can be expressed by a single lexeme does not demonstrate that it is not an Adverb-Adjective collocation (i.e. only that its meaning can be expressed in other ways). It is also, secondarily, a fact that وسيمٌ (*wasīm*) 'handsome' does not mean the same as جَمِيلٌ جداً (*jamīl jiddan)* 'very beautiful'.

One noteworthy example related to both the Adverb-Adjective and Verb-Adverb collocations is the set of rules for using adverbs in Arabic; in English, the adverb precedes the adjective it modifies, which is in not the case in Arabic. To illustrate this, Diab (1997) performed an error analysis study on Lebanese EFL students' written essays in which he found word-order errors specifically in the placement of adverbs.

e.g. *here are three rules **very** helpful > here are three **very** helpful rules*.

The Arabic version of this sentence is:

1. *فيما يلي ثلاث قواعد مفيدة للغاية* [11]
2. فيما يلي ثلاث قواعد جداً مفيدة

As seen in Diab's example, the adverb 'very' in English should be placed before the adjective 'helpful' followed by a noun being modified by the Adverb-Adjective (e.g., 'a very helpful tool'), the error in the first version of the example above being caused by L1 transfer from Arabic to English. It shows that learners experience problems with the placement of adverbs due to L1 transfer (Al-Shormani and Al-Sohbani, 2012; Rajab et al., 2016). Arabic dialects play a role in this error as the reversed order can also be true as جداً مفيدة (*jidda mufiydat*) 'very helpful'. It is, perhaps, possible that the respondents here are being influenced by their own Arabic dialect in addition to English (or, even, rather than English).

[11] There are many possible ways to formulate this sentence in Arabic. This is a translation of a sentence stated by a learner in Diab, N. 1997. The transfer of Arabic in the English writings of Lebanese students. *The ESPecialist.* **18**(1), pp.71-83.

2.3.2 Verb-Noun Collocation

In this section, I illustrate the use of the Verb-Noun collocation, which requires me to discuss the two basic Arabic sentence[12] structures. The Arabic language has two basic sentence types, nominal and verbal, their names being based on the class of the first word in the sentence (Schulz, 2004, p.170; Dickins, 2020, p.3). Thus, the nominal sentence contains a predicand[13] (a noun) and a predicate, and the verbal sentence contains a verb, a subject (optional) and an object (optional). The structure of the two sentence types is as follows:

- Nominal clause/sentence: Subject-Verb-Object (SVO)
- Verbal clause/sentence: Verb-Subject-Object (VSO)

The subject in an Arabic nominal clause/sentence is nominal, and equivalent to a noun or other nominal (e.g. noun phrase) in English. A nominal in Arabic and English can be of different types, such as a pronoun, demonstrative or noun (e.g. names of people, animals, and inanimate objects). In Arabic, a nominal sentence requires a predicate to complete its meaning. Dickins explained that predicate can take five forms including nominal phrase, adjectival, verb phrase, adverbial, clausal (2010a, p.240). For example a predicate as nominal can be either a single اسم (*ism)*, i.e. noun which is an adjective in English, rather than a noun (e.g., جديدةٌ (*jdydat*) '*new'* in السيارة جديدةٌ (*al-syārat jadydat*) *'the car is new')*, or 'quasi-sentence' which consists of a preposition and a noun (e.g*.,* ساره في البيت (*sara fiy al-bayt*) *'*Sarah is in the house'), or a verbal sentence (e.g., ساره تحبُ السفر (*sara tuḥibu al-sfr*) *'*Sarah loves travelling').

While there other structures, Schulz explained that the Arabic verbal sentence can have one of three basic structures, which can be regarded as the most common; either (1) consisting of only a verb, (2) or a verb + subsequent subject, (3) or a verb +subject+ object (2004, p.174). The following lines give some examples of these structures:

---

12 The word 'sentence' is used because in Arabic grammatical tradition, the word 'sentence' does not distinguish between the two notions of 'clause' and 'sentence' (Dickins, J. 2010b. Junction in English and Arabic: Syntactic, discoursal and denotative features. *Journal of Pragmatics.* **42**(4), pp.1076-1136.)

13 The predicand refers to المبتدأ به (*mubtada' bi-h*) 'begun with-it', unambiguously distinguishing it from فاعل (faʔil) 'the subject of a verbal clause/sentence' (Dickins, 2010, p11).

1. Verbal sentence consisting of only a *verb*:
    - رَكَضَ (*rak*ḍa) ‘He ran’.
    - وَصَلَتْ (*waṣalt*) ‘She arrived’.
2. Verbal sentence consisting of a *verb* + *subsequent subject*
    - ذَهَبَ الرَّجُلُ (*ḏahab al-rrajul*) ‘The man has gone’.
    - وَصَلَت السَّيِّدَةُ (*waṣalt al-ssyadat*) ‘The lady has arrived’.
3. Verbal sentence consisting of a *verb* (+*subject*) + *object*
    - شَرَبَ (الوَلدُ) العَصيرَ (šarab (al-waladu) al-ʕaṣiyr) ‘He (the boy) drank the coffee’.
    - طَلَبَتْ (السَّيدةُ) الرِسَالة (ṭalabat (al-ssayadatu) al-rriysaala) ‘She (the lady) requested the letter’.

Without a verb, it is impossible to construct a meaningful clause/sentence in English (discounting some marginal exceptions such as ‘The more the merrier.’). In English, the verb acts as the primary clause/sentence component, as it does in the verbal clause/sentence in Arabic. There are, of course, differences between the two languages in word order and clause/sentence structure (Siddig et al., 2022, p.4201). The Arabic verb has two basic tenses: perfect tense and the imperfect tense (Schulz, 2004, pp.13-14). The perfect tense is roughly equivalent to the past tense in English, most basically denoting an action that is done and completed in the past (e.g., ‘he <u>wrote</u>’, and ‘he <u>has written</u>’ both have كَتَبَ (*katab)* as their most obvious Arabic equivalent). The imperfect tense most basically denotes an action taking place in the present (e.g., يَشْربُ (*yašrabu*) *‘he <u>is drinking</u>’*) or in the future (e.g., سَيَشْربُ (sa*yašarabu*) ‘he <u>will drink</u>’ for which another possible version is سَوْفَ يَشْربُ (swfa *yašarabu*). Schulz explained that سَوْفَ (*swfa)* is a particle and its abridged form سَـ (*s*) acts as the direct equivalent of *will* in Arabic, placed at the beginning to emphasize the future meaning (2004, p.14). It also, however, frequently denotes an ongoing or unfinished action in the past. Saleh and Fatah explained that the purpose of the perfect and the imperfect tense in Arabic is to indicate whether the action is been performed or not (2022, pp.66-67).

2.3.3 Verb-Adverb Collocation

In this section, I explain the use of the Verb-Adverb collocation in Arabic and English. The adverb position within the English language is movable under specific conditions (McConnell-Ginet, 1982; Radford, 2009). This is in line with Arabic adverbs as they can also occur in multiple positions in the sentence, leading to several possible potential error types that can be caused by L1 interference (Al Aqad, 2013, p.73). One difference between the two languages is that the adverb placement in Arabic is more flexible than it is in English, and its position does not affect the meaning. In English, when we need to express a special impression or emphasis, the adverb precedes the verb, as in 'strongly agree*'* (Nofal, 2012, p.82). As for the use of verbs in Arabic, they are an essential element in verbal sentence formation (Alhuthaily, 2000). Arabic words typically involve radicals, i.e. word-roots made up of three consonant (or in some cases four consonants), traditionally represented in Arabic grammar using the root فَعَلَ (*f-ʕ-l*) (Abu-Chacra, 2007). Thus, radicals are the basis, in almost all cases, for forming verbs, nouns, adverbs and adjectives. As seen in example 2.1, below, the verb position varies in Arabic and the verb can be placed before or after the subject. Unlike English, Arabic is 'a free word order language' in which this kind of sentence could have four different forms of word order (Al Aqad, 2013, p.71):

Example 2.1 Possible Arabic Sentence Structures (taken from Al Aqad, 2013, p.71)

Subject (S)-Verb (V)-Object(O),	e.g., الولد يكتب بسرعة (*al-wld yktb bisurʕah*) 'the boy writes ***quickly***'
Verb (V)-Subject (S)-Object (O),	e.g., يكتب الولد بسرعة (*yktb al-wld bisurʕah*) 'writes the boy ***quickly***'
Verb (V)-Object (O)-Subject (S) ,	e.g., يكتب بسرعة الولد (*yktb bisurʕah al-wld*) 'writes ***quickly*** the boy'
Object (O)-Verb (V)-Subject (S)	e.g., بسرعة يكتب الولد (*bisurʕah yktb al-wld*) '***quickly*** writes the boy'

The adverb بسرعةٍ[14] (*bisurʕah*) 'quickly' can be placed either in initial, medial, or final position in an Arabic sentence, creating four possible positions within an Arabic sentence structure, which explains the source of difficulty for Arab EFL learners in using adverbs.

[14] Al Aqad, 2013 refers to the adverb بسرعةٍ as an object in the above examples, but it is considered as a subjunct by Quirk, R. and Greenbaum, S. 1973. *A Concise Grammar of Contemporary English.* Harcourt School. Subjuncts include adverbs that express an intensification of a condition.

Therefore, the main difference between the two languages in terms of using the Verb-Adverb collocation is the unfixedness of the placing of the placing collocation components in Arabic. In addition, in Arabic, the adverb position does not necessarily affect the intended meaning of the collocation, as seen in example 2.1. Sometimes, however, it does. Consider, for example, the difference between the two sentences below:

Sentence one:

لم يتكلم بصراحة (*lm yatakalamu biṣaraāḥ*) 'he did not speak frankly' (i.e. he spoke but what he said was not frank').

Sentence two:

بصراحة لم يتكلم (*biṣaraāḥ lm yatakalam*) 'frankly, he did not speak' (i.e. I am telling you frankly that he did not speak (i.e. he did not say anything)).

Therefore, the position of the adverb in both languages contributes significantly in the delivering the intended meaning. The two examples given for understanding the differences and similarities between the two languages explain some of the potential errors that could be attributed to the Arab EFLLs' L1. The possibilities of using Verb-Adverb collocation are vast in both languages, which contributes to the difficulty for EFLLs in using this collocation. After this discussion of Verb-Adverb collocations, in the following section (§2.3.4), I explain the fourth and final collocation type investigated in this study, Adjective-Noun collocations.

2.3.4 Adjective-Noun Collocation

In Arabic, the word order of the Adjective-Noun collocation is the reverse of the equivalent structure in English. For example, the collocation 'key role' is said as دوراً رئيسيا (*duwr raʾīsiya)* 'role key' in Arabic. Some authors view Adjective-Noun collocations as weak because they are easily predicted by EFLLs (El-Dakhs, 2015; Alangari, 2019). This collocation type can be used literally in Arabic and English. For example, the collocation 'good boy' has the Arabic equivalent ولدٌ صالح *(walad ṣāliḥ)*. Another literal Arabic example is حربٌ قوية (*ḥarb qawiyah)* 'strong war'. Noun-Adjective collocations can also be used figuratively in both languages,

as in Arabic حربُ طاحنةٍ (*ḥarb ṭāḥinaat)* (literally: ‘A grinding war), which means figuratively ‘devastating war’ or ‘damaging battle’. An example of a figurative English Adjective-Noun collocation is ‘strong tea’, which corresponds to the also figurative Arabic form شاهي ثقيل (*šāhiy ṯaqiyl*) ‘heavy tea’. This collocation seems to be easier than other types of lexical collocation used by EFLLs from different L1s (see, for example, for Dutch EFLLs, (Peters, 2016), or for Arab EFLLs (Shehata, 2008; Youcef, 2022).

To summarise this section (§0), I created table 2-4 to highlight the main similarities and differences in using the four collocations in Arabic and English.

The significant findings of this review of collocation in Arabic and English are:

- Collocations are predictable in both languages and are present in our daily life (Abu Khaled, 2020). For example, in English, ‘take a shower’. In Arabic, we have the most well-known Islamic collocations, such as القران الكريم(*al-qurā’ān al-lkarīm*) ‘the holy Quran’ and مكة المكرمة (*Mākkah al-Mukrmh*) ‘the holy Makkah’.

- Trying to use the accurate and appropriate collocate is regarded as a continual struggle for language learners (El-Dakhs, 2015).

- All four collocations exist in Arabic, except the Adverb-Adjective collocation which is often claimed as not being an Arabic collocation (Ghazala, 1993).

Table 2-4: Summary of the use of the four lexical collocations in Arabic and English

Collocation in English	Collocation in Arabic	Justification
Adverb-Adjective Collocation (e.g. *absolutely delighted*)	Null	The combination of Adverb-Adjective does not exist in Arabic Brashi (2005) states that this collocation exists in Arabic in the reverse order, as an Adjective-Adverbial phrase (i.e., preposition+noun) such as مُستنكرٍ بِشِدَةٍ (*mstnkr bšdat*) 'strongly condemns'
Verb-Noun Collocation (e.g. *ask a question*)	The Verb-Noun collocation in Arabic can take the following forms: 1. Verb-Noun collocation 2. Noun-Verb Collocation 3. Verb only	The combination exists in Arabic either in the same word order or with the order of components reversed. For example: - Verb-Noun Collocation (e.g. *book an appointment* يحجزُ موعداً (*yḥajiz mwʕid*). - Noun-Verb Collocation (e.g. مواء القطط (*miw'ā al-qiṭṭ*) 'cats mew'). - Verb-only collocation without a noun (e.g. يُصافح (*yuṣaāfiḥ*) 'shake hands').
Verb-Adverb Collocation (e.g. *play carefully*)	In English, the adverb can be placed at the beginning, middle or end of a clause. - Mid position. (e.g. 'you <u>*can definitely*</u> pass this exam'). - The adverb precedes the verb when adding emphasis or a special feeling (e.g. 'strongly suggest').	- The position of adverbs is flexible in Arabic. They can occur either before or after the verb they modify. For example: □ فرِحَ فرَحاً (*fariḥ faraḥ*) translates into English as '*joyful joy*' □ مسرورٌ للغايةِ (*masruūr illġaāyat)* 'absolutely delighted' □ جداً سعيد (*jida saʕiyd)* 'very happy' can be said in reverse order without affecting the meaning.
Adjective-Noun Collocation (e.g. *medical care*)	Adjective-Noun collocations can have figurative and literal meaning: □ Figurative meaning (e.g. *strong tea,* in Arabic شاهي ثقيل (*šāhy ṯqīl)* 'heavy tea'. □ Literal meaning (e.g. حرارةٌ مرتفعه (*ḥrārat mwrtfʕh)* 'high temperature'.	Adjective-Noun collocations exist in Arabic but in the reverse Noun-Adjective order. For example, the collocation 'good reason' isسببٌ وجيه (*sabab waājiyh*) is said literally as 'reason good' in Arabic. In Arabic, the Adjective-Noun collocation can have both figurative and literal meaning (§see examples given in the column on the left for the Adjective-Noun collocation in English).

- However, some of the literature reviewed claim the Adverb-Adjective collocation is present in Arabic when the order of the elements is reversed (Grimm, 2009; Khaled, 2020).

- Word-order issues pose a difficulty for some collocations, but should be considered a positive point that language learners should be made aware of to help them predict collocations.

- The Verb-Noun collocation exists in Arabic, according to Ghazala's (2007) classification. However, the equivalent use of the Verb-Noun collocation in Arabic varies from English because the noun in the collocation is occasionally treated as an object. Therefore, a word-order issue arises from L1 influence on the use of the Verb-Noun collocation. According to Ghazala (2007), Verb-Noun collocations can be used in Arabic following the same order in English, for example, يحجزُ موعداً (*yaḥjizu mawʕid*) 'book an appointment' or, in the reverse order, Noun-Verb collocations that correspond to Verb-Noun collocations, for instance, نباح الكلاب (*nibāḥ al-kilāāb*) 'dogs bark' or مواء القطط (*miwāā' al-qiṭṭ*) 'cats mew' (Ghazala, 2007). In addition, there could be an omission of the noun in the collocation in Arabic when the noun is viewed as the object in the collocation (Dickins et al., 2016). For example, the collocation 'shake hands' can be said as يُصافح (*yuṣāfiḥ*) 'to shake' in Arabic without saying the noun *hand*. The second collocation pattern is mostly restricted to the names of sounds. In this thesis, the Verb-Noun collocations under investigation include collocations with/without an article between the verb and the noun (§Chapter 4 and Chapter 5 for the results and discussion of the use of the four collocations by Arab EFLLs and NBESs).

In the previous sub-sections (§2.3.1-2.3.4), I have further illustrated lexical collocation differences between Arabic and English and the effects of L1 interference. In addition, I have illustrated some issues that should be highlighted for Arab EFFLs when using word classes in English (i.e., nouns, verbs, adjectives, and adverbs).

2.4 Previous Studies on the use of collocation by learners of English

In this section, I provide a review of the literature on the concept of collocation and studies of collocational patterns by learners of English. Collocation is part of how we use language to communicate, either in spoken or written form. Collocations are a crucial element in writing, and employing them correctly will help a learner's writing appear to be like that of a native English speaker (Zhang, 1993; Namvar, 2012). Accurate use of collocations by EFLLs is a key indicator of proficiency, or fluency in English (Pawley and Syder, 1983; Lewis and Conzett, 2000).

Namvar (2012) investigated the written collocation use of Iranian EFLLs using multiple-choice tests and writing tasks. This study focused on grammatical and lexical collocations in the students' written work. The study included 15 participants at postgraduate level to assess their collocation errors and to identify the source of their collocation difficulties. The findings suggest that learners encounter difficulties with both types of collocation and that L1 transfer affects the learners' use of collocations. Namvar found that for Iranian EFLLs the Verb-Noun collocation was the easiest category, and that the Adjective-Noun collocation was the most challenging (2012, p.49). Namvar suggests that learners' L1 contributed positively to the use of collocations, mainly when there was an equivalent of a collocation in Persian.

A more recent study by Boonraksa and Naisena (2022) on Thai EFLLs' use of collocations also revealed that collocational errors and difficulties are induced by the influence of the learners' L1. In Boonraksa and Naisena's study, the collocations were given as multiple-choice questions to three groups of learners of high, medium, or low proficiency, plus another translation test for collocations. The words for the collocations were taken from a wordlist that basic English users can understand and should be familiar with. The error rate for lexical collocations was high for all three groups of learners. The most challenging collocation type for Thai EFLLs was the Adverb-Adjective collocation. Most high-proficiency learners performed better in grammatical collocations such as Verb-Preposition,

and both medium and low-proficiency learners did better for Adjective-Preposition collocations. L1 interference is observed in the learners' errors in which they adhere to the same rules of their L1 regardless of English structure. For instance, verb choice in Verb-Noun collocations was influenced by the learners' L1, e.g. the English expression 'save a seat' is said as '*book at place for me' in Thai. Also, direct synonyms from the learners' L1 were an additional cause of collocation errors. For example, Thai EFLLs tend to say '*straight flight' as opposed to 'direct flight'.

Many research studies have identified L1 influence as the major source of English written errors for Arab EFLLs: i.e., the collocations they produce might sound unnatural to a native speaker of English (Abushihab et al., 2011; Hago and Ali, 2015). Mahmoud (2005) created a study to gather, categorise, and analyse learners' collocation errors in three writing samples of learners at two different levels: post-intermediate and advanced. Forty-two written samples of the weekly homework produced by third-year male and female university students[15] were collected without telling the participants the main reasons for the study, as this might have led to the overuse or underuse of collocations in the samples. The study found usage of both grammatical and lexical collocations, totalling 420, in the participants' written essays at levels of 20% and 80% respectively. The most common types of lexical collocation were combinations of verb and noun, and noun and verb. The learners used 336 lexical collocations, of which 269 (64%) tended to be incorrect by the native raters' judgment. Mahmoud suggested that while correct usage of collocations could be a result of either the learners' exposure to the English language or of positive transfer from Arabic, incorrect usage could be the result of interlingual transfer from Arabic; occasionally such interlingual errors would be due to problems learners have in their first language, as they might get confused by two similar words in Arabic. Thus, Mahmoud indicates that errors in the use of collocations depend on learners' reliance on interlingual and intralingual strategies, which could ease usage in the case of linguistic similarities but might in turn, lead to problems in the case of differences. From my perspective, Mahmoud's study aimed to find the types of collocations

[15] Third year students in this study are in their pre-final year in the Saudi Arabian system. Bachelor's degrees consist of 4 years along with a preparatory year.

which Arab learners tend to employ when using English; however, the interpretation of the findings was built on the author's assumptions. Mahmoud (2005) should have interviewed the learners to gain a deeper insight into their intentions in employing a specific choice of wording in the collocations they used.

It seems very probable that the L1 of Arab EFLLs affects the production of their English lexical collocations. Abu Khaled (2020) investigated the use of lexical collocation by Arab EFLLs from Palestine through interviews and written tests. Abu Khaled's study found that incorrect use of collocations was higher than correct attempts, accounting for 73.3% and 24.2%, respectively. This study found that learners find it difficult to select the accurate collocate. The Palestinian EFLLs tended to rely on their L1 in the choice of lexical items. For example, the collocation 'wear makeup' is said as 'put makeup'. This demonstrates that choosing collocates based on the L1 may not be an appropriate strategy, in line with Zughoul who stated that "what collocates in one language does not necessarily collocate in another" (1991, p.52). However, in English, the two collocations are semantically different: 'wear makeup' is habitually use makeup; 'put makeup on' is to apply makeup. Regardless of the difference in meaning, in Arabic, while the verb 'wear' is commonly used with clothes, the verb 'put' is commonly used when applying makeup and using perfumes.

2.4.1.1 Corpus-based error analysis studies of EFLLs' use of collocations

In this section, I review some previous corpus-based error analysis studies of EFLLs whose L1 is not Arabic. Collocations place a certain burden on language learners, who tend to make errors when using collocations regardless of their level of language competence (Nesselhauf, 2005; Deveci, 2018). Previous studies have explored the use of Verb-Noun collocations by EFLLs other than Arabs. A well-known study by Nesselhauf (2003) found that even advanced German EFL learners encountered difficulties with collocation, specifically the Verb-Noun collocation, which was the main type examined in her study. In a similar vein, Laufer and Waldman (2011) investigated the use of Verb-Noun collocations by Hebrew learners of English using a learner corpus. The results of both these studies were somewhat similar to those obtained by Mahmoud (2005),

in that the two studies found that even learners with advanced proficiency in English tended to commit interlingual errors when using collocations. Laufer and Waldman (2011) also found that learners utilise fewer collocations overall than native speaker users of English, regardless of their proficiency level. It seems that although the two studies concentrated on learners who had different first languages, Arabic and Hebrew, both sets of learners tended to rely on inter-lingual strategies during second language use.

Hong et al., (2011) performed a corpus-based error analysis on Malay EFLLs to investigate the types and sources of written Verb-Noun collocational errors, allowing for possible intermediates such as a determiner or a preposition. Hong et al., used Gass and Selinker's (2008) error analysis framework. The results suggest that choice of intermediate within a collocation has the highest error rate, including wrong choices of preposition or determiner. The second error encountered was in the choice of verb in a collocation. For example, Malay EFLLs would use the verb 'drop*'* instead of 'fall' in the collocation 'fall in the river'. The same was observed in nouns, as the learners tended to misuse them. It was found that intralingual transfer was the primary category of collocational errors: the learners' ignorance of the grammatical restrictions on collocations contributed towards the inaccurate use of English collocations. Two of the intralingual error sources found in Hong et al.'s study were *approximation* and *overgeneralisation*. Approximation is used when learners are unsure about the target language structure. For example, 'picking flowers' is said as 'cutting flowers' by Malay EFLLs (2011, p.40). Overgeneralisation is when learners deviate from the appropriate language use because learners are applying familiar language structures: for instance, the use of the past *-ed* ending on irregular verbs. Hong et al.'s results align with Nesselhauf's (2005) results on German EFLLs that the major source of collocational errors is intralingual.

A great deal of previous research into the use of the Adjective-Noun collocation by other learners of English as a foreign language found similar results to those of Arab EFLLs scholars in terms of choice of lexical collocations. Indeed, the choice of lexical items within the Adjective-Noun collocations by Croatian EFLLs was found to be highly influenced by the topics chosen for the learners' essays

compiled within the corpus (Takač and Lukač, 2013). Takač and Lukač used three corpora: the Croatian corpus of English Learner Essays (CELE) compared with two reference corpora, which were the BNC and the International Corpus of Learner English (ICLE). Some examples of deviant choice of adjectives within the Adjective-Noun collocation, according to Takač and Lukač, were ‘poor students’ and ‘inappropriate clothes’. Takač and Lukač explaines that the collocations ‘inappropriate clothes’ is based on learners L1 corresponding to more detailed collocations such as “outrageous, wrong, gaudy," clothes (2013, p.392). The choice of adjectives in the collocations given creates semantic vagueness in the meaning of the collocations. Although neither collocation seems deviant in isolation, Takač and Lukač claim that there is no equivalent to these collocations in the reference corpus used in the study (i.e., the BNC). It should be noted that there is a potential risk of flagging collocation errors when the two-word combinations are actually used correctly. This danger arises primarily due to the lack of contextual information provided by the authors. As a result, without the proper context, the collocations may appear to be acceptable as individual units. Takač and Lukač found that Croatian EFLLs overuse general adjectives (e.g. *bad, different, good,* etc), more than other specific adjectives. Takač and Lukač were not clear as to what the specific adjectives in their study were. Also, they found that concordance analysis provides more insight into syntagmatic relations. Interestingly, this study was one of the main studies that used statistical significance testing for Adjective-Noun collocations. The main hypothesis was to test whether the Croatian EFLLs Adjective-Noun collocations were statistically significant in the assigned threshold in the BNC and the ICLE corpora. The results found that some collocations were not statistically significant different in either the BNC or the ICLE. However, some collocations were more statistically significant in the ICLE than the BNC due to the difference in size between the two reference corpora.

2.4.1.2 Corpus-based error analysis studies of the use of collocation by Arab EFLLs

In this section, I review the previous literature of corpus-based error analysis studies of Arab EFLLs. Several scholars (Farooqui, 2016; Alangari, 2019; Khoja, 2019; Youcef, 2022) have examined the use of collocation by Arab EFLLs to identify areas of difficulty and potential causes of errors. There have not been many corpus-based studies of Arab EFLLs' use of collocation. The most extensive study is that of Farooqui (2016), who aimed to investigate the difference between the use of academic written collocations by native and non-native speaker computer science students. Farooqui aimed to explore the reasons behind the under/overuse of 24 selected lexical collocations. She focused on two types of collocations: Noun-Preposition-Noun and Noun-Adjective-Noun collocations. Farooqui found that topic and discipline-related collocations were the main factors contributing to the learners' lack of knowledge of collocations.

Adopting a phraseological approach, Alangari (2019) performed a corpus-based study of Adjective-Noun and Verb-Noun collocations on Arab EFLLs from Saudi Arabia. The results indicate that, due to L1 influence, Arab EFLLs misuse the Verb-Noun collocation far more than the Adjective-Noun collocation. Also, Alangari found that Saudi advanced learners' choice of verbs was limited to a small set of general verbs. The frequency of the correct Adjective-Noun collocations indicated no statistically significant difference between advanced Arab EFLLs and native speakers. A major finding was that Arab EFLLs tend to repeat some common English Adjective-Noun collocations such as 'long time', 'social life' and 'best way'.

Another corpus-based study of Arab EFLLs from Saudi Arabia is that of Khoja (2019). She investigated four collocation types: Verb-Noun, Noun-Verb, Adjective-Noun, and Noun-Noun. The data were taken from 16 written essays of learners from two proficiency levels: pre-intermediate and intermediate. The researcher extracted the collocations manually and then compared them using the BNC as the reference corpus to determine the statistical association measure

(i.e., LogDice[16]) for judging the learners' collocation use. Khoja also implemented a phraseological approach using native speaker intuition to assess the accuracy of collocation use. She found that Saudi Arabian EFLLs produced a high percentage of acceptable collocations without explicit teaching. The results indicated that the pre-intermediate learners tended to repeat the same words within the collocations. Yet, the intermediate-level learners differed as they used a variety of lexical items leading to their producing a higher degree of accuracy in the collocations used. Khoja in her study claimed that though learners were based in their home country with little exposure to English, their collocation use was more successful than those based in an English-speaking country, contradicting many other studies that believe lack of L2 exposure leads to greater use of unacceptable collocations (Fan, 2009). This only a claim that she made without reporting any results for those based in an English-speaking country.

A more recent corpus-based study by Youcef (2022) of the use of Adjective-Noun collocations by Arab EFLLs from Algeria found that learners were restricted to certain lexical choices. Youcef found that the number of correct Adjective-Noun collocations was high, accounting for 59% of the total used collocations in the corpus. However, collocational errors were seen in the choice of the collocates, such as 'great combination' instead of 'ideal combination'. The errors were attributed to negative transfer between modern and non-standard Arabic. For example, the collocation 'historical landmarks' is said by Algerian Arab EFLLs as 'traditional places' in the context '*traditional places related to town edges to remember their ancestors' (Youcef, 2022, p.932). The other errors were attributed to intralingual factors, such as dependence on synonyms and overgeneralization.

Mammeri (2015) implemented a corpus-based study on morphosyntactic errors made by Arab EFLLs also from Algeria. Morphosyntactic errors are those caused by misuse of syntactic rules. Mammeri collected written samples from 120 second-year university students. The results revealed eight types of errors: "word

[16] LogDice is explained in 2.1.11.3.

order, subject-verb-agreement, verb structure, noun/adjective/adverb structure, word/morpheme addition, word/morpheme omission, short forms/abbreviation, and inaccuracy of use of conversational and informal words" (2015, p.117). For example, word order was seen in the use of the following combinations:

> *- *He introduced expressions very clear – He introduced very clear expressions.*
>
> *- *And always he felt that he will soon died – and he always feels that he will die soon.*

There seems to be a problem for Arab EFLLs with the placement of adverbs in English, which usually precedes the subject in Arabic. Mammeri found many errors related to the inaccurate use of the noun/adjective/adverb structure, such as '*his long sufferance' instead of 'his long suffering' and '*mental sick' instead of 'mental sickness' (2015, p.119). The errors are caused by both intralingual and interlingual factors, such as overgeneralization, interlanguage interference (i.e., mainly with French), and ignorance of L2 rules.

2.5 Conclusion

To conclude, in this section (§2.1), I have explained and defined some terms and concepts used in this thesis. In section (§2.1.11.3), I have discussed two approaches to collocation. On the one hand, the frequency-based approach is associated with statistical criteria in defining collocation. Through this approach, collocation can be defined as the association between words that occur more often than expected by chance (Biber and Conrad, 1999). In opposition to the frequency-based approach, the phraseological approach depends on several factors to identify a collocation. Such factors include relying on collocation dictionaries (Laufer and Waldman, 2011, p.657), depending on native speaker knowledge of a language (Greenbaum, 1974, p.83), or, rather, depending on both to identify collocations, as suggested by Nesselhauf (Nesselhauf, 2003, p.50). Combining both approaches when investigating collocation can help clarify collocational patterns more precisely (Gablasova et al., 2017). One example of a collocation study is Nesselhauf (2005), who applies the phraseological approach to German learners of English. Unlike Nesselhauf (2005), both approaches have been used on Hebrew learners of English (Laufer, 2011) and on Arab postgraduate students in an English university (Farooqui, 2016). A fuller

discussion regarding the two approaches will be given in Chapter 3. In section (§2.2), I have explained the theoretical framework. In section (§2.3), I have provided an overview of the four lexical collocation use considering the differences between Arabic and English.

To sum up section (§2.4), considering all the corpus-based error analysis studies of collocation by EFLLs listed above, it seems that:

- initially, many studies found that the choice of lexical item for the verbs in Verb-Noun collocations, mainly trying to use the right equivalent, is challenging for many EFLLs from different L1s (Hong et al., 2011; Namvar, 2012). The right choice of collocate is a problem for many learners, yet we must consider that the topics of the learners' essays may govern their lexical choices. For example, Takač and Lukač stated that "Different topics and genres of texts influence lexical choice – which accounts for the presence or absence of lexical collocations" (2013, p.390).

- many EFLLs are not aware of the collocation restrictions in English (Hong et al., 2011; Takač and Lukač, 2013).
- EFLLs tend to overuse general vocabulary when using the collocations (Takač and Lukač, 2013).

To summarise the reoccurring themes in corpus-based studies of Arab EFLLs' use of collocations, the studies listed above suggest that:

- Arab EFLLs tend to have difficulty in using English verbs (Abu-Chacra, 2007; Chaaraoui, 2017). This difficulty with verbs was observed particularly in the choice of verb tenses.

- The collocational errors observed in most of the corpus-based studies are the use of incorrect synonyms, and that learners tend to overgeneralise from their experience.

Chapter 3 Methodology

Introduction

In this chapter, I present the methodology for this research project. I will explain the methods implemented to investigate the use of lexical collocations by Arab learners of English from Kuwait and Dubai compared with the usage of native British English speakers. The core of this PhD thesis follows a quantitative approach that discusses collocation use in terms of their frequency. I will elaborate on the methods, tools and processes followed. I will outline how corpus linguistics can be used as a methodology and describe the corpora used for this study. In addition, I will provide a critical evaluation of the selected approaches with reference to the previous literature.

In this thesis, I tackle lexical collocation from two theoretical perspectives, a corpus-based approach to frequency and contrastive error analysis (§3.1.2.1 and 3.1.2.2). Many works in the literature value the combination of both approaches in collocation studies (Nesselhauf, 2005; Farooqui, 2016; Chang, 2018). The corpus-based approach to frequency focuses on the number of occurrences; therefore, in brief, the higher the number, the more acceptable is the word combination. The problem with the corpus-based frequency method is that it only considers quantitative data, ignoring the grammatical and semantic significance of collocations. This may be mitigated by combing the corpus-based frequency method with contrastive error analysis (this is explained in 3.1).

The outline of this chapter is as follows: in Section 3.1, I will discuss the methodological approaches implemented in this thesis, including the corpus-based frequency and the intuition-based contrastive error analysis approaches. After that, the population and sampling design will be described in Section 3.2 and the tools used for this study detailed in Section 3.3. In Section 3.4, I will go through the analytical procedures—the statistical tests used--in depth.

3.1 Methodological Approaches

In this section, I will explain the methodological approaches employed in this thesis. To have a comprehensive overview of lexical collocation usage by Arab EFLLs and NBESs, a systematic two-level, mixed-methods methodological approach was designed.

The first level was to investigate frequencies. I will investigate the differences between Arab EFLLs' and NBESs' collocations in two corpora using frequency association measures. The corpus linguistic approach and the statistical analysis of frequencies approach will be merged into a single approach referred to, henceforth, as the corpus-based frequency approach. The second level will be a contrastive error analysis approach. The benefit of considering a contrastive error analysis approach alongside the initial frequency approach relies on the fact that the "syntactic relationship between the elements [in the frequency approach] does not play a role in deciding whether they form a collocation or not" (Nesselhauf, 2005). Therefore, having a two-level approach will be beneficial in identifying issues that could be attributed to L1 interference.

The second level also includes a descriptive-contextual approach. The descriptive-contextual approach is a combination of the *intuition-based contrastive analysis approach* with some reference to phraseology when needed in the analysis (§3.1.2.3). Corpus-based approaches have usually been contrasted with intuition-based approaches, which can be explained as approaches in which language researchers can create "purer examples instantly for analysis because intuition is readily available and invented examples are free from language-external influences existing in naturally occurring language" (McEnery et al., 2006, p.6). However, in this research project, the corpus-based and intuition-based approaches both involved using real language use for data analysis.

The following section describes how each approach was employed.

3.1.1 Benefits of a Two level, Mixed-methods Approach

In this section, I will explain the rationale for choosing a two-level, mixed-methods approach. Implementing a two-level, mixed-methods approach allows the overall research methodology to be explained from both quantitative and qualitative perspectives, with the emphasis on the quantitative. The quantitative analysis will be conducted using the *corpus-based frequency approach*. The qualitative analysis will be conducted using the *intuition-based contrastive error analysis approach* (§3.1.2.2). In brief, a qualitative approach investigates the 'quality' of how something is. It is much concerned with characteristics and qualities. In contrast, a quantitative approach is usually concerned with 'quantity' and explores how many entities there are, or how large an entity is, in relation to a specific item or characteristic (Rasinger, 2010, pp.49-56). Although a qualitative approach leads to richness and precision of analysis, a quantitative approach can lead to more reliable and generalisable statistical results. Through the qualitative approach, I aiming to understand how the Arab EFLLs and NBESs use lexical collocations and what might contribute to the differences between them. In terms of the quantitative research perspective, descriptive statistics will be used and will be given to explain the differences in the word frequency lists between NBESs and Arab EFLLs.

The above outline highlights that a researcher is limited when using a corpus-based approach, as some research questions cannot be entirely addressed by this approach alone. For example, the role of L1 interference is difficult to identify through a corpus-based approach to collocation alone. An intuition-based approach is needed to determine whether the concordance lines show collocational differences from a semantic or a grammatical perspective, (McEnery et al., 2006). Researchers working on corpus data need to balance the two approaches, the corpus-based and intuition-based, because both complement each other and can potentially give rise to a new research area in the field of linguistics (Fillmore, 1992; McEnery et al., 2006; Hu, 2020).

The main advantage of the corpus-based approach over an intuition-based approach derives from its ability to provide strong empirical evidence as opposed

to intuition. Yet, an intuition-based approach could still assist in bridging the gaps left by corpus approaches. Compared to an intuition-based approach, the corpus-based frequency approach is more objective in its analysis (Guz, 2016). These two approaches should therefore be viewed as complementary, as (Fillmore, 1992; Leech, 1992c). Fillmore stated that “intuitive linguists and corpus linguists need each other. Or better, […] the two kinds of linguists, wherever possible, should exist in the same body” (1992, p.35). This view was supported by Leech who clearly stated that “Neither the corpus linguist of the 1950s, who rejected intuitions nor the general linguist of the 1960s, who rejected corpus data, was able to achieve the interaction of data coverage and the insight that characterize the many successful corpus analyses of recent years” (1992c, p.14). Other scholars such as McEnery and Wilson (2001) and McEnery et al. (2006, p.7) also believe that the two approaches can be considered complementary and should be so if a conceivable wide scope of research questions is to be addressed by language specialists.

3.1.2 The Two-level Mixed-methods Approach

In this section, I will discuss the two-level, mixed-methods approach adopted. The first level is the corpus-based frequency approach. The second level is the contrastive error analysis approach, which depends on the first level; it examines the outcomes of the corpus-based frequency analysis. The second level requires and includes the third approach, which I refer to as the descriptive-contextual approach. The descriptive-contextual approach combines the intuition-based approach of contrastive analysis with phraseology.

3.1.2.1 Level 1 - The corpus-based frequency approach

The first level of the mixed methods approach involves running statistical analyses on collocational frequency patterns generated from the research corpora. Because *corpus linguistic tools and methods* and the *statistical analysis of frequencies* are approaches that complement each other, I refer to this first level as the *corpus-based frequency approach*. I will comment briefly on the relationship between the two approaches. Corpus-based analysis is used widely

in language pedagogy and linguistics. A corpus-based approach can be explained as a method utilising computational tools to gather spoken or written data to study a linguistic phenomenon.

In this thesis, a corpus-linguistic methodology is essential as a first step. There are computational tools available that can quickly search a tagged corpus to quantify collocation use and provide association measures (Biber et al., 1994; Dazdarevic et al., 2015). The association measures are fed into the statistical frequency analyses, which constitute the next step, where a suite of statistical calculations are used to test the statistical properties of the collocational frequencies, and to determine which collocations found in the Arab EFLLs corpus are significantly different from those in the NBES corpus.

The first step in the research process comes from corpus linguistics and involves choosing the node word for the collocation search from a frequency list of the top 200 words for each word class. The second step is to use the log-likelihood (LL) statistic as an association measure for choosing the collocates of the node words, and then other analyses are carried out (see section 3.5.1)

3.1.2.2 Level 2 - The contrastive error analysis approach

Contrastive analysis is an approach that aims to show the structural differences and similarities between two languages (Johansson, 2008, p.9). I will use the contrastive analysis in this research because it helps to explain some of the lexical collocation errors in terms of L1 interference. A contrastive interlanguage analysis essentially means comparing the various ways in which native and non-native speakers of English differ in terms of linguistic background and foundation (Granger, 1998a). To achieve this, cluster and contrastive analyses underpinned by L1 interference and error analysis theories will be used. Therefore, I entitle this approach the '*contrastive error analysis approach'* because the contrastive analysis is fundamental to the error analysis. Cluster analysis is a statistical classification method used to group (cluster) a population into several groups (clusters) based on shared similarities, rather than on being identical (Moisl,

2015). This method is widespread in the field of corpus linguistics (Gries, 2010, p.22). It has demonstrated several advantages in picking out differences in frequency word lists between native and foreign speakers of English. This type of analysis is considered appropriate for the purpose of understanding the lexical and grammatical differences and similarities that exist between native English speakers and Arab foreign language learners of English. The clustering of these differences will give us insights regarding the major difficulties of Arab EFLLs when using English lexical collocations.

3.1.2.3 Level 2 extended – The descriptive-contextual approach

The descriptive-contextual approach is a fundamental part of the contrastive error analysis. The contrastive error analysis will include a *descriptive contextual analysis* of the collocations used by Arab EFLLs. The term *context* refers to the surrounding text of the collocations searched in the corpus (Storjohann, 2003, p.759). The *descriptive-contextual approach* concurrently combines two approaches: an intuition-based approach and a phraseological approach. I label it as the *descriptive-contextual approach* because both approaches within it adopt a descriptive approach to the context of language use (Cowie, 1998). Because linguists depend on their intuition when scanning concordance lines for patterns and examples of language use, McEnery and Hardie regard the intuitive approach as “non-statistical” (2011, p.126). However, the intuition-based approach reveals issues that are occasionally hard to verify with the corpus-based approach alone. Intuition, here, takes the form of researchers’ judgements of language use in terms of grammar and lexical meaning (McEnery and Hardie, 2011, p.209). Some authors refer to the intuition-based approach as introspection, where a researcher depends on his or her own intuition to judge language use (Nesselhauf, 2005; McEnery and Hardie, 2011; Willems, 2012). The intuitive approach is crucial as it provides a qualitative explanation of the data.

Intuition is used to identify instances of AEFLL collocations that might contain errors. For example, the corpus-based approach may identify and quantify the AEFLLs’ use of the English article system, such as *a*, *an*, *the* or the capitalisation

issues within the learners' essays. However, without the intuition-based approach, it will not be possible to explain the reasons for the relative lack of use of articles by Arab EFLLs (articles are considered an absent category in the Arabic language). I will use the intuitive approach to determine whether the Arab EFLLs' collocations seem acceptable to a native English user. The approach will require the use of other reference corpora and native English speakers. For this study, two British English native speakers will be recruited to judge Arab EFLL collocations that do not occur in the BAWE corpus or the BNC. The main impetus behind having native English-speaking judges is to have the data examined from a native-speaker perspective and to enhance the use of the intuitive-based approach (Wray, 2004).

The phraseological approach (which is more semantic, and focuses on fixedness and degrees of fixedness of word combinations that function as a unit) contradicts Firth's view of collocations (which is broader and more stochastic, being only concerned with habitual co-occurrences). Gyllstad (2012) wrote in his entry for the Encyclopaedia of Applied Linguistics. He neatly and succinctly defined the phraseological approach to collocations as an "approach focusing on collocations as word combinations governed by semantic aspects, such as idiomaticity and restricted substitutability" (2012, p.1). Idiomaticity refers to the phrases where their meaning cannot be inferred from the words literal meaning. Restricted substitutability means that some words or phrases cannot be easily replaced or modified within these units without altering their meaning or violating their grammatical structure. For example, 'break a leg' is a phrase that is commonly used to wish someone good luck. Therefore, the meaning cannot be inferred from the literal meaning of the single words. This example also has restricted substitutability, as replacing 'leg' with another part of the body, would result in a non-idiomatic phrase. Hunston (2002) views phraseology as an important approach to frequent words and especially to fixed phrases such as idioms. A well-known phraseological example is 'strong' vs. 'powerful' with the noun *tea*, where 'powerful tea' is not acceptable to native speakers (Halliday, 1976; Zernik, 1990). However, the use of these two adjectives differs when used with another noun such as 'computer'; a 'powerful computer' is the right expression, whereas a 'strong computer' is not acceptable (Duan and Qin, 2012).

Another example of a fixed phrase is ‘a piece of cake’. Without context and intuition, Arabic speakers might incorrectly interpret this literally as a portion of cake (Khuwaileh, 2000, p.102). The intended meaning can only be determined by the context in which it is used and the reader's knowledge of common idiomatic expressions in English. Phrasal verbs are another topic discussed in phraseology. In this thesis, phrasal verbs, idioms and fixed phrases are not part of the investigation but there are occasions where learners use a fixed phrase but use it incorrectly because of the influence of their L1. Therefore, describing collocational errors from a phraseological point of view can be helpful in identifying and explaining another source of errors, and this can be useful in teaching collocations. The native speakers’ intuition and knowledge of phraseology will be used to check whether the AEFLL collocations are acceptable or not.

Two phraseological aspects that will be used to determine the more phraseological type of collocations are *recurrence* and *usage restriction*. Recurrence is related to the work of Firth on collocation--his view that there is a tendency of words to co-occur, along with his most-cited definition “The habitual collocations in which words under study appear are quite simply the mere word accompaniment, the other-word material in which they are most commonly or most characteristically embedded” (1968, p.180). Usage restriction is the second aspect of identifying collocations from a phraseological perspective. Usage restriction is mostly associated with the work of Cowie who explained that there is a syntagmatic and paradigmatic relationship between the words in the collocation. The syntagmatic relation is related to the semantic meaning of collocations in which the meaning of one element is determined by the other element (Cowie, 1998). The paradigmatic meaning is related to the re-combinability of the two elements creating more fixed collocations, such as ‘kick the bucket’ (Cowie and Howarth, 1996). Therefore, the phraseological approach to EFLLs’ collocations use will be mainly concerned with the semantic and syntactic analysis of collocations.

Summary

In this research project, I aim to employ all four approaches outlined above in a balanced manner. I use the corpus data and the statistical measures already built into the web interfaces for the corpora used in this study, namely IntelliText and Sketch Engine. The first step in comparing how Arab EFLLs and NBES use lexical collocations will be to apply a corpus-based frequency approach to extract lexical collocations from the research corpora and to evaluate the statistical significance of the null hypothesis that there is no difference between the two groups. The lists of Arab EFLLs' collocations will be manually compared to those from the NBES corpus and the BNC. If the Arabic-speakers' collocations do not occur in either of these two corpora, they will be evaluated by native British-English speakers for their acceptability in English. Then, a contrastive error analysis will be carried out on the Arab EFLLs' collocations lists. The analysis will be descriptive, taking into account the context of the collocations, with some attempts, when necessary, to investigate some of the errors from a phraseological perspective.

The investigation of collocations through a quantitative approach alone will not yield results as meaningful as one based on linguistic understanding through the use of the two-level, mixed-methods approach. Greenbaum claims that "[a] more valuable, if more modest, contribution might be made to the study of collocations if a relatively homogeneous class of items were selected and an investigation undertaken of the collocation of each item in the class with other items that are related syntactically in a given way." (1970, p.13). Therefore, following the corpus-based frequency analysis, contrastive error analysis using the descriptive-contextual approach will allow for a more adequate explanation for the differences in use and the errors.

3.2 Sample and Sampling Design

In this section, I will discuss the samples used to investigate the use of written lexical collocations by Arab EFLLs. At the end of this chapter, in Section 3.6, I will also present an experimental study that was carried out to determine the corpus size effect on lexical collocation within the research corpora I used.

3.2.1 Corpus data

The primary focus of research in this study involves the analysis of four types of lexical collocation use by Arab English as a Foreign Language Learners (Arab EFLLs) compared against their use by Native British English Speakers (NBESs) who are also university students. The analyses of the collocations involve the use of corpora that have been designed to allow fair comparisons between native and non-native speakers of English in their use of written English. The type of corpus considered for the present study is a learner corpus, a type that has already been specifically developed and used to identify differences between native and foreign speakers of English (§2.1.3 and 2.1.4 for more explanation of the corpus).

The learners studied were Arab foreign learners of English from Kuwait and Dubai in the United Arab Emirates. The written data for these Arab EFLLs had already been collected by other researchers, and were made available to me. The counterparts to these Arab EFLLs are Native British English-speaking students, and their English is represented by the BAWE/NBES corpus (though other reference corpora will be used as well for some analyses). Table 1 summarises the data sources and the sizes of the different corpora.

(a) The Arab EFLLs Corpus:

- The Arab Learner English Corpus (ALEC) [17]
- The BUiD Arab Learner Corpus (BALC) [18].

[17] For more information about the ALEC corpus: https://dspace.auk.edu.kw/handle/11675/1757

[18] For more information about the BALC corpus: http://ucrel.lancs.ac.uk/publications/cl2009/54_FullPaper.doc

(b) The NBES corpus:

- The British Academic Written English Corpus (BAWE)[19].

Two existing corpora of Arab learners of English were combined to form the Arab EFLLs Corpus used in my research. The first corpus was the Arab Learner English Corpus (ALEC), which is a collection of four types of written essays by first-year university students in Kuwait whose first language is Arabic (the number of words for each type of essay can be seen in Table 3-1). Each type of essay served a different purpose but required careful consideration of the topic and effective use of evidence to communicate a message. I will briefly explain each type. Analysis essays focus on breaking down a subject or text to understand its meaning. Narrative essays tell a personal story with a message. Synthesis essays bring together multiple sources to make an argument. Argumentative essays present a specific viewpoint with evidence.

The second corpus that I will use for the Arab EFLLs is the British University in Dubai (BUiD) Arab Learner Corpus (BALC) (Randall and Groom, 2009), which consists of 1,865 written essays by two levels of learners who were either in their first year at university or in their last year of secondary school. The BALC corpus has 287,227 tokens and 20,275 word-types. The ALEC and the BALC corpora were combined and together formed my Arab EFLLs Corpus. The total size of the Arab EFLLs Corpus is 1,268,975 tokens and 883,141 word-types.

[19] For more information about BAWE corpus: https://warwick.ac.uk/fac/soc/al/research/collections/bawe

Type of corpora / Info	ALEC	BALC	BAWE (NBES)
Origin of Learners	Kuwait (70%) Egypt, Lebanon, Syria	Emirates	British English
Corpus Size	510,589	287,227	6,506,995
Types of essays	Analysis (184,749), Narrative (67,527), Synthesis (66,015), Argumentation (192,298)	Essays under examination	Narrative, case study, critique, design specification, empathy writing, essay, exercise, explanation, literature survey, methodology recount, problem questions, proposal, research report
Learners Level	Freshman level- first year university students (undergraduates)	- First-year university students (undergraduates) - Last year of high school students	Undergraduates Master's students
Teachers and type of curriculum	American teachers American model of higher education	- Most teachers are British, but they could be from other BANA countries (British, Australasian, and North American) or even fluent non-native speakers. - British published textbooks	British teachers

Table 3-1 Summary of information about the corpora selected for this PhD research project

The British Academic Written English Corpus (BAWE) will be used in this thesis to represent the written English of British students (NBES). The BAWE corpus contains 6,506,995 words, including 8,336,262 tokens and 6,968,089 word-types which were collected from the assignments of good-standard students. Token can be explained as the total counts of words or a word sequence in a corpus. Word type refers to a unique form or lexeme regardless of how many times it occurs in a corpus. The NBES corpus comprises of nearly 3000 high-quality student assignments. To guarantee a good standard of writing, only learners achieving grades of 60 or more (on the UK 0-100 system) were included in the corpus (Alsop and Nesi, 2009). The academic levels of these students were across four levels, from undergraduate to Master's level. The collection is well-balanced across four primary fields of study (Arts and Humanities, Social Sciences, Life Sciences, and Physical Sciences). To make my comparisons with the Arab EFLLs Corpus clearer, in my research I will refer to BAWE as the NBES corpus. Together, the Arab EFLLs and NBES corpora will allow for direct comparisons of language use.

Table 3-2 The learner corpora sizes

Corpora	Token	Type
Arab EFLLs Corpus	1,268,975	883,141
NBES corpus	8,336,262	6,968,089

Table 3-2 illustrates the number of tokens and types for the Arab EFLLs and the NBES corpora. The NBESs Corpus has a much larger stock of tokens and types than the Arab EFLLs Corpus. Although the lower number of tokens and types for the Arab EFLLs Corpus might be considered a limitation, it nevertheless provides sufficient data for this study, and I will provide evidence in a later section that the difference in corpus size did not affect the validity of my findings on collocations.

3.2.2 Access to existing corpora

To gain access to the data for both the AEFLL corpora, direct contact was made with the professors who worked on compiling them, Dr Inas Mahfouz for the ALEC corpus and Dr Mick Randall for the BALC corpus. The compilers kindly shared a

copy of their data with me, and I had the corpora uploaded to the Sketch Engine and IntelliText web interfaces.

The proficiency levels of the Arab learners of English that contributed to the Arab EFLLs Corpus were intermediate to upper-intermediate. Some of the learners were University students who had had enough experience in using English to have reached that level. The selection of such a group of students was also intended to ensure that the language had a high enough level of morpho-syntactic complexity so that the focus of research could be on the more discourse-oriented linguistic issues rather than just vocabulary or single words. The learners' levels in the BALC and the ALEC corpora that made up the Arab EFLLs Corpus[1], are detailed in table 3-3.

The decision to combine the data for the ALEC and BALC corpora into a single corpus, i.e., the Arab EFLLs Corpus, was made for the following reasons. First, the learners in both corpora are roughly of the same age. Second, the L1 for both is Arabic, and Standard Arabic is commonly spoken and taught across the UAE (Al-Mahrooqi et al., 2016). The two corpora are from the same region, the Arabian Gulf. The essays in both corpora are the learners' first drafts. Based on my contact with the compilers of the two Arab corpora, there are some differences between the two corpora that need to be acknowledged. The ALEC corpus was compiled from learners who wrote for their writing course assignments using a word processor, and they were able to use the spell-checker. In contrast, the BALC corpus consists of essays handwritten under examination conditions.

The data for the native English speakers are already accessible via the available web interfaces, such as Sketch Engine. I also requested that the BAWE corpus be uploaded to IntelliText (http://corpus.leeds.ac.uk/itweb/htdocs/Query.html#) for the purpose of this research, along with all the texts in the Arab EFLLs Corpus, so that there was a common interface for the research corpora.

[1] The AEFLL corpus created from combining the data for the BALC and ALEC

Table 3-3: Learners' levels in the Arab EFLLs Corpus

BALC	ALEC
- Various sources of teaching materials such as national textbooks used in the UAE as well as those from British publishers of standard intermediate ESOL courses. - The learners' levels vary from beginners to upper intermediate. - Final year of secondary school or first-year university level of education. - The data come from a single essay written under exam conditions. - The lengths of the essays are 150 words maximum.	- The University follows the American model of Higher Education. Courses are taught by American professors. - All Kuwaiti students are required to take an Accuplacer test to assess their English language proficiency level. The compilers used the Accuplacer placement test (https://accuplacer.collegeboard.org) to only include the work of students who scored 250 or more in reading and writing on the test. The meaning of the 250-band score is explained below (a screenshot taken from the Accuplacer Skills Insight[2]) All the collected essays are first drafts and vary in length. **Score range: 250–262** Students scoring in this band can typically demonstrate the following additional skills and knowledge in moderately challenging texts: ▪ Locate and interpret subtly stated information in and determine an implicit central idea, theme, or purpose of a text ▪ Determine a straightforward function of a part of a text in relation to the whole text ▪ Make somewhat challenging connections between multiple texts on the same topic ▪ Determine the meaning of a common high-utility academic word or phrase using context clues

In this section, I have explained the contents of the corpora, the size of each corpus, and the age and proficiency level of the learners for each corpus. The NBESs Corpus is significantly larger than the Arab EFLLs Corpus. The NBESs Corpus contains more than six million words, while the Arab EFLLs Corpus contains approximately 800,000 words. Based on the corpus size differences, I wanted to check whether there was any effect that could be attributed to corpus size. The results of the experimental study investigating the corpus size effect on collocations between the Arab EFLLs and the NBES corpora are presented in 3.6.

[2] More details can be found via this link: https://accuplacer.collegeboard.org/accuplacer/pdf/consolidated-skills-insight-statements.pdf

3.3 Tools

In this section, I outline the tools that were adopted for my research. There are two sorts of tools, (i) general research tools and (ii) corpus-based tools. For general research tools, I used software packages such as Microsoft Excel, SPSS and the Jisc online survey tool. The corpus-based tools used were web interfaces such as Sketch Engine and IntelliText.

3.3.1 General research tools

For this study, I will use some general office software and common research tools such as Excel spreadsheets, SPSS, and the Jisc online survey tool. I will provide an explanation of how these tools were used in the thesis in Section 3.5.

3.3.1.1 Microsoft Excel

Microsoft Excel is a spreadsheet system that I will use to document my data and their frequencies. Its spreadsheet tools were used throughout the study to log the data and arrange them based on the variables of interest. This software will help in creating tables to visual the data and performing some of the basic calculations such as those involving descriptive statistics (§3.4.2).

3.3.1.2 SPSS

The statistical software that I will use throughout the study will be the standard software Statistical Package for Social Science (SPSS). SPSS will be used to carry out statistical significance tests that cannot be performed through Microsoft Excel. For example, I will use to investigate the significance of the overall frequency of the collocations using tests such as the Mann-Whitney U test and the Kruskal-Wallis H test (§Explained in 3.4.3.2 and 3.4.3.3).

3.3.1.3 Jisc online survey tool

The Jisc online survey tool (formerly BOS), is a tool used in academic research for conducting surveys and questionnaires. I will use this tool in this thesis to

collect native speakers' judgments about the Arab EFLLs collocational choices. To do this, the collocations will be listed in a survey form and will be given three values to be rated on: either acceptable, questionable, or not acceptable (§see 3.5.2 for further explanation about the choice of the three levels of acceptability).

3.3.2 Corpus-based tools

In this section, I present the corpus-based research tools employed in this PhD thesis. As explained above, the available web interfaces that I will use for the corpus searches were mainly IntelliText and Sketch Engine. These two web interfaces were designed to allow researchers to build, access and exploit corpora. The University of Leeds IntelliText[3], referred to hereafter as IntelliText, is the result of an AHRC-funded research project carried out at the Centre for Translation Studies (CTS) in 2010-2011. The first version of the interface was released in 2011 and further functions have since been added. Sketch Engine is a language exploration tool developed for the English language (Kilgarriff et al., 2014). I will be using these two interfaces for the following reasons. Both online interfaces enable analysis through algorithms offering basic and complex computations, such as frequencies and association measures for collocations and other word combinations. Also, I used the IntelliText because of the built-in functions that enable me to extract the required data for my study including some functions that are not available in other interfaces such as the compare frequency tool which I have explained in (§3.3.2.2). Three commonly used corpus functions available through these web interfaces are frequency lists, collocations, and concordances. These three search functions typify the basic analytical tools of corpus linguistics that were used in this thesis.

3.3.2.1 The Keyword-in-Context (KWIC) concordance

The Keyword-in-Context (KWIC) concordance and collocation view functions were used to extract typical examples of language use (O'Donnell, 2008, p.108).

[3] The University of IntelliText web interface can be accessed via the link: (http://corpus.leeds.ac.uk/itweb/htdocs/Query.html)

A Keyword-in-Context (KWIC) concordance shows the search word in its immediate context. The concordance tool is a search function within Sketch Engine and IntelliText that I used to search for the words that are under study. IntelliText was particularly frequently chosen for this task, because it offers a concordance view that classifies the search results according to the Common European Framework of Reference for Languages (CEFR) word difficulty classification levels. This provides information for evaluating learners' collocation use by showing which words fall under which CEFR level. The concordance view was used to automatically generate examples of the selected searched-for word or 'node', the term used to refer to a searched-for word in a corpus, which is usually displayed in the centre of the concordance lines (which are the search results). Besides showing the node word, the concordance lines display the context that precedes and follows the node word.

The concordances still require a manual analysis to interpret the large amount of information displayed. Even frequency and collocation lists need manual editing or manipulation due to the number of unusual forms that can occur due to bugs in a corpus/interface (i.e. collocation and frequency lists need to be "cleaned" before they are fit for use in the analysis). For example, due to errors in the corpus or the interface used for extracting frequency and collocation lists, unusual forms of words may occur, often due to tagging issues. In the Arab EFLLs corpus, a few instances were removed as they contained errors or nonsensical characters like "&equo," "&mdash," and "&hellip" which stand for quotation marks, the em dash, and the horizontal ellipsis, respectively. Faaß and Eckart (2013) explained different methods for cleaning corpus lists, including identifying and removing duplicates. I used Microsoft Excel in this study to manipulate and further interrogate the data generated as it aids the labelling and annotating of the different variables of interest, which can then be output and loaded into statistical software packages such as SPSS. A concordance analysis is the preferred approach because it allows the qualitative examination of the collocations (Potts et al., 2015, p.154). Therefore, I used this tool in the contrastive error analysis of the Arab EFLLs' use of collocations. Also, the concordance lines that were generated from the searches were the data that were presented to the native-speaker raters (this will be discussed in 3.5.2).

3.3.2.2 Frequency lists and the compare frequency function

The *compare frequency* corpus function is an IntelliText function used to compare the frequencies of specific queries in a corpusor across different corpora. This can help the user analyse and compare how frequently certain words or phrases appear in different sets of texts (§Figure 3-1).

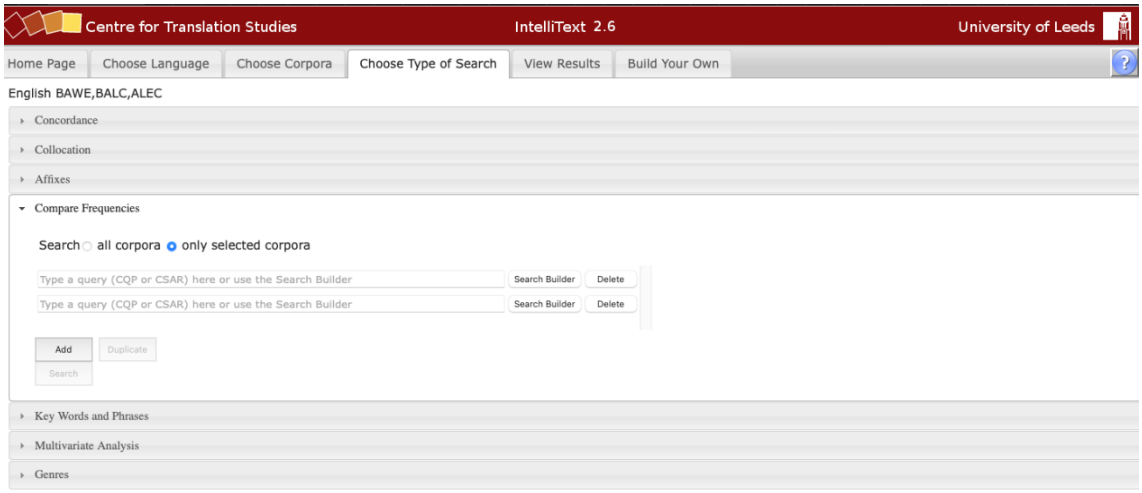


Figure 3-1: A screenshot of the *compare frequency* function in IntelliText

The function allows you to specify which corpora you want to compare and which queries (words or phrases) you want to search for. For example, the frequency of three collocations ‘strong coffee’ vs. ‘powerful coffee’ vs. ‘heavy coffee’ in the BNC and I-EN is shown in Figure 3-2. You can then view the results in a table or graph format, which shows the frequency of each query in each corpus, making it easy to compare and contrast the results (§Figure 3-2).

Corpus	strong coffee		powerful coffee		heavy coffee	
	Instances	IPM	Instances	IPM	Instances	IPM
BNC	19	0.171	0	0.000	0	0.000
NEWS-GB	46	0.210	0	0.000	3	0.014
INTERNET-EN	14	0.074	0	0.000	2	0.011
WIKI-EN	31	0.022	5	0.003	7	0.005
UKWAC	207	0.098	4	0.002	8	0.004
Totals	317	0.078	9	0.002	20	0.005

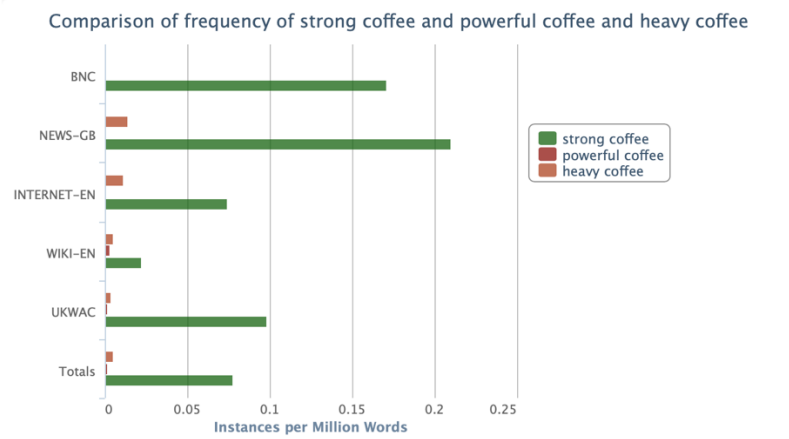


Figure 3-2: A screenshot of the output of the compare frequency function in IntelliText.

In Sketch Engine, the normalised frequencies of each word under the selected word classes were recorded. The output of collocation searches in IntelliText were arranged according to multiple association measures such as Mutual Information (MI), log-likelihood (LL), and *t*- score (§3.3.2.1).

The *frequency list* is a function used to list items based on their frequency of occurrence within a corpus. Frequency lists from unequally sized corpora can be made comparable by normalisation, with unnormalised frequencies being referred to as 'raw' or 'absolute' (Gries, 2010; Brezina, 2018b). Normalisation is usually expressed as ratios of instances per thousand or million words (Gries, 2010; Gries and Newman, 2017). The term 'relative frequency' is another synonym used in the literature for normalised frequency (Gries, 2010; Brezina, 2018b). The absolute (raw) frequency represents the actual number of occurrences of a word in a corpus, usually reported when working on its frequency in a single corpus. In this thesis, since the Arab EFLLs Corpus and the

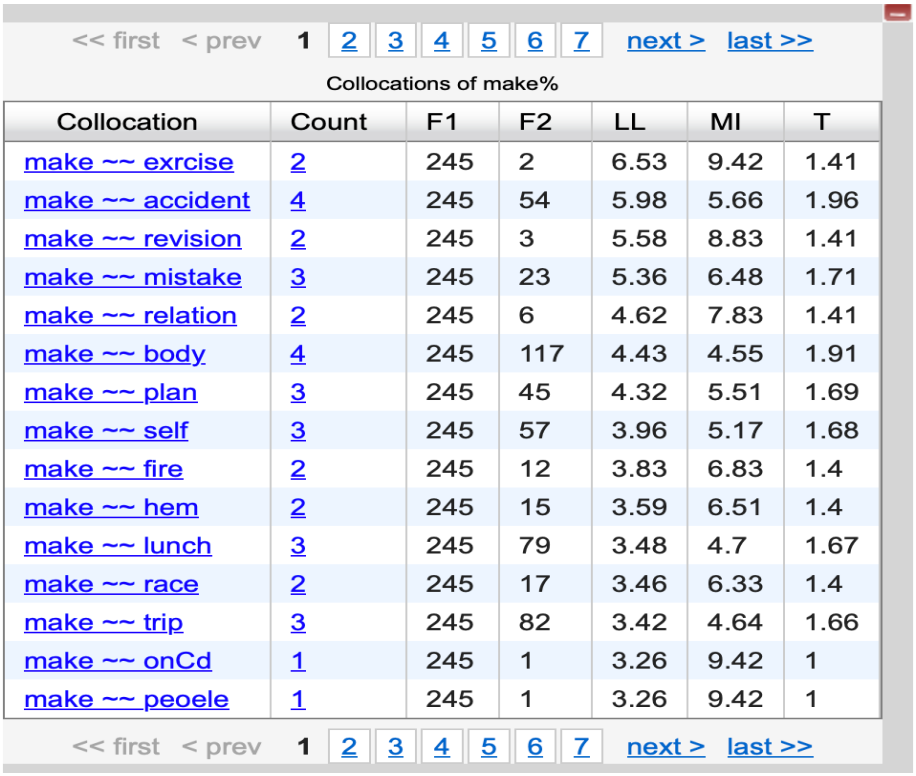
NBESs Corpus differ in size, relative/ normalised frequencies are also reported to allow fairer comparisons. One important aspect related to frequency is the word frequency list. Word frequencies are central to the frequency dictionaries designed to help learners: for example, Buckwalter and Parkinson (2011) for Arabic, Sharoff et al. (2013) for Russian and Leech et al. (2001) for English.

3.3.2.3 The collocation function

The *collocation* search function not only extracts collocations from a corpus, it also outputs statistical information about how strongly words are associated with each other or how strongly they tend to co-occur. The function allows different spans of word combinations to be called up, and thus is not limited to only two words next to each other.

Figure 3-3: A screenshot for the collocation query function taken from IntelliText

The screenshot in Figure 3-3 illustrates the collocation query function in IntelliText. The first box allows you to investigate the desired node words. For example, if I want to search for the possible noun collocates for the verb ‘make’, I will type into the box the word ‘make’ with the percentage symbol (i.e., ‘make%’). The percentage symbol then allows for a search of the possible forms of the verb ‘make’.



<< first < prev 1 2 3 4 5 6 7 next > last >>

Collocations of make%

Collocation	Count	F1	F2	LL	MI	T
make ~~ exrcise	2	245	2	6.53	9.42	1.41
make ~~ accident	4	245	54	5.98	5.66	1.96
make ~~ revision	2	245	3	5.58	8.83	1.41
make ~~ mistake	3	245	23	5.36	6.48	1.71
make ~~ relation	2	245	6	4.62	7.83	1.41
make ~~ body	4	245	117	4.43	4.55	1.91
make ~~ plan	3	245	45	4.32	5.51	1.69
make ~~ self	3	245	57	3.96	5.17	1.68
make ~~ fire	2	245	12	3.83	6.83	1.4
make ~~ hem	2	245	15	3.59	6.51	1.4
make ~~ lunch	3	245	79	3.48	4.7	1.67
make ~~ race	2	245	17	3.46	6.33	1.4
make ~~ trip	3	245	82	3.42	4.64	1.66
make ~~ onCd	1	245	1	3.26	9.42	1
make ~~ peoele	1	245	1	3.26	9.42	1

<< first < prev 1 2 3 4 5 6 7 next > last >>

Figure 3-4: Screenshot of the collocation query tool output in IntelliText.

Figure 3-4 illustrates the output of the search on the query 'make' in IntelliText. The first column shows the possible noun collocates for the query searched for, which is, in this case, the noun collocates of the verb *make*.[4] If I want to limit the node word into a specific content class, I will type in the box the following syntax [lemma='make'&tag!='V.*']. Here, 'make' is the node word, and 'tag' represents the selected word class. For example, 'V.*' would represent all verb forms. The collocates can be limited to only nouns by choosing the syntactic category from the PoS Editor button. The number in the second column shows the frequency of each node-collocate pair, i.e. the total count of the collocation in the given corpus. Column F1 gives the frequency count for the first word (the node), and F2 for the second word (i.e., the collocate of the node). The last three columns show the association measures of the collocations, such as the log-likelihood (LL), Mutual Information (MI), and *t* scores. An explanation of the available association measures is given in the following section.

[4] The odd spellings 'exrcise' 'peoele' are in the original learners' essays.

3.4 Statistical tests used in this thesis

In this sub-section, I will demonstrate how ordinal statistics are applied in collocation hypothesis testing (§3.4.1). Then, I provide an overview of the available collocation association measures and justify my choice of the primary association measure used in this study (§3.4.1.1). In the last two sub-sections, I will explain the statistical terms and tests used in this thesis: descriptive statistics (§3.4.2) and inferential statistics (§3.4.3).

3.4.1 Ordinal Statistics and choices made for the available collocation association measures

The frequency-based analysis in this thesis is based on ordinal statistics. Ordinal statistics are a type of statistics that analyse variables that are categorised based on a particular order. In this thesis, I investigate the use of collocation through hypothesis significance testing on ordinal collocation data. Ordinal statistical data is what we get after collocation extraction as the collocations are displayed in a ranked order, according to their log-likelihood score. Based on my knowledge of the field, there has been no discussion about the use of ordinal statistics for investigating collocations. Organising ordinal data means that collocations are arranged according to their log-likelihood score. Then, descriptive and inferential statistics are performed on these ordinal Log-Likelihood scores. However, prior to conducting the appropriate statistical significance tests on the lexical collocation data, the data distribution needs to be determined (see §3.4.3.1 for an explanation of data distribution). As mentioned earlier, the data in my thesis are lexical collocations. These lexical collocations are extracted and tabulated based on their log-likelihood test statistic. The log-likelihood (LL) test is a statistical association measure for collocation that predicts the probability of the two words occurring together, or the strength of a word association based on the frequency of co-occurrence between the node word and its collocate, as well as the frequency of each word occurring independently (Oakes, 1998; Brezina, 2018a). When the collocation extraction tool in IntelliText is used, the collocations are arranged according to their frequency or log-likelihood.

While studying collocations based solely on the raw frequency of the collocate is one option, association measures that rank collocates based on their association strength are better, because these association measures give finer, more nuanced information, enabling corpus linguists to evaluate the significance of the collocations more accurately (Evert et al., 2017, p.531). These collocation association measures are used to measure the strength of a combination of words using their occurrence and co-occurrence in a given corpus. The well-known association measures used are the *log-likelihood* (LL) score, *mutual information* (MI), *pointwise mutual information* (PMI), *t-score* and the *Fisher-Yates exact test* (Gries, 2010; Gries and Newman, 2017; Brezina, 2018b).

The MI is an association measure for the interdependency of two random variables. MI measures the joint probability of (A) and (B) and the probability of observing the two independently by chance (Church and Hanks, 1990, p.23). The PMI is mainly used for the analysis of “two-way co-occurrences” (Van de Cruys, 2011, p.16). According to Van de Cruys, the PMI is an association measure of two variables (A) and (B) and calculates the likelihood of “their co-occurrence given their joint distribution and the probability of their co-occurrence given the marginal distributions” of (A) and (B) assuming the two random variables are independent (2011, p.17). Thanopoulos considers PMI as a measure of “independence” (2002, p.621). Both the MI and PMI are association measures of co-occurrences in a corpus, but MI is used to measure the association of co-occurrence across the entire corpus, whereas the PMI measures the association in a specific context.

All collocational strength calculations take into account the following: (i) Number of word tokens in the corpus, (ii) Number of tokens of the node word in the corpus, (iii) Number of tokens of the collocate (on its own) in the corpus, and (iv) Number of tokens of the node and collocate co-occurring within a specified range/span. The log-likelihood ratio and the t-score are better association measures because they take frequency into account (Gries, 2010, p.14). The chi-squared statistic $\chi 2$ test is based on two assumptions, the normal distribution and the binominal

distribution, that cannot be applied to less-frequent words. Therefore, Dunning (1993) introduced an alternative, the log-likelihood ratio, also known as the G2 test. Having low-frequency words is a considerable issue when providing mutual information scores (Schmitt, 2012, p.5; Garner et al., 2018). Corpus size is a significant factor that must be taken into account prior to selecting the appropriate association measure: whereas the t-score depends on corpus size, the MI does not (Hunston, 2002).

The LL score is used to measure observed/actual number of co-occurrences is then compared to the expected-by-chance number of co-occurrences (i.e. what we would have expected by chance given the size of the corpus and the number of occurrences of the each of the two words as separate items). The different scores (t-score, LL, MI, etc.) all have different ways of calculating "expected by chance" (Evert, 2008, p.18), and that explains why they give different results when used to extract the collocates of a node word from a corpus. One of things they take into account is Zipf's law[5], and there are other considerations, such as how unevenly language items are distributed in any collection of texts (due to the word's part of speech, genre considerations, etc.), and how much weight should be given to rare words.

3.4.1.1 Justification for LL as the primary collocation association measure

Brezina's view is that there is "no best association measure" (2018a, p.100), and I agree that we should choose the collocation association measure that best serves the research objectives. For this current study, log-likelihood will serve as the primary extraction criterion because it serves the intended purpose of hypothesis testing. Through hypothesis testing, I am aiming to find out whether there is enough evidence in the data to reject the null hypothesis. One way to achieve this aim is through the use of log-likelihood figures that can simply provide a yes or no answer to the hypothesis being tested (Brezina et al., 2015, p.161). In addition, log-likelihood scores tell us the significance of a specific collocation based on its rank in a corpus; the LL statistical test is useful in ranking

[5] For more information about Zipf's Law: Gries, S.T. and Ellis, N.C. 2015. Statistical measures for usage-based linguistics. *Language Learning.* **65**(S1), pp.228-255.

the collocates of the node word. The test is also useful when comparing frequencies across two or more corpora. In this research, LL scores help us quickly and objectively identify all the significant collocational differences between the two corpora, something that would not otherwise be possible through manual observation alone. Log-likelihood is preferred over other statistical measures to avoid possible skewness from low-frequency words, as the LL ratio favours high-frequency words (Evert, 2010; Potts et al., 2015). What makes it special as a collocation association measure is that it is sensitive to both frequency and rarity of co-occurrence of the words. Additionally, the Log-likelihood score is known for being reliable and effective even for small datasets, making it a valuable tool for language analysis in various fields. This has been confirmed by Kunchukuttan and Damani (2008) who found that LL is a better association measure for considering collocations with very low-frequency counts.

The frequency threshold for the LL of collocations needs to be explained. After surveying the literature, there seems to be no agreement over the cut-off point for collocations extraction, and, as Biber and Barbieri (2007) explained it, their particular choice was ‘somehow arbitrary’. Setting the LL score threshold to (LL ≥ 5) is used by some scholars because this threshold assists in filtering out any random co-occurrences (Evert, 2005; Pecina, 2005; Potts et al., 2015). In this study, the frequency threshold was set to ≥ 5 because the data were taken from a learner corpus, which implies that there will be some instances of language use with low frequency.

3.4.2 Descriptive Statistics

The first sub-section below is about the descriptive statistics used for explaining the maximum and minimum values for the collocation LL scores and the standard deviation. Descriptive statistics can be explained as the measures used to collect, group, analyze and interpret results; it is the type of statistical calculation used in the descriptive analysis of the data (Byrne, 2007). Through descriptive statistics, one can justify issues within data using numbers.

3.4.2.1 Maximum and Minimum value

Maximum and minimum values are also reported in the collocation LL scores when using SPSS tool. The maximum value expresses the highest LL score of a collocation use, where the two words have the strongest collocation. The minimum score expresses the minimum LL score of the collocations. The minimum possible collocation LL score was set to (LL ≥ 5).

3.4.3 Inferential Statistics

In this section, I will explain the inferential statistical tests employed in this thesis. Inferential statistics are testing methods applied to make inferences or predictions of data (Byrne, 2007). Inferential statistics are used in hypothesis significance testing (Trafimow and MacDonald, 2017), which, in this thesis, is used as an important indicator to examine the differences in the collocation LL scores for the Arab EFLLs and the NBESs. Hypothesis testing sets a *p*-value to measure significance. The '*p'* stands for the probability of the results of a significance test ranging from 0 to 1, in which 0 suggests that there is a significant difference and the null hypothesis should be rejected, and 1 suggests there is not enough evidence to reject the null hypothesis. It is regarded as the "strength of evidence" for testing the statistical null (H_0) hypothesis (Ludbrook, 2013, p.357). Ludbrook (2013) argues that if the alternative hypothesis (H_1) simply asserts that the two groups' means or proportions differ, a two-sided *p*-value is appropriate.

3.4.3.1 Data distribution

Data distribution and normality testing are fundamental to determining an accurate inferential statistical test (Haden, 2019; Mishra et al., 2019). The norm for many statistical studies is to use histograms in their conclusion of the data distribution. Using histograms is one possible way to claim whether the data is normally/not normally distributed. The most important assumption of normality is that the means of independent samples are spread out normally. Usually, histograms are the first indicators for assessing normality used by researchers, the bell-shaped curve of the histogram indicating a normal distribution of the data.

However, since most of the data in this thesis did not follow a normal distribution, this indicated a need to apply a non-parametric test such as the Mann-Whitney U test and the Kruskal-Wallis H test (§3.4.3.1 and 3.4.3.3). These two tests each have their specific assumptions described in §3.4.3.2 and 3.4.3.3.

Rather than histograms, statisticians use skewness and kurtosis as two indicators for variable distributions, whether symmetrical (normal) or asymmetrical (non-normal) (Hair et al., 2017). Skewness is used to determine to what extent a variable distribution is normal/symmetrical. Kurtosis is used to measure the shape of a distribution if it is too peaked. If the results were zero for both skewness and kurtosis, then this would mean that the data had a normal distribution. However, the general role for determining the skewness of a distribution is the value □ 1, indicating a skewed distribution (Jones, 1969; Harlow, 1986). For kurtosis, the distribution is too peaked if the value is > +1. However, if the value is < -1, then this indicates a flat distribution.

Examining the shape of the data distribution is a fundamental step in choosing the accurate inferential statistical test (Brezina, 2018a). It has been assumed that parametric tests are the most powerful tests used in social science (Saldanha, 2009, p.5). However, with ordinal data and the application of statistical significance testing, using a parametric test is not possible for linguistic data. This is in line with other scholars who agree with the use of non-parametric tests for ordinal linguistic data (Sorace and Keller, 2005; Oakes, 2019).

In the field of statistics, there are unlimited statistical tests and listing them all would require much time and effort to explain them. Therefore, in the following section, I explain only the statistical tests used in this research.

3.4.3.2 Mann-Whitney U test

The Mann-Whitney U test is a non-parametric test used for data in two groups; it is the non-parametric test equivalent of the two-sample *t*-test. The Mann-Whitney U test assumes that data for the two groups do not have a normal distribution

which is an essential assumption in conducting this test (Gibbons and Chakraborti, 1991; Nachar, 2008). Four assumptions must be considered for conducting the Mann-Whitney U test. First, the data must be from two independent observations. Second, the two observations should not be related. Third, the type of observation should be ordinal. The fourth assumption is that the data are not normally distributed but must have a similar distribution. This assumption is crucial to conducting a Mann-Whitney U test. One way to assess whether this assumption holds is through applying Levene's Homogeneity of Variance test.

Levene's Homogeneity of Variance test is used to determine equality of variance between two groups. Some statistical tests have specific assumptions related to the data distribution and whether the data is homogenous, which means that both data sets have an equal distribution shape. Levene's test is considered an inferential statistical test that measures the "equality of variance" (Vergura et al., 2008, p.2). The Levene is a test that aims to determine whether the variances being analysed, the distribution of scores around the mean between two or more groups, are considered equal (Davis, 2010). Levene's test is used in this thesis to measure the assumption of the equality of variance for the collocation data to determine the accurate statistical tests. This test yields a *p*-value. If the *p*-value is > 0.05, the H_0 is rejected.

If the fourth assumption of equal variances in the Mann-Whitney U test is not satisfied based on a Levene test, an alternative approach may be to use the Independent Samples t-test with bootstrapping as an estimation technique. The Independent Samples t-test is a parametric test that compares the means of two independent groups (Gerald, 2018, p.50). The core concept of bootstrapping lies in the lack of additional information or prior knowledge about a population (Ahad et al., 2012, p.43). Therefore, this test would be the alternative to the Mann-Whitney U test when the assumption of homogeneity of variance is not satisfied. The advantage of using bootstrapping as a technique is that it allows for the estimation of the sampling distribution of a statistic without relying on the assumptions required by parametric tests.

Therefore, the Mann-Whitney U test can be used on the two collocation lists from the Arab EFLLs and NBES corpora, which are two data sets both ordinal and independent from each other. The fourth assumption was tested on each collocation set through Levene's test to determine whether the Mann-Whitney U test was applicable in investigating the hypotheses. Usually in a Mann-Whitney U test the null hypothesis is rejected if one group is significantly larger than the other (Nachar, 2008). In the Mann-Whitney U test, H_0 is that the median of the two data is not different. In non-parametric tests, the median is used instead of the mean.

3.4.3.3 Kruskal-Wallis H test

The Kruskal-Wallis H test is a non-parametric test, also described as the One-way ANOVA for testing the ranks of a sample of three or more groups (Kruskal and Wallis, 1952; Ostertagova et al., 2014; Verma and Abdel-Salam, 2019). The Kruskal-Wallis H test is also used when the data does not show a normal distribution. The assumptions of the Kruskal-Wallis H test are the same as those of the Mann-Whitney U test.

The One-way ANOVA on ranks is the analysis of variance test used to assess the fourth assumption to conduct the Kruskal-Wallis H test. It is also used to assess whether there is a statistically significant difference between the means of three or more independent data sets (Ostertagova and Ostertag, 2013). The One-way ANOVA on ranks was used in my study when I included data from three corpora to investigate the corpus size effect (§3.6.2.2) to assess the fourth assumption of the Kruskal-Wallis H test.

3.5 Quantitative and Qualitative Procedures

In this section, I will explain the procedures followed for investigating the use of lexical collocations in the research corpora. The two sub-sections below describe the procedures followed for the quantitative analysis (§3.5.1) and the qualitative analysis (§3.5.2).

3.5.1 The procedure for the quantitative analysis

Given the nature of this thesis, a statistical frequency analysis approach on the output from corpus-based searches is the initial stage for investigating the lexical collocations in the academic writing of language learners. This approach depends on choosing the right node words, the first word in the collocations, based on their frequencies. For example, for Verb-Noun collocations, I select a high-frequency verb and use the collocation tool to extract the possible noun collocates that are used with that verb in both the Arab EFLLs and NBES corpora. The results are an initial database of the most frequently occurring lexical collocations for both NBESs and Arab EFLLs.

The Arab EFLLs Corpus consists of nearly 800 thousand words in total. Due to this relatively small size, several steps will be followed to collect enough data. Accordingly, the first step consists of obtaining a list of the most frequent words of each word class in the study, i.e., *verbs*, *nouns*, *adjectives*, and *adverbs*. This step gathers the top 200 most frequent words of each word class, which are then compared against the BNC as the English reference corpus. Most of the 200 most frequent words were common to all three corpora (Arab EFLLs Corpus, NBESs Corpus, and the BNC). Several considerations led to the decision to limit the search to the top 200 most frequent words in the lemma form. First, there is the limited use of many words in the word lists in the Arab EFLLs Corpus. Thus, having the 200 most often occurring words will assist in obtaining enough collocations to analyse. Second, limiting the search to a general query for collocations within the corpus will lead to obtaining fewer collocations. For example, performing a general query for the two-word classes within the collocation types will produce a limited list of collocations. Therefore, I have taken the step of including the top 200 most frequent words in each word class. In the

literature, there are some variations in the initial criterion for selecting the top most frequent words. Shin and Nation (2008) focused on the most frequent 1,000 content words in spoken English and then on their top 50 collocations. Farooqui (2016) extracted the top 100 most frequent words and then searched for their possible collocates.

3.5.1.1 Collocation identification and extraction in this thesis

Two steps were followed to identify a collocation in the corpus. Using both the IntelliText and Sketch Engine web interfaces, the first step was to search for the node word using the lemma option, then to select the part of speech (PoS) that collocates with that node word to make concordances and frequency lists. To explain this, a collocation in this study comprises two components A + B, in which the node word is the first part of the collocation (i.e., Component A). The node word is referred to as the pivot word by some scholars (Shin and Nation, 2008; Balcı and Çakır, 2012). Frequency lists include a list of the top 200 most frequent words which were selected as the baseline for analysis. In this step, a reference corpus was used to determine the top 200 most frequent words in each word category (i.e., nouns, verbs, adverbs, and adjectives). These 200 top words represent Component A in the collocations.

The second step was to refine the collocation lists based on raw frequency co-occurrence and association strength thresholds within the corpus. In this case, the association strength threshold for collocation extraction in the corpus was set to a co-occurrence of five times or more as a threshold score (LL ≥ 5). This means that only words that have a minimum LL score of 5 had been included in the lists. Also, the raw frequency threshold is set to five occurrences (F ≥ 5). One of the benefits offered by using web interfaces is their ability to filter by frequency and association strength (F ≥ 5) automatically. Collocations that occur only once should be disregarded because it is the repetition of a collocation that makes it more noticeable and significant. Sinclair (2004) stated that the minimum co-occurrence to establish a language pattern is two.

To explain the extraction procedure, I take the *Verb-Noun* collocation as an example. The top 200 verbs in the BAWE were collected alongside their frequencies and compared against the 200 verbs in the BNC. The frequencies were recorded in Microsoft Excel Spreadsheet files based on their frequency in the corpora. Then the top 200 verbs were compared to the top verbs in the Arab EFLLs corpus. This step generated the collocation lists that were used to investigate how the lexical collocation is used by Arab EFLLs. The main impetus behind this step was to increase the procedure's reliability. This approach of analysing the top most frequently-used words in word classes (e.g. verbs) has been commonly followed by some scholars, such as Doró (2007), Gardner and Davies (2007), Liu (2011) and Zhou (2016). Also, comparing the Arab EFLLs top 200 verbs to the BAWE list was important as it showed that Arab EFLLs' use differs from that of NBESs. The differences between the two were removed and I selected a number of verbs over the top 200 that must occur in top 200 in the BNC. Then, the noun collocates were extracted based on these top 200 verbs. This procedure was repeated for the remaining collocations in this study.

3.5.1.2 Statistical analysis of the collocations

The collocations were collected and tabulated in Excel spreadsheets. The first step was to quantify the number of collocations based on their range of occurrences. Then, the LL scores for the listed collocations were entered into SPSS. The analysis provided some visualization of the data that explained the data distribution so that the right statistical test, such as box plots and histograms, could be selected. Based on the data distribution for each collocation, the appropriate statistical significance testing was then carried out to investigate the formulated hypothesis.

3.5.2 The procedure for the qualitative analysis

The second part of the analysis focuses on the errors present in the collocations used by the Arab EFLLs. The steps followed in the application of Contrastive Error Analysis are laid out in Figure 3-5.

Steps	Procedure
Step 1: Data collection	**Collect lists of** - Adverb - Adjective collocations - Verb - Noun collocations - Verb - Adverb collocations - Adjective - Noun collocations
Step 2: Error identification based on native judgments of a sample of the collocation lists (i.e., classifying the errors into types):	**A measure of collocation acceptability (Nesselhauf, 2005).** - Acceptable - Questionable - Not acceptable **It must be noted that not all errors are collocation errors there** **Error classification into categories[6]:** - Grammatical errors (e.g. the use of incorrect tense of one item in the collocation). - Lexical and Semantic errors (e.g. the meaning is accurately conveyed). The use of appropiate lexical item to convey the meaning accurately. - Spelling errors (e.g. including typographical and orthographic errors)
Step 3: Analyse the source of errors.	**What is the cause of errors? (e.g. differentiation, new category, absent category, omission, or addition)** - L1 interference errors which are errors that are caused by the interferneces of the learners' first language when using the second language. - There are two types of errors: interlingual or intralingual errors.
Step 4: Error quantification through frequency analysis	Frequency analysis of the number of errors.

Figure 3-5: The combined Contrastive Error Analysis framework used in this thesis (adapted from Gass and Selinker, (2008))

The contrastive error analysis framework used in this thesis was adapted and modified from Gass and Selinker, (2008) and consisted of four steps (§Figure

[6]The errors, although not collocation errors, they are being made within the collocations (e.g., spelling errors, for example errors in using compound words when used in the collocation where learners tend to split the compound words into two different words such as 'break fast'). In addition, some of the collocations were correctly used but were placed in an incorrect position in the sentence (e.g., 'Very wonderful my summer holiday').

3-5). The procedure followed for this approach was to create lists for the Adverb - Adjective, Verb - Noun, Verb - Adverb and Adjective - Noun collocations in the Arab EFLLs Corpus. The second step required manual error identification to determine whether the errors are grammatical, such as incorrect tense or semantic errors, and whether the meaning was accurately conveyed in the collocations. This second step relied on English corpora and native-speaker intuitive judgment for the measurement of collocation acceptability, meaning which collocations are acceptable or "permissible to native speakers of the language" (Lau et al., 2017, p.1218). For the second source of acceptability, two native English speaker raters from Britain (a male and a female, and of similar age to the learners being investigated) were consulted to determine the acceptable collocation patterns. The choice of raters in this study may enhance the reliability and validity of the results, as their familiarity with the language and their academic background could contribute to a more accurate assessment of acceptability. Having two raters enhanced the reliability and validity of this study, following many earlier studies that also used two native-speaker raters to investigate collocation acceptability (Nesselhauf, 2005; Futagi et al., 2008; Chang, 2018; Ding and Reynolds, 2019). The collocation lists were compared with an English reference corpus and the NBESs Corpus. When a collocation did not occur or occurred less than twice in both corpora, they were shown to native speakers of English for further judgments of acceptability. Learners' collocation use was then tabulated and analysed based on the degree of acceptability. The scale of acceptability adopted was similar to that suggested by Nesselhauf (2005) with a slight difference in the notations for the degrees of acceptability.

*	Clearly Unacceptable
(*)	Largely unacceptable
?	Questionable
(+)	Largely Acceptable
+	Clearly acceptable

Figure 3.4: Degrees of determining the acceptability of collocations (taken from Nesselhauf (2005))

Nesselhauf (2005) suggested five levels of acceptability, as presented in figure 3.4. However, I believe having an acceptability measure of three levels is enough to convey the acceptability rate adequately. In the literature, binary judgments, including yes or no questions, have disadvantages, as choosing one value over another can have a large impact on the results. Therefore, I allowed for another degree of acceptability, the “questionable”, which was also used in Nesselhauf's (2005) study. The main advantage of having a three-level acceptability scale is that it makes the judgment process natural and straightforward without having the raters decide if a particular sub-level is the more suitable (Dąbrowska, 2010). Therefore, setting three levels of acceptability made it easier for the native raters to judge the data and for the researcher to obtain a gist of the native speakers’ judgments about Arab EFLLs' use of collocation. The three-level scale for measuring acceptability, known as the 3-point Likert scale, allows for more satisfactory quantitative prediction, and for researchers to determine which of the three describes the acceptability of the collocation with high fidelity (Lau et al., 2017).

It is important to acknowledge that a three-level system may not capture the nuances of acceptability as effectively as the five-level system used in Nesselhauf’s study. Future research could consider conducting a sensitivity analysis to determine the extent to which the results may differ if a five-level system had been used. Overall, the choice of rating system should be carefully considered based on the research question, context, and the intended use of the results.

The second step of the Contrastive Error analysis framework consisted of classifying the errors into categories (§Figure 3-5). The higher the number of errors in a category, the more important it is for the researcher to investigate the issue. Then, the next stage was to examine the potential causes of these errors, whether caused by the L1 or other factors. The final step was to provide a descriptive frequency analysis of the errors.

Before giving the data to the native speaker raters, it was necessary to implement a sampling approach because of the high number of collocations in the data under investigation. There were 146 examples of the Verb-Noun collocation, 666 examples of the Adjective-Noun collocation, and 144 examples of the Verb-Adverb Collocation. Therefore, it would not have been possible to present all those collocations to the native-speaker language raters, as it would cost too much time and effort. Also, it could have had an adverse effect on the accuracy of the study, as the raters would have got tired (§4.2.3.1 for the sample size).

Regarding the procedure followed for error classification, I used mainly two approaches for classifying errors. First, I used the grammatical approach that aims to investigate the use of collocation from the perspective of grammatical accuracy. Arab EFLLs tend to need help with the use of the right tense and form in English (Chaaraoui, 2017). Second, I used the lexical approach that assumes that the meaning of a collocation is determined by the co-occurrence of the two words in the collocation. Sinclair stated that "There is no assumption that meaning attaches only to the word. It is anticipated that meaning also arises from the loose and varying co-occurrences of several words, not necessarily next to each other" (1996, p.99). In this approach, lexis is viewed as the central element in creating meaning, not grammar. This approach is mostly associated with the work of Firth (1957). For example, the lexical error is seen in the choice of lexis within the collocations, such as 'handsome man' could be said by EFLs as '*beautiful man'. While 'beautiful man' occurs 11 times in the BNC and 17 times in the BNC2014, it is deemed unacceptable by my raters' native judgement. It is commonly argued that 'man' collocates with 'handsome', while 'beautiful' is typically associated with 'woman'. Skitina et al. (2017) claims that the use of 'beautiful man' occurs more often in colloquial speech (§the Arabic use of this example has been discussed in 2.3.1). Another measure within error analysis is the classification of semantic errors, which is a part of the lexical approach in that it separates grammar from semantics in its analysis. The semantic errors aims to find out why certain words tend to co-occur with certain other words. For example, the collocations 'hot food' is often incorrectly used to mean 'chilli food'. Semantics is a factor in collocation. The lexical approach requires a meaning to be generated between the co-occurring lexemes. An analysis based on this approach might contribute to

understanding L1 interference or other intralingual factors that could affect the use of English collocations by Arab EFLLs.

Summary

Based on the previous literature, I found that it is useful to follow a systematic process. I initially looked for any evidence of collocation use from the concordances of my selected words (taken from the lists of the most frequent words in the language). The next step was to systematically identify the collocations using software tools, then classify the collected collocations and group them based on significant features (i.e., that explain how words are strongly associated in a combination using LL scores). The final step was to interpret the patterns found based on the previous step. So, the data were gathered based on what was displayed in the KWIC view through the web interfaces, and then Microsoft Excel spreadsheets were used to arrange the output of the concordance lines of interest. The concordance tool was used to search for the chosen word combination or collocation pair. Statistical association measures were automatically generated for the word combinations, and then Microsoft Excel was used to create tables and to perform some of the mathematical calculations required. The data tabulated in the Excel Spreadsheets were loaded into SPSS to conduct the statistical significance tests.

To conclude, in Section 3.5, I have explained the research procedure and how the research tools were employed. In the following section, I will present the results for an exploratory study for investigating if there is corpus size effect because of the different sizes of the two corpora (the AEFFLs and NBES corpora).

3.6 Exploratory Study: Determining corpora size effect on the collocation patterns in the Arab EFLLs and the NBES corpora

In this thesis, to compare the use of English lexical collocations, I use two corpora that differ in size. The AEFFLs corpus, containing the data of the ALEC and BALC corpora combined, is around 800,000 words in size, whereas the NBESs Corpus is much larger, at 6.5 million words, more than 8 times the size of the Arab EFLLs corpus. Due to this imbalance, additional analyses were done to ensure the validity of the data analysis and interpretation. Frequencies can be normalised to account for the differences in corpus size, but the overall corpus size effect is another phenomenon that needs to be account for. Another impetus for this investigation goes back to differences in the initial results in which the log-likelihood results for the NBES collocations were consistently higher than those for the Arab EFLLs, which could be because of the relative imbalance in the sizes of the two corpora.

The main hypotheses tested for determining the corpus size effect are the following:

> **Null Hypothesis (H_0):** Corpus size does not affect the collocation patterns in the Arab EFLLs and NBES corpora.

> **Alternative Hypothesis (H_A):** Corpus size affects collocation patterns in the Arab EFLLs and NBES corpora.

The following section explains the method and results for determining the corpus size effect.

3.6.1 Method for determining corpus size effect

The method for determining corpus size effect involves the selection of a small British corpus similar in size to the Arab EFLLs corpus to compare against the NBES (BAWE) corpus. The British English 2006 corpus (BE06), a one-million-word corpus, was chosen for this reason. The rationale for this exploratory study of the size effect is this: If the collocation patterns for "BE06 corpus versus NBESs Corpus" (both by native speakers of British English) are similar to each other and differ from the Arab EFLLs corpus, then it justifies my use of the large NBESs Corpus as a comparison corpus against the smaller Arab EFLLs corpus, and my research results are safe from "contamination" by any possible corpus size effects. In order to examine the most frequently occurring verbs in English, I used the Wordlist tool in Sketch Engine to search the verbs in the BE06. I chose five verbs from the top 10 most frequent verbs in the BE06, which were MAKE, HAVE, DO, GET, and TAKE. Then, the noun collocates for these five verbs were extracted and analysed, with the frequency threshold set to f ≥ 5 (meaning the frequency of each noun collocate has to be equal to or greater than five occurrences).

To summarize what has been explained above, the following steps were followed to determine the size effect of the corpora on collocation use.

Step One: A small native-speaker English corpus, in this case, the BE06 was chosen to match the size of the NBESs Corpus.

Step Two: The collocations in the three corpora (BE06, NBES, and Arab EFLLs) were searched for the verbs MAKE, HAVE, DO, GET, and TAKE.

Step Three: The first 20 instances of noun collocates (*f* ≥ *5*) for each verb were selected.

Step Four: The output of the BE06 was compared to the results of the NBES and the Arab EFLLs.

Step Five: A table was created to summarise and give an overview of the collocations of the five verbs within the three corpora.

3.6.2 Results for exploratory study of corpus size effect

This section presents the results for the study of whether there is corpus size effect on collocations. I will first report the descriptive statistics of the data (i.e., the results of the frequency approach), and then the inferential statistics of the data (i.e., Kruskal-Wallis H test results).

3.6.2.1 Descriptive statistics results (Frequency approach)

In this section, I present the results of the corpus-based frequency approach for the noun collocates of the five listed verbs extracted. I will report some descriptive statistics of the extracted collocations, such as quantifications and differences in the five collocation sets among the three corpora. First Table 3-4 illustrates the normalised/relative frequencies of the five selected verbs.

Table 3-4. The relative frequency (per million words) for the five verbs within the three corpora

Verb	BE06 (ipmw)	Arab EFLLs (ipmw)	NBES (ipmw)
MAKE	1868	3208	1736
HAVE	11589	10865	8273
DO	4758	6192	2470
GET	1855	2527	355
TAKE	1546	2303	1227

The frequency counts of the five verbs were higher for the Arab EFLLs than either the BE06 or the NBES in four cases, the exception being the verb HAVE, which was still higher for the Arab EFLLs than the NBES, but not higher than BE06. The Arab EFLLs' very frequent use of these top verbs therefore seems to be generally

higher than that of native speakers in NBES and the BE06, which suggests that Arab EFLLs might overuse these basic verbs over other verbs.

The output for determining the size effect is given in Table A-1 to Table A-5 in the Appendices. These tables present the extracted noun collocates for the five verbs within the three corpora. The initial result from Table 3-6 below gives the collocation instances (tokens). The data for the verbs MAKE and HAVE within the three corpora illustrates that there were 20 collocation instances within the set threshold (i.e., NF counts of ($f \geq 5$)). For the verb DO, there were 20 collocation instances in BE06, 9 in NBES, and 16 in Arab EFLLs within the assigned threshold. As for the verb GET, there were 20 collocation instances in BE06, 20 in NBES, and 9 in Arab EFLLs corpus within the set threshold. The verb TAKE had 20 collocation instances in BE06, 17 in NBES, and 17 in Arab EFLLs. It must be mentioned that, mainly in the BE06 and the Arab EFLLs corpora, some examples were taken from below the threshold to increase the number of samples for each corpus. The reason for there not being more examples in BE06 and Arab EFLLs could be due to the small corpus size for both data sets.

3.6.2.2 Inferential Statistics Results

Table 3-6 reports the results for the five sets of collocations. The data for all three corpora did not have a normal distribution, which explains why a non-parametric test was chosen to investigate the differences between the three corpora (§Figure A-1 to Figure A-5 in Appendix II). The Kruskal-Wallis H test is the most suitable test for analysis of this data type and for investigating the size effect of the research corpora. To further determine whether the five collocation data sets met all the four assumptions of the Kruskal-Wallis H test, another statistical test needed to be performed. The four assumptions are that the data i) are from three independent groups, ii) have no relation between the variables, iii) display non-normal distribution, and iv) most importantly, should display a similar manner of distribution even though it is not normal. The data met the first three assumptions, but to check whether the data satisfy the fourth assumption, I analysed the variances through the One-way ANOVA on ranks test. The One-way ANOVA on

ranks is used to determine the equality of means for three or more datasets (Ostertagova and Ostertag, 2013). The results are given in Table 3-5, which shows the F-value and the significance *p*-value. The F-value is used to explain the variance between the groups' means, and the Sig. value (is significance *p*-value) is the evidence measure used against the null hypothesis. The usual threshold for the p-value is set at 0.05, indicating that the null hypothesis is rejected if the calculated p-value is lower than this threshold. On the other hand, the F-value in ANOVA measures the variation between the means of multiple groups relative to the variation within each group. A threshold F-value of 1 or less indicates that the difference in variation between the groups is not significant compared to the variation within each group. The interpretation of this test results requires analysing the results of the p-value.

Table 3-5: One-way ANOVA for the five collocation sets

Collocation	F	Sig.
Make_Noun	0.605	0.549
Have_Noun	0.265	0.775
Do_Noun	1.353	0.269
Get_Noun	1.960	0.153
Take_Noun	0.009	0.991

It can be concluded from the table that all five collocations are all larger than 0.05, satisfy the assumption of homogeneity and fail to reject the $H_0 > 0.05$ *p*-value that would indicate that all the samples have the same mean. Therefore, the Kruskal-Wallis H test was suitable to be used as a test for statistically significant differences between the five collocations within the three corpora. This test was conducted using SPSS and the results are tabulated in Table 3-6.

The use of collocations in both native-speaker corpora, BE06 and NBES, were higher than the Arab EFLLs in the results for the verbs MAKE and HAVE. Table 3-6 illustrates that the Arab EFLLs corpus has higher frequency counts than the BE06 and NBES corpora for the verbs *make* and *have,* accounting for 49.82 and

70.20, respectively. However, there is not much difference between the highest and lowest scores.

Table 3-6: Statistical significance testing for the five verbs (*MAKE, HAVE, DO, GET, and TAKE*) in the BE06, BAWE and Arab EFLLs corpora

Corpora/Verb	Stats		BE06	NBES	Arab EFLLs
MAKE_N_collocations	Number of instances		20	20	20
	Minimum		4.00	3.08	4.53
	Maximum		24.00	39.85	49.82
	Median rank		33.58	30.58	27.35
	Kruskal-Wallis H Test	K-W H	1.273		
		df	2		
		Asymp. Sig.	0.529		
HAVE_N_collocations	Number of instances		20	20	20
	Minimum		4.00	5.23	4.53
	Maximum		44.00	68.15	70.20
	Median rank		25.2	33.25	33.05
	Kruskal-Wallis H Test	K-W H	2.767		
		df	2		
		Asymp. Sig.	0.251		
DO_N_collocations	Number of instances		20	9	16
	Minimum		4.00	4.00	4.53
	Maximum		49.00	22.00	27.18
	Median rank		27.10	21.33	18.81
	Kruskal-Wallis H Test	K-W H	3.749		
		df	2		
		Asymp. Sig.	0.153		
GET_N_collocations	Number of instances		20	20	9
	Minimum		4.00	0.62	4.53
	Maximum		16.00	4.46	10.19
	Median rank		35.85	10.70	32.67
	Kruskal-Wallis H Test	K-W H	34.282		
		df	2		
		Asymp. Sig.	0.000		
TAKE_N_collocations	Number of instances		20	17	17
	Minimum		4.00	2.31	4.53
	Maximum		103.00	28.62	22.65
	Median rank		30.45	35.65	25.40
	Kruskal-Wallis Test	K-W H	3.449		
		df	2		
		Asymp. Sig.	0.178		

The results for the BE06 and NBES corpora are relatively similar in terms of the collocation relative frequencies. The NBES has relatively more collocations than the BE06; however, the number of instances for the NBES remains low considering the corpus size. Therefore, the findings of this comparison show

similar results in BE06 as in NBES, suggesting that native speakers use more collocations.

The Kruskal-Wallis H test results for the verb GET showed a statistically significant difference in collocation frequency between the three corpora (X^2= 34.282, p = 0.000), with a median rank score of 35.85 for BE06, 10.70 for NBES and 32.67 for Arab EFLLs. Despite the Kruskal-Wallis H test showing a statistically significant difference, the median rank for the Arab EFLLs and the BE06 were similar, at 35.85 and 32.76, respectively.

However, the Kruskal-Wallis H test results showed no statistically significant difference in collocation use between the three corpora for the remaining four verbs: *MAKE*, *HAVE*, *DO,* and *TAKE.*

3.6.3 Concluding remarks on the experimental study

The difference in collocation use between native and non-native speakers is further reinforced by the comparison with a second native corpus that is similar to the Arab EFLLs corpus size. The results above suggest that the different corpus sizes do not affect the LL scores of use of collocations. The two native-speaker corpora, the smaller BE06 and the much larger NBESs Corpus, have similar collocational frequency profiles (native-speaker English students use more collocations in their writing than Arab EFLLs). Additionally, the hypothesis test indicated that four out of the five selected verbs did not show statistical significance in the results. Therefore, it can be concluded from the statistical analyses that corpus size is not a factor in the differences between Arab EFLLs and NBES. It must be mentioned that some collocation results collected in this investigation will be further used and analysed in the error analysis section of the results chapter (§4.2.3.2).

3.7 Conclusion

This chapter has described the methodology adopted for this thesis. The reasons for choosing the main approaches adopted, the corpus-based frequency approach and the contrastive error analysis approach have been given. The statistical tests performed to measure the associations in collocations for the two groups of English users have been described. In the next chapter, I present an exploratory data analysis of native English speakers to check for potential differences for both English varieties between British and American English before using the I-EN corpus as a second reference corpus.

Chapter 4 Results

4.1 A Corpus-based frequency analysis comparing collocation use in British and American English

Introduction

Having presented the methodology of this research project in the previous chapter, in this chapter, before presenting the results of the research project, I will present an analysis of a comparative study of the use of lexical collocations in British English (BE) and American English (AE). This comparative study was carried out before conducting the comparative study between Arab EFLLs and NBESs to evaluate the feasibility of the research design and methodology, identify potential problems, refine the research questions and hypotheses, and provide a basis for understanding the context of the research. It was important for determining the robustness of the research method and for ensuring that the results obtained are reliable and valid. The primary objective was to establish the comparability of the AE and BE data. If no significant differences were found, I planned to use the I-EN as the second reference corpus, as it is more contemporary.

To test the possibility of merging the British and American English native-speaker corpus texts, the following procedure was applied to determine the reliability and validity of the methodology implemented. Two comparison corpora were used in this investigation, the British National Corpus (BNC) and Internet Texts in English (I-EN). The I-EN corpus is generally thought to be predominantly composed of American English texts. To verify this claim, I selected two words with different spellings in BE and AE: ‘center/centre’ and ‘colour/color’. Frequencies of these spellings in the BNC and I-EN were compared, and I also compared the spellings

in the BE06 corpus and AME06[1] corpus as well, to add validity to the approach being taken. The normalised frequencies in the four corpora are provided in Table 4-1.

Table 4-1: Normalised Frequency Comparison Results: BE vs. AE Usage in the Selected Corpora

	BNC	I-EN	BE06	AME06
Color (AE)	0.48	48.28	0.86	72.30
Colour (BE)	95.03	18.72	75.32	0
Center (AE)	0.74	65.827	6.06	88.46
Centre (BE)	143.258	36.94	129.86	1.70

The frequency comparison revealed that the BNC versus I-EN comparison yielded similar results to BE06 versus AME06. The American spellings in the BNC and the BE06 were low based on their normalised frequency. However, the American spellings in the I-EN and the AME06 were higher than the British spellings. Based on these findings, I concluded that the I-EN corpus is predominantly written in American English. The BNC and I-EN corpora were specifically chosen to represent the two varieties of English because of their large sizes: 100 million words and 150 million words respectively. It is now necessary to formulate the hypotheses which will be considered in this comparison:

- **Null hypothesis (H_0)**

  There are no significant differences observed between BE and AE in the use of lexical collocations.

- **Alternative hypothesis (H_A)**

  There are significant differences observed between BE and AE in the use of lexical collocations.

In the following section, I will provide a short review of the literature on the use of collocations in BE and AE.

[1] The AME06 is the American English corpus from 2006 and can be accessed via IntelliText: http://corpus.leeds.ac.uk/itweb/htdocs/Query.html#

4.1.1 Previous literature on the differences between BE and AE in the use of collocations

Greenbaum (1974) investigated differences between BE and AE by focusing on verb-intensifier collocations. He performed experiments on university-level students from the United Kingdom (UK) and the United States (US). These experiments were designed to trigger the most-used verbs by giving the participants sentences that included the following six headwords, 'badly', 'entirely', 'greatly', 'utterly', 'completely', and 'very much'. These sentences then had to be completed with a verb. The results in Greenbaum's paper in which all the eight collocations identified in the British tests were also found in the American test indicate that there were remarkable similarities between the two varieties. The most notable example was seen in the word 'entirely' which collocated with the verb 'agree'. Therefore, 'entirely' was mainly associated with agreement and disagreement for both groups. The only difference noted was observed in the use of 'entirely' and 'completely' by BE students due to some distinctions between the two intensifiers in the UK data, a difference that was not clear in the US data. In the UK, the word 'entirely' usually collocated with verbs that indicate agreement whereas *completely* collocated with verbs that indicated a failure to get to a specified aim or purpose. The US data was different for 'entirely' as it collocated with 'forget', and with verbs that indicated both agreement and disagreement (1974, 86). Greenbaum's findings suggested that there was not much difference between the two varieties unless there was a semantic difference in one of the words in the collocation. As a result, the only considerable difference in the use of collocation between BE and AE was in collocates for which the node has a different meaning in the two varieties. In a more recent study, Deáková (2013) aimed to determine and explain the lexical convergence between BE and AE. She found that when there is a difference in the meanings of a word, there was bound to be a difference in the use of collocations. She pointed out that the differences were not restricted to word choices, as the two populations of speakers were "also of a different outlook and culture" (Deáková, 2013, p.52). Her research depended on the BNC and the Corpus of Contemporary American English (COCA), choices which posed some limitations on her findings. Concerning these, although the BNC text were collected between 1970 and 1994, Deáková claimed that there could still be some

sort of comparability to current BE with respect to some of the selected adjectives such as ‘cute’ and ‘cool’. A previous work by Darragh (2000) stated that there are 4,000 frequent and common words in BE that could not be found in AE or are used differently. This is in line with Deáková’s (2013) finding that there are differences between BE and AE in the meanings of particular words. Overall, these studies suggest that when there are differences between BE and AE in the meaning of particular words, there are likely also to be differences in the use of collocations. For example, Trudgill and Hannah (2013, p.90) showed how the word ‘homely’ has different meanings in British and American English. In BE, it means ‘down to earth’ whereas in AE, it means ‘ugly’. I must point out that the findings above are exceptional cases that do not form a consensus and that for these scholars, collocation use appears to be broadly similar in both of the two English varieties.

4.1.2 Methods

The method of this pilot study was designed to investigate three sets of collocations: Verb + Noun, Noun + Noun, and Adjective + Noun. The initial step was to identify the most common/frequent words for the verbs, nouns, and adjectives, and to create a frequency list of them. The words selected for each part of speech will be the node words for the collocation searches. The observed/raw frequency of each word was recorded in a separate table at the beginning of the results section for each set of collocations. Using the collocation tool in the web interface, the 1R (one word to the right) noun collocates for these word categories were then searched for in the corpora. Then, the top twenty frequent collocates were counted for each corpus so that the differences and similarities could be analysed. The query outputs were arranged based on their log-likelihood scores in Intellitext. When choosing the top twenty, some examples had to be deleted because they were typos or meaningless words such as *&equo*, *&mdash, and &hellip* representing quotation marks, the use of the long dash and horizontal ellipsis respectively. Some prepositional phrases that were disregarded in the output as the focus was on nouns collocating with the node words.

4.1.2.1 Procedure

The initial step was to find the most frequent words for the categories of Verb, Noun, and Adjective. The top four words of the three categories are listed in the tables below. The most frequent verbs were *DO*, *MAKE*, *GET*, and *TAKE*, the most frequent nouns were *TIME*, *WORK*, *LIFE*, and *WAY* and the most frequent adjectives were *NEW*, *GOOD*, *FIRST* and *OTHER*. The following section will present the results in detail. As the web interface only shows the raw frequency counts, and the research corpora were all different in size, the frequencies were normalised separately to instances per million.

4.1.2.2 Research corpora used

The two comparison corpora used to perform this comparison were the BNC and the I-EN. The BNC was used to provide data representing the British English variety (§2.1.4 for more information about the BNC). The I-EN was used to represent the American English variety. This corpus consists of 150 million words of online texts produced predominantly by American writers, containing 126,643,151 word tokens and 2,003,056 word types.

4.1.2.3 Research tools

The research tools used for the comparative study between BE and AE were IntelliText, Microsoft Excel, and SPSS. The IntelliText web interface has built-in access to the two selected comparison corpora, and it offers automatic collocation extraction and ranking using different association measures such as log-likelihood scores. The other two tools were used for storing and manipulating the data and performing the statistical testing. After extracting the most frequent words and their collocations, the next step was to include the compiled frequencies, and the textual data (collocation examples) were put into a separate Excel file. Then, the relevant numerical data were exported to SPSS to run statistical analyses and create graphs and diagrams.

4.1.3 Results

The tables below show the top four verbs, nouns, and adjectives in the two corpora (with absolute and normalised frequencies). These were thus the node words that were later examined for their *noun collocates*.

The first table, Table 4-2, shows the frequency data for the most frequent verbs in BNC and I-EN. The results indicate that these verbs occur far more in I-EN than than the BNC. The differences in the relative frequency between the BNC and I-EN is probably linked to differences in formality levels. The BNC contains formal data than the I-EN which contains Internet data that include less formal genres. Therefore, less formal genres tend to have far higher density of basic verbs (Sharoff, 2021, p.17).

Table 4-2: Most frequent verbs in BNC and I-EN

Most frequent Verbs	BNC (count)	BNC (IPMW)	I-EN (count)	I-EN (IPMW)
DO	94794	843	252507	1993
MAKE	65739	585	118367	934
GET	72598	646	140618	1110
TAKE	59240	527	82219	649

The results for the top twenty noun collocates for the four verbs are listed and illustrated in Table A-7: Verb-Noun collocations in the BNC and I-EN, in the appendices. Descriptive statistics were employed to explain the differences and similarities in collocations between the two groups. The collocation output in IntelliText was ordered based on the collocations' log-likelihood (LL) scores, which is a statistical measurement used to indicate the probability of words co-occurring. Despite the BNC having lower relative frequencies of the collocations of the verbs than I-EN, the collocation output reveals that BE and AE are alike to a considerable degree, with *DO*, *MAKE*, *GET* and *TAKE* having respectively 80%, 75%, 85%, and 75% similar collocations across both corpora. These percentages

mean that the four verbs had very similar noun collocates across the two corpora, with 70% of the top 20 collocations being similar for both corpora.

For the Noun-Noun collocations in the BNC and I-EN, I first extracted the top four most frequent nouns in both corpora (*TIME*, *WORK*, *LIFE*, and *WAY*) to act as node words. The raw and normalised frequencies for these nouns are recorded in Table 4-3. The table shows that these four nouns have a high overall frequency in both corpora ranging from 1800ipmw to 500ipmw. The normalised frequencies for the I-EN remain higher than those of the BNC.

Table 4-3: Most frequent nouns in BNC and I-EN

Most frequent Nouns	BNC (F)	BNC (NF)	I-EN (F)	I-EN (NF)
TIME	146974	1308	237244	1873
WAY	91810	817	132210	1043
WORK	63814	568	96362	760
LIFE	52439	466	84066	663

After generating the noun collocates of these four nouns, it was found that *TIME, WORK, LIFE*, and *WAY* collocate with similar nouns in both corpora. The percentage of the top twenty ranged from 45% up to 60%. This result is quite different from that for Verb-Noun collocations as it was difficult to find enough examples to be put in the top twenty most frequent nouns collocates (see, for example, the §*way*+ *noun* column in Table A-8 Noun-Noun Collocation list in the BNC and I-EN in the appendices). This difficulty could be because the topics discussed in the different texts compiled for the two corpora were probably very different. Another factor could be that Noun-Noun collocations in English are limited in number anyway, and have rarely been investigated in the literature (Balıkçı, 2011).

As for the results for the Adjective-Noun collocation, the top four adjectives in both corpora were *NEW, GOOD, FIRST*, and *OTHER* and their frequencies are shown below (§Table 4-4).

Table 4-4: Most frequent adjectives in BNC and I-EN

Most frequent Adjectives	BNC (F)	BNC (NF)	I-EN (N)	I-EN (NF)
New	104240	927	156182	1233
Good	76470	680	131841	1041
First	93727	834	118582	936
Other	139245	1239	262348	2071

The collocation output for the Adjective-Noun collocation yielded a large number of examples of similarly frequent collocations in both corpora. The percentage of the top twenty nouns that collocate with the adjectives *NEW*, *GOOD*, *FIRST*, and *OTHER* were 55%, 65%, 70%, and 80% respectively (§Table A-9 Adjective-Noun collocation in the BNC and I-EN in the appendices). The collocations for both corpora were therefore quite similar, with slight differences in their frequency rankings.

4.1.3.1 Statistical analysis

A statistical normality test was performed by means of the Shapiro-Wilk test on the results tabulated below (§Table 4-5). This step is a prerequisite to determining whether to apply parametric or non-parametric tests (for normally distributed or non-normally distributed data respectively).

Table 4-5: Normal distribution in the collocation sets investigated for the comparison between BE and AE

Verb-Noun collocation	BNC Shapiro-Wilk	I-EN Shapiro-Wilk
DO + Noun	0.000	0.000
MAKE +Noun	0.000	0.000
GET + Noun	0.002	0.009
TAKE + Noun	0.000	0.003
Noun-Noun collocation		
TIME + Noun	0.002	0.001
WORK + Noun	0.000	0.000
LIFE + Noun	0.000	0.000
WAY + Noun	0.000	0.000
Adjective- Noun collocation		
NEW + Noun	0.000	0.000
GOOD + Noun	0.000	0.000
FIRST + Noun	0.000	0.000
OTHER + Noun	0.000	0.000

The results were all below the threshold (p<0.05) indicating that the data significantly deviate from a normal distribution. Figure 4-1 is a histogram of the collocates of *WORK-Noun* which is representative of how the data are distributed.

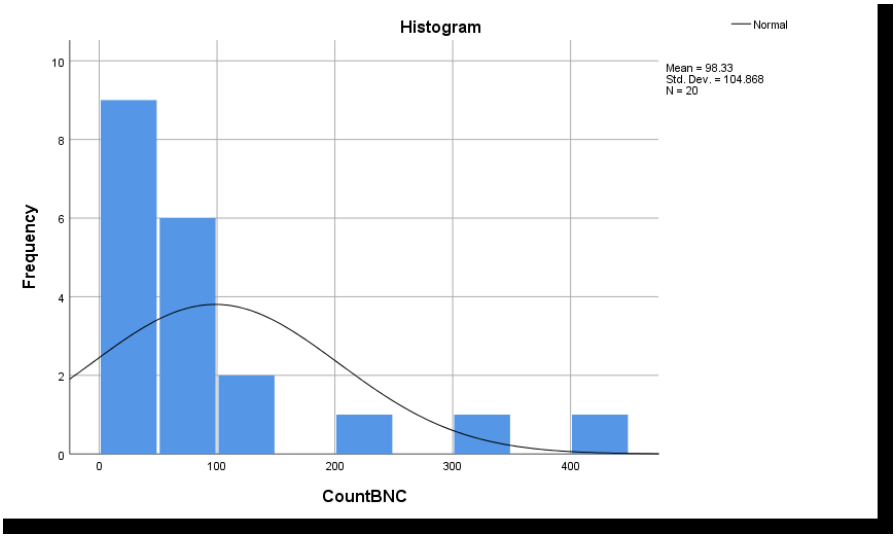


Figure 4-1: Histogram of *WORK-Noun* collocations

The curve in Figure 4-1 is right-skewed for this set of collocations, telling us that the data do not have a normal distribution. Because the data are not normally distributed, a non-parametric test had to be used for the data analysis. The Mann-Whitney *U* test was performed on the two corpora and the results were compared against the results of each list of each corpus in order to see whether there is a significant difference between BE and AE. The mathematical formula for calculating the Mann-Whitney *U* is presented in Equation 4.1:

$$U = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - \sum_{i=n_2+1}^{n_2} R_i \qquad \textbf{Equation 4.1}$$

Where:

U is the Mann-Whitney *U* test.

N_1 stands for sample size 1

N_2 stands for sample size 2

R_i stand for the rank of the sample size

The Mann-Whitney *U* test leads to a definition of the significance level in the form of a *z*-score. The *z*-score is used when comparing a sample mean with the hypothetical mean of a population. It is performed by calculating the sample mean and then subtracting the population mean over the standard deviation (§Equation 4.2).

$$z = \frac{U - m_U}{\sigma_U} \qquad \textbf{Equation 4.2}$$

Where:

U= is the sample mean

m_U= is the population mean

σ_U= is the standard deviation

Overall, the results show a significant difference between the two groups (§Table A-9:, Table A-10 and Table A-11 in the appendices). For the Verb + Noun collocations, the results for the four verbs were U= 94, 40, 43, and 84 < U =127 at α=0.05 (a two-sided level of significance). For the Noun + Noun collocations, the results for the four selected nouns were U= 14, 64, 76, and 71< U =127 at (α=0.05 two-tailed). For the last collocation, Adjective + Noun, the results were U= 35, 54, 88, and 91< U =127 at α=0.05 (a two-sided level of significance).

However, there are four collocates in the I-EN that indicate that there are no significant differences with p set at > 0.05 which confirms that evidence does not allow the rejection of the H_0. These were found in the following collocations:

- *DO* + Noun p=0.70 > 0.05
- *TAKE* + Noun p=0.13 > 0.05
- *TIME* + Noun p=0.234 > 0.05
- *WORK* + Noun p=0.088 > 0.05

4.1.4 Discussion

The results reported above were generated using descriptive statistics. Owing to the fact that the collocation outputs were arranged based on their probability of co-occurrence, the results for each set of collocates were expected to yield similar results. This turned out to be a robust procedure for conducting a preliminary comparison between BE and AE. Nevertheless, the statistical test unexpectedly yielded some results that put into dispute what has been found by other scholars. One big difference was in the case of Noun-Noun collocations, mainly those consisting of *WAY-Noun,* for which it was difficult to find similar collocations between BE and AE in the two corpora. One possible reason is in the fact that the noun *WAY* is frequently associated with words that indicate places, locations, or a description of an action.

Therefore, though there were two opposing results of the Mann-Whitney U test, I do reject the H_0, as most *p-values were* <0.05. The two opposing U test results

explains that there is a possibility of similarities between the BE and AE in the use of lexical items within collocations regardless of their significance within the corpora. I therefore claim that there is no significant difference between BE and AE in terms of the choice of collocations as the results for both corpora were quite similar. This finding somewhat contradicts Darragh (2000), who found that there are 4,000 frequently used words in BE that are either not that much used or used differently in AE, and also Deáková (2013), who found that there are differences between BE and AE in collocations when there are differences in the meaning of particular words.

4.1.5 Limitations and conclusion

This comparison has certain limitations. The BNC is based on data collected from 1970 to the early 1990s (Leech, 1992a). As any language change since the 1990s is not documented in the BNC, contemporary language use cannot be observed in the data collected in the BNC (Sharoff, 2006). Therefore, this may have affected the comparison between the two English varieties. This study was designed initially as a pilot study for this PhD, but is used here to introduce the investigation of collocations used by Arab EFLLs and NBES.

4.2 The results of the comparative corpus-based frequency analysis of the use of lexical collocations by Arab EFLLs and NBESs

Introduction

Having presented the results of the comparison between British English and American English use of lexical collocation above, in the rest of this chapter I will present the statistical analysis and the contrastive error analysis of the use of lexical collocations by Arab EFLLs and NBESs. I will then discuss the important findings of this research.

As with the study above, I will present the results for the use of Adverb-Adjective collocations[2], Verb-Noun collocations, Verb-Adverb collocations, and Adjective-Noun collocations. The results for all three collocation types will be given in two separate sections based on the two analytical approaches mentioned above: quantitative and qualitative.

4.2.1 Quantitative results: a Corpus-based Frequency analysis

In Chapter 2, I presented the previous literature on the Verb-Noun collocation and its theoretical background (§2.3.2). Here, I will further present an in-depth discussion of the findings in line with the literature relating to the Verb-Noun collocation. A general commentary on the results and a final discussion will be given in a separate section at the end of this thesis (§Chapter 5).

I entered a verb as the node word, but used the 'lemma' option in the search (available in both the IntelliText and Sketch Engine web interfaces). Using the lemma option saves time by capturing all possible inflections of the node word. Then, when searching for collocates of this verb, I used the option for specifying

[2] The results were presented in a published papar: Alshammari, R.R.M. 2021. Adverb-Adjective Collocation Use by Arab EFLLs and British English Native Speakers: a comparative corpus-based study. *The Language Scholar.* **9**, pp.24-65.

the part of speech (POS) and chose ‘nouns’, and set a span of two words to the right (meaning the noun can be one or two words after the verb). In the Verb-Noun collocation study, I searched for the verb lemma and extracted noun collocates with/without an intermediate word, and a separate search for the infinitive form of the verb followed by a noun (with/without an intermediate word). The intermediates refer to any possible word that might occur within a collocation, such as the articles in ‘break a record’ or ‘break the law’. Therefore, this case study reports the results for the four following patterns of the Verb-Noun collocation:

- *VERB-Noun* collocation
- *VERB-(Intermediate)-Noun* collocation
- *to* + *VERB - Noun* collocation
- *to* + *VERB- (intermediate)-Noun* collocation

The use of infinitives in VERB-Noun collocations was separately investigated (e.g. ‘to book an appointment’ or ‘to play football’). When used with the infinitive form (*to* + Verb-Noun collocation), the Verb-Noun collocation is still considered as a lexical (not grammatical) collocation, as stated by Benson et al. (2009). The reason for separately examining this combination is that Arabic has no infinitives (Hafiz et al., 2018), which makes it problematic for Arab EFLLs to accurately use expressions such as ‘أَكَلَ’ (*ʾakala*) ‘to eat’ in sentences like ‘he likes to eat’. To demonstrate this, I searched the corpus for such errors in the data:

Arab EFLLs	Corrected vestion
- *We went to the Zoo 9:00 morning. I sew in the Zoo the animal Examples the rabhet, tiger, Leon, snek,sonic , gazil , bear, fox, and bird Examples falckon. In the afternoon went ***eat the*** lunch with my famely . We fi nsh lunch falow way[3]	- In the afternoon, we went to **eat** lunch with my family
- *Will go to me brather he smayll for me. And my father ***go get*** food	- My father went to **get** food

[3] The use of (*) indicates that the text in the written data has spelling and grammatical mistakes made by the learners.

The above verbatim texts from the AEFLL Corpus demonstrate the type of error Arab EFLLs make when using the infinitive verb in Verb-Noun collocations. Both examples show the absence of the infinitive marker '*to*' with the main verb in the sentence. According to Schulz (2004), Arabic has a rough equivalent of the English infinitive, known in Arabic as the مصدر (literally: 'source'), a term which is sometimes translated into English as 'verbal noun'. The مصدر 'verbal noun' can be considered to be the nominal form of the Arabic verb, i.e. it expresses what the verb expresses, but in the form of a noun. For example, قراءة (*qiyr'āat*) 'reading' (from the root q-r-a). Schulz (2004) states that there are 40 infinitive patterns[4] in Arabic. One of these, which is very rare in Arabic, could include the use of the Arabic equivalent of the '*to'* infinitive with a verb, such as 'to become poor' one equivalent of which is تَمَسْكُنٌ (*tamasskun*) in Arabic. The use of the infinitive with a verb is problematic for Arab EFLLs because it is relatively rare in Arabic. Even if we accept Schulz (2004) contention that the infinitive structure does exist in Arabic, it is still, by all accounts, very rare, and thus the problem that Arab EFLLs have with the infinitive in English is worth further investigation.

4.2.1.1 Verb-Noun collocation results

In this section, I will provide the results for the frequency analysis of Verb-Noun collocations found in the data. The extraction procedure was planned and executed in four steps to ensure accuracy, reliability, consistency, and clarity. The steps were the following:

> **Step One:** Using the Wordlist tool in the Sketch Engine web interface, the top 200 most frequent verbs in the NBES and the Arab EFLLs corpora were compiled (§Table A-14 and Table A-13 in the Appendices).

[4] Patterns in Arabic are represented using the modal root فعل (*f-ʕ-l*) (Dickins, J. 2011. Faʿ* l forms in Sudanese Arabic: the reassertion of morphology. *Zeitschrift für arabische Linguistik.* **160**(53), pp.36-67.). Patterns in Arabic refer to the arrangement of vowels and consonant around the root. The root is used to represent the set of consonants that indicate the core meaning of the word. For example, the word كتب (*katab*) 'he wrote' has the pattern فعل (*f-ʕ-l*) as (k-t-b).

Step Two: The frequency lists were organized based on their normalised frequency of occurrences (§see the definition and the formula for calculating the normalised frequency in Chapter 2).

Step Three: Using the collocation tool in the IntelliText web interface with these verbs as search terms (node words), the co-occurring nouns (the collocates) were retrieved.

Step Four: The collocations were loaded in Excel spreadsheets and analysed in SPSS.

It should be noted that no extraction process is devoid of error. The association strength threshold in this study was set at LL ≥ 5 which is considered to be very low compared to other studies where the cut-off point was set to 30, as by Shin and Nation (2008), and to 20 by other scholars, such as Nesselhauf (2005) and Farooqui (2016). Various factors play a role in deciding the threshold, such as corpus size, window size[5] and the ranking method (Seretan, 2018). Therefore, considering the small size of the Arab EFLLs Corpus, I used the LL threshold and was set to LL ≥ 5 to increase the number of extracted collocations. The results for setting the LL to ≥ 10 would lead to 49 Verb-Noun collocations for the Arab EFLLs and 237 for the NBESs. When the LL score is set to LL ≥ 5 the number of collocations is high in both corpora, allowing Verb-Noun collocations to be investigated thoroughly.

The results for the Verb-Noun collocations will be discussed in four sub-sections:

- The results for the Verb-Noun collocations will be given in Sub-section (§4.2.1.1.1).
- The results for the Verb-(Intermediate)-Noun collocations will be presented in Sub-section (§4.2.1.1.2).

[5] Window size indicates the collocation span: a large window or span allows for intermediates between the node and collocate.

- The results for the *to* + Verb-Noun collocations will be presented in Sub-section (§4.2.1.3).
- The results for the *to* + Verb-(Intermediate)-Noun collocations will be presented in Sub-section (§4.2.1.1.4)

4.2.1.1.1 Statistical Analysis Results for Verb-Noun collocations Results

The analysis was carried out on the generated top 200 most frequent verbs from the Arab EFLLs Corpus, and then use them as nodes for the Arab EFLLs Corpus. Then I generated the top 200 verbs from the NBESs Corpus and then use them as nodes for the NBESs Corpus. These two lists were common as the top 200 verbs to the BNC as the reference corpus, and to discover and compare their potential for forming Verb-Noun combinations (§Table A-13 and Table A-14 for the list of the 200 verbs in each corpus in the appendices).

Table 4-6: Percentage of Verbs of with threshold noun *collocates* in Verb-Noun Collocations

	No. of verbs that has a noun collocate in the corpus	%
Arab EFLLs	53	26.5%
NBES	61	30.5%

Table 4-6 illustrates the number of verbs that had noun collocates that emerged in the extraction phase from the Arab EFLL and NBES corpora. There were 53 verbs that had noun collocates within the threshold, in the AEFLL Corpus accounting for 26.5% of the total 200 verbs searched for. However, there were 61 verbs that had noun collocates within the threshold, in the NBESs Corpus accounting for 30.5% of the total verbs searched for. The number of collocates extracted for Arab EFLLs appears small compared to the NBESs, but that is because many of the collocations were below the frequency threshold used. Nonetheless, the number of collocations extracted for these verbs is smaller for the Arab EFLLs when compared to the NBESs. The total numbers of Verb-Noun collocations extracted for these verbs are 145 for the Arab EFFLs and 393 for the NBESs (§Table 4-7).

Table 4-7: Total number of Verb-Noun collocations within Arab EFLL and NBES corpora

LL score cut-off point	Arab EFFLs corpus	%	NBESs corpus	%
above 300	0	0%	3	0.76
above 200	1	0.6%	2	0.50%
above 100	0	0%	8	2.03%
above 50	6	4.13%	31	7.88%
above 25	7	4.82%	55	13.99%
above 20	6	4.13%	30	7.63%
above 10	29	20%	108	27.48%
above 5	96	66.20%	156	41.98%
Totals	145	100%	393	100%

Table 4-7 illustrates the range and distribution of the collocation LL scores within the Arab EFLL and NBES corpora. Table 4-7 is a simplified version derived from the extracted collocations presented in Table A-15 and Table A-16 in the Appendix that are arranged according to the log-likelihood measure. Table A-15 shows the Verb-Noun collocations extracted from the AEFLL Corpus along with the LL scores of their occurrences in the second column. Table A-16 illustrates the extracted Verb-Noun collocations in the NBESs Corpus for the NBESs. Table 4-8 shows a sample of the extracted collocations for the Arab EFLLs and the NBESs. The AEFLL Corpus shows similar collocations to NBES in the choice of collocates but Arab EFLLs use fewer collocates. For example, the verb ‘have’ in the AEFLL Corpus had two noun collocates as in *have fun* and *have nothing*. However, the NBESs were using other nouns, such as ‘have difficulty’, ‘have implication’, ‘have trouble’, etc.

Table 4-8 Verb-Noun collocations in the Arab EFLL and NBES corpora[6]

Arab EFLLs	LL score	NBESs	LL scores
Have ~~ nothing	8.21	have ~~access	89.61
Have ~~ fun	74.04	have ~~difficulty	51.18
Do ~~ anything	54.85	have ~~nothing	46.28
Do ~~ something	34.34	have~~ implication	34.14
Do ~~everything	11.92	have ~~child	18.74
Do ~~ something	8.07	have~~ trouble	13.7
Do ~~nothing	8.04	do ~~something	85.52
Go ~~ shopping	35.75	do~~ anything	55.22
Save~~ money	11.5	do ~~nothing	49.9
Save ~~time	8.25	do~~ thing	49.02
Save ~~life	5.88	do~~ everything	15.28
Make ~~people	20.22	do~~ mathematics	10.98
Make ~~sense	14.35	make~~ sense	197.1
Make~~ decision	8.7	make~~ decision	118.3
Make ~~*exrcise	6.02	make~~ use	51.34
Make ~~ life	5.77	make~~ mistake	49.67
Make ~~revision	5.07	make~~ profit	27.37
Spend ~~time	6.39	make~~ reference	27.16
Spend ~~money	5.32	save~~ money	42.18
Become ~~part	15.81	save~~ time	30.7
Cause ~~problem	41.95	cause~~ problem	62.77
Cause ~~obesity	19.74	cause~~ damage	24.45
Cause ~~air	13.43	cause~~ harm	21.94
Cause ~~injury	13.16	keep~~ track	60.23
Keep~~ fit	13.21	keep ~~pace	25.9
Keep ~~silence	7.52	keep ~~inflation	12.89

The first notable difference in Table 4-7 is in the total number of collocations extracted, which shows that Arab EFLLs have a lower use of Verb-Noun collocations than NBESs; there were 145 collocations for Arab EFLLs and 393 collocations for NBESs. Though the LL score cut-off point was set to a relatively low level (LL ≥ 5) because selecting a higher LL score would have excluded many collocations for the Arab EFLLs, the NBESs had more collocations with higher LL scores than the Arab EFLLs. The accuracy of automatic extraction depends on several factors, for example, filtering out collocation output generated due to part-of-speech (POS) tagging errors. Therefore, during the collection phase, I carefully filtered collocations to eliminate any inappropriate collocations or tagging issues. Verb-Noun collocations that had LL scores above 20 number 129 for NBESs and 20 for Arab EFLLs. There were many Verb-Noun collocations that were in the range of (*LL* ≥ 5 - ≥10) accounting for 66% of the total extracted collocations of the Arab EFLLs and 41.98% of the NBESs. The highest LL score for the Arab

[6] The sign (~~) indicates that a span is possible between the node word and the collocate.

EFLLs is 230.2 (M = 13.98, SD = 22.30), and the highest for the NBESs is 957.87 (M = 27.19, SD =62.24). The other Verb-Noun collocation LL scores remain higher for the NBESs than the Arab EFLLs (§Table 4-9).

Table 4-9: Descriptive statistics for the use of Verb-Noun collocations

Descriptive statistics	Maximum LL score	Mean	Standard deviation	Skewness	Kurtosis
Arab EFLLs	230.20	13.98	22.30	7.006	62.511
NBESs	957.87	27.19	62.24	10.282	136.592

The following is a report on the results of the statistical significance testing using SPSS based on the two tables in the appendices that show the results of the LL scores for the use of the Verb-Noun collocation for Arab EFLLs and NBESs (§Table A-15 and Table A-16 in the appendices). The first step was to perform a statistical test based on the collocations' frequencies for the Arab EFLLs and NBESs. Before taking this step, to choose the correct statistical significance test it needed to be determined whether the distributions of the data follow a normal or non-normal distribution. This step was performed using histograms that help in visualising the data distribution of both Arab EFLLs and NBESs.

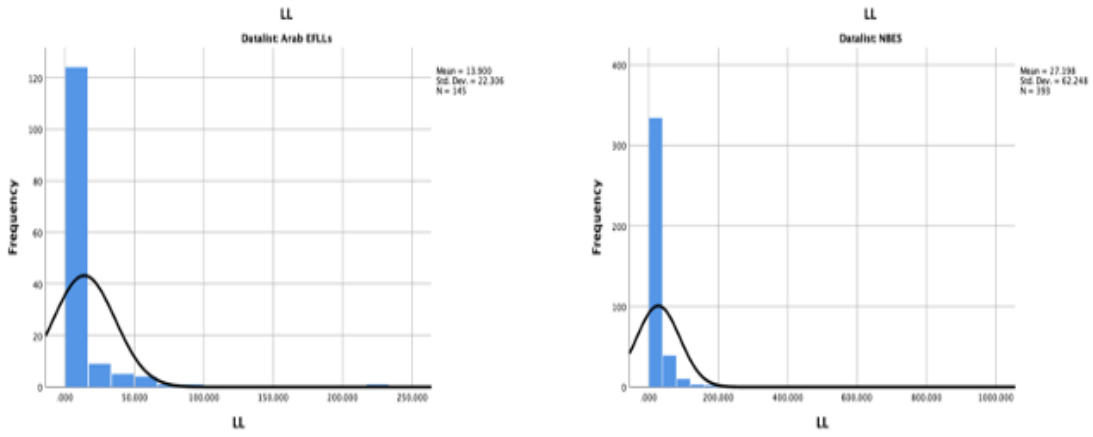


Figure 4-2: Histograms for the Verb-Noun collocation frequencies in Arab EFLL and NBES corpora

The histograms in Figure 4-2: Histograms for the Verb-Noun collocation frequencies in Arab EFLL and NBES corpora show that the data for both Arab EFLLs and NBESs

do not follow a normal distribution. Other indicators for the shape of the distribution were the results for skewness and kurtosis which are given in Table 4-9. The results indicated that the distribution was positively skewed, amounting to 7.00 for Arab EFLLs and 10.28 for NBESs. The results for kurtosis suggested that the distributions are heavily tailed compared to the normal distribution. They were found to be 62.51 for Arab EFLLs and 136.59 for NBES. Therefore, since both collocation data lists are positively skewed this confirms the need to carry out a non-parametric test.

There are four assumptions that need to be met before conducting a Mann-Whitney *U* test. The Verb-Noun collocation is distributed between two groups and is ordinal which satisfies the first two assumptions listed in (§3.4.3.2). The third assumption is the non-normality of the distribution. If the data distribution does not show a normal distribution yet has a similar shape of distribution, this meets the vital fourth assumption that justifies the need to use the Mann-Whitney *U* test over any other non-parametric test. Therefore, to test the assumption of a similar distribution, a homogeneity of variances test is needed, the results of which are presented in Table 4-10.

Table 4-10: Levene's Test of Homogeneity of Variances for Verb-Noun collocations

Test of Homogeneity of Variances					
Verb – Noun collocation LL scores	Based on Median and with adjusted *df*	Levene Statistic	*df*1	*df*2	*Sig*.
		4.882	1	427.099	0.029

According to the results in Table 4-10, the significance of Levene's statistic test was (p = 0.000 > 0.05). This means that the assumption of homogeneity of distribution between the Arab EFLLs and the NBESs is not satisfied (F (1.427) = 4.882, p = 0.029). This implies that the non-parametric version of homogeneity of variances has been rejected, so the *p*-value of the non-parametric Mann-Whitney *U* test could not be interpreted. Therefore, an independent sample t-test with bootstrapping as the estimation technique had to be used (§see Chapter 3,

section 3.4.3 for further discussion regarding the choice of the right statistical test). The advantage of the bootstrapping technique is its ability to estimate the distribution of a sample without relying on the assumptions required by parametric tests. This makes it a powerful tool for statistical inference, particularly in situations where the underlying population distribution is unknown or non-normal (§3.4.3.2 for more explanation about the independent t-test with bootstrapping technique).

Table 4-11: Independent Sample t-Test of the Verb-Noun collocation

	No. of Collocations	Mean	Std. Deviation	t-test			
				F	*Sig.*		*df*
Arab EFLLs	145	13.8996	22.30559	8.129	0.005	-3.648	535.700
NBES	393	27.1981	62.24791				

An Independent Sample *t*-test was conducted to compare the LL scores of the Verb-Noun collocation for Arab EFLLs and NBESs. Table 4-11 illustrates the SPSS output for the independent sample t-test using bootstrapping as the estimation technique. There was a significant difference in the LL scores of the Verb-Noun collocation for Arab EFLLs (M = 13.8996, SD= 22.30559) and NBESs (M = 27.1981, SD= 62.24791); t (535.700) = -3.648, p = 0.005. The p-value for the t-test suggests that we reject the null hypothesis that claims there is no difference between Arab EFLLs and NBESs in their use of Verb-Noun collocations.

4.2.1.1.2 Statistical Analysis Results for Verb-(Intermediate)-Noun collocations

In this section, I investigate the results for the Verb-Noun collocation list when allowing intermediates (i.e., Verb-(intermediate)-Noun collocations). There were 31 collocations with this pattern for both the Arab EFLLs and NBESs. Those 31 collocations are listed and presented in Table 4-34.

Table 4-12: Descriptive statistics for the use of Verb-(Intermediate)-Noun collocations

Descriptive statistics	Number of collocations	Maximum LL score	Mean	Standard deviation	Skewness	Kurtosis
Arab EFLLs	31	6.04	0.86	1.19	1.424	11.425
NBESs	31	61.39	5.09	10.65	5.243	28.468

The LL scores for these collocations were below the threshold in most cases. This is because I needed to increase the number of extracted collocations to investigate the possible differences between Arab EFLLs and NBESs. The highest LL score of these collocations is 6.04 (M = 0.89, SD = 1.19) for the Arab EFLLs and 61.39 (M = 5.09, SD =10.65) for the NBES (§Table 4-12).

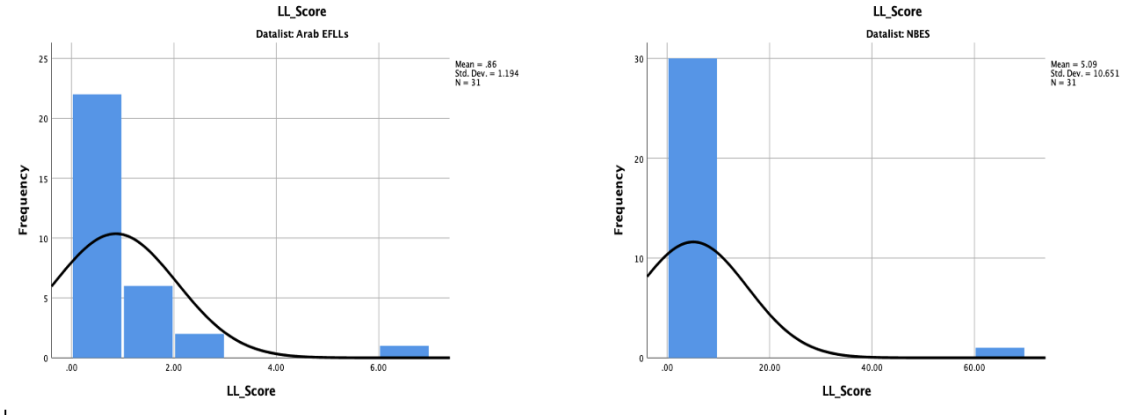


Figure 4-3: Histograms for the Verb-Noun collocation frequencies allowing intermediates in Arab EFLL and NBES corpora

Figure 4-3 shows a histogram of the data distribution for the Verb-Noun Collocation when allowing intermediates within the collocation in the two corpora. The results for skewness and kurtosis given in Table 4-12 indicate that the distribution was positively skewed, amounting to 1.42 for Arab EFLLs and 5.24 for NBESs. The results for kurtosis are 11.42 for Arab EFLLs and 28.46 for NBESs suggesting that the distribution is heavily tailed compared to a normal distribution. Based on Figure 4-3 and the results for skewness and kurtosis, the data were not normally distributed; therefore, as stated earlier, a non-parametric test was needed to perform the statistical analysis, which was, in this case, the Mann-Whitney *U* test. As illustrated previously, one assumption that must not be

violated when performing the non-parametric Mann-Whitney *U* test is that the distribution has to follow a similar pattern in both data-sets.

Table 4-13: Levene's Test of Homogeneity of Variances for Verb-(Intermediates)-Noun collocations

Test of Homogeneity of Variances					
Verb-(intermediates)-Noun collocation LL scores	Based on Median and with adjusted *df*	Levene Statistic	*df*1	*df*2	*Sig.*
		2.133	1	30.643	0.154

Table 4-13 shows the results based on the median and with an adjusted degree of freedom *df*. Levene's statistic test showed that the assumption of homogeneity of variances was not violated for the Verb -(intermediates)-Noun collocation data (F (1.30) = 2.133, p = 0.154). Thus, the Mann-Whitney *U* test could be used, and the results are illustrated in Figure 4-4.

Mann-Whitney Test

Ranks

	Datalist	N	Mean Rank	Sum of Ranks
LL_Score	Arab EFLLs	31	19.13	593.00
	NBES	31	43.87	1360.00
	Total	62		

Test Statistics[a]

	LL_Score
Mann-Whitney U	97.000
Wilcoxon W	593.000
Z	-5.400
Asymp. Sig. (2-tailed)	.000

a. Grouping Variable: Datalist

Figure 4-4: Mann-Whitney *U* test for the use of the Verb-Noun collocation when allowing intermediates in Arab EFLLs and NBES corpora.

When adding an intermediate to the Verb-Noun collocation, the Mann-Whitney *U* test indicates that the NBESs have the highest median rank. The Test Statistics table in Figure 4-4 illustrates that the use of the Verb-Noun collocation when allowing intermediates for NBESs (*Mdn*=43.87) differed significantly from its use by Arab EFLLs (*Mdn*=19.13), *U* =593, *p* = 0.000. Therefore, the *p*-value suggests that we reject the null hypothesis and conclude that there is a statistically significant difference between the two corpora in the use of Verb-(intermediate)-Noun collocations.

4.2.1.1.3 Statistical Analysis Results for to + Verb-Noun Collocations

In this section (§4.2.1.1.3), I present the results for the last two combinations: for the use of the Verb-Noun collocation with infinitives and for when intermediate words are included in the same collocation combination; the first is *to* + Verb-Noun collocation and the second is *to* + Verb-(intermediates)-Noun collocation.

Table 4-14: Descriptive statistics for the use of *to* + Verb-Noun collocations

Descriptive statistics	Number of collocations	Maximum LL score	Mean	Standard deviation	Skewness	Kurtosis
Arab EFLLs	13	13.44	3.36	11.71	2.397	6.631
NBESs	14	7.74	3.23	3.19	1.790	2.627

Table 4-14 shows the general descriptive statistics for the use of *to* + Verb-Noun collocations (§the collocations themselves are listed in Table 4-35 and will be analysed in the contrastive error-analysis section 4.2.3.2). The number of *to* + Verb-Noun collocations is 13 for the Arab EFLLs and 14 for the NBESs. The maximum LL score was 13.44 (M = 3.36, SD = 11.71) for the Arab EFLLs and for the NBESs the highest LL score is 7.74 (M = 7.74, SD = 3.19). For this collocation, the LL score is higher for the Arab EFLLs, which was not an expected result.

The results for skewness and kurtosis indicate that the distribution was positively skewed, amounting to 2.39 for Arab EFLLs and 1.79 for NBESs. The results for kurtosis are 6.63 for Arab EFLLs and 2.62 for NBESs suggesting that the

distribution is more tailed than the normal distribution. This result is also confirmed in the two histograms below that show a non-normal distribution in the data requiring the use of a non-parametric test (§Figure 4-5).

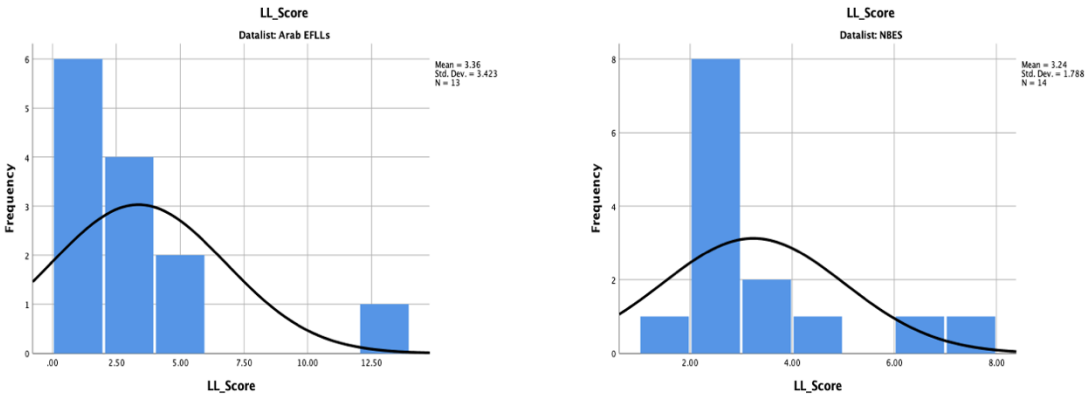


Figure 4-5: Histograms for the frequencies of the use of *to* + Verb - Noun collocation in Arab EFLL and NBES corpora

The distributions in both graphs show a non-normal distribution in the data requiring the use of the Mann-Whitney *U* non-parametric test. Prior to performing the Mann-Whitney *U* test, the vital assumption of homogeneity of variances must be justified (§Table 4-15).

Table 4-15: Levene's Test of Homogeneity of Variance for *to* + Verb-Noun collocations

Test of Homogeneity of Variance					
to + Verb-Noun collocation LL scores	Based on Median and with adjusted *df*	Levene Statistic	*df*1	*df*2	*Sig.*
		1.430	1	19.865	0.246

Table 4-15 illustrates the results of the test of homogeneity of variances to determine whether the data distribution is of a similar nature. Levene's statistic test revealed that the assumption of homogeneity of variances was not violated for the *to* + Verb-Noun collocation data (F (1.19) = 1.430, *p* = 0.246). Thus, the Mann-Whitney *U* test could be used, and the results are illustrated in Figure 4-6.

Mann–Whitney Test

Ranks

	Datalist	N	Mean Rank	Sum of Ranks
LL_Score	Arab EFLLs	13	13.04	169.50
	NBES	14	14.89	208.50
	Total	27		

Test Statistics[a]

	LL_Score
Mann–Whitney U	78.500
Wilcoxon W	169.500
Z	-.607
Asymp. Sig. (2–tailed)	.544
Exact Sig. [2*(1–tailed Sig.)]	.550[b]

a. Grouping Variable: Datalist

b. Not corrected for ties.

Figure 4-6: Mann-Whitney *U* test for the *to* + Verb-Noun Collocation for Arab EFLL and NBES corpora

Figure 4-6 illustrates the SPSS output of the Mann-Whitney *U* test for the *to* + Verb-Noun collocation. Because the infinitive form does not exist in Arabic, the main aim of investigating the use of *to* with the Verb-Noun collocation was to see whether there is a difference between Arab EFLLs and NBESs. The results of this collocation demonstrate that the difference is not significant as the median rank results were 13.04 for Arab EFLLs and 14.89 for NBESs. The Mann-Whitney *U* test results suggest that there is no difference in the use of *to* + Verb-Noun for Arab EFLLs and NBESs thus accepting the null hypothesis at $U = 78.5$, $p = .550$.

4.2.1.1.4 Statistical Analysis Results for to + Verb-(intermediate)-Noun collocations

In this section (§4.2.1.1.4), I provide the statistical significance testing for the use of the *to* + Verb-(Intermediate)-Noun collocation by the Arab EFLLs and NBESs. Table 4-16 illustrates that the number of instances within the two corpora is low amounting to 7 collocations for the Arab EFLLs and 13 for the NBESs. The highest LL score for the Arab EFLLs is 3.09 (M = 1.11, SD = 1.12) and for the NBESs 17.06 (M = 4.29, SD = 4.33). These results indicate a very low use of

these collocations within both corpora, but I must explain that the main reason for investigating the last two collocations is to have a general overview of the use of these collocations before widening the investigation of these collocations through contrastive-error analysis.

Table 4-16 Descriptive statistics for the use of *to* + Verb-(Intermediates)-Noun collocations

Descriptive statistics	Number of collocations	Maximum LL score	Mean	Standard deviation	Skewness	Kurtosis
Arab EFLLs	7	3.09	1.11	1.12	1.030	1.587
NBESs	13	17.06	4.29	4.33	2.450	1.191

Table 4-16 illustrates the values for skewness and kurtosis, which indicate that the distribution was positively skewed, amounting to 1.03 for Arab EFLLs and 2.45 for NBES. The results for kurtosis are 1.5 for Arab EFLLs and 1.19 for NBESs suggesting that the distribution is tailed to the right compared to the normal distribution. This result is also confirmed in the two histograms below (§Figure 4-7).

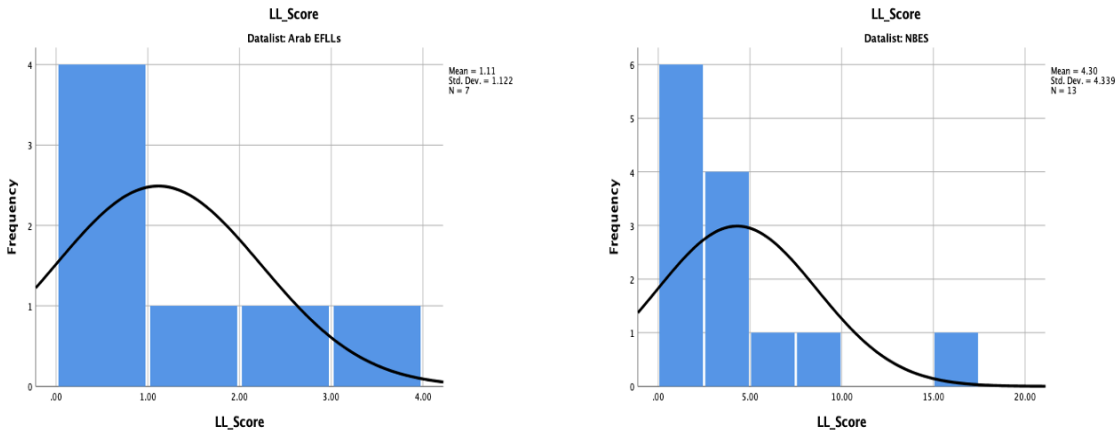


Figure 4-7: Histograms for the frequencies of the use of *to* + Verb-(intermediates)-Noun collocations in the Arab EFLL and NBES corpora

The histogram distributions and the results for skewness and kurtosis demonstrate that the data have a non-normal distribution, necessitating the use of the non-parametric Mann-Whitney *U* test. The crucial assumption of homogeneity of variances must be supported before running the Mann-Whitney *U* test. The results for the homogeneity of variances test are given in Table 4-17.

Table 4-17: Levene's Test of Homogeneity of Variances for *to* + Verb-(Intermediate)-Noun collocations

Test of Homogeneity of Variances					
to +Verb-(intermediates)-Noun collocation LL scores	Based on Median and with adjusted *df*	Levene Statistic	*df*1	*df*2	*Sig.*
		1.408	1	12.495	0.257

Levene's test revealed that the assumption of homogeneity of variances was not violated in the *to* + Verb-(Intermediate)-Noun collocation data (F (1.12) = 1.408, *p* = 0.257). Thus, the most appropriate test for this type of distribution was the Mann-Whitney *U* test, the results of which are illustrated in Figure 4-8.

Mann–Whitney Test

Ranks

	Datalist	N	Mean Rank	Sum of Ranks
LL_Score	Arab EFLLs	7	5.57	39.00
	NBES	13	13.15	171.00
	Total	20		

Test Statistics[a]

	LL_Score
Mann–Whitney U	11.000
Wilcoxon W	39.000
Z	–2.734
Asymp. Sig. (2–tailed)	.006
Exact Sig. [2*(1–tailed Sig.)]	.005[b]

a. Grouping Variable: Datalist

b. Not corrected for ties.

Figure 4-8: Mann-Whitney *U* test for the use of *to* + Verb-(intermediates)-Noun collocation in Arab EFLL and NBES corpora

The Mann-Whitney *U* test results for the *to* + Verb-(intermediates)-Noun collocation suggest that the NBESs have 13.15 as the median rank which is significantly different from the results for the median rank of the Arab EFLLs at 5.57. The comparison suggests that there is a significant difference in using the *to*+ Verb-(intermediates)-Noun collocation between Arab EFLLs and NBESs, thus rejecting the null hypothesis at *p*-values of $U = 11$, $p = 0.005$.

4.2.1.2 Verb-Adverb Collocation Results

In this section (§4.2.1.2), I present a statistical analysis for the Verb-Adverb collocation study. The combinations that were investigated are the Verb-Adverb collocation with/without an intermediate word. The verbs used in this study are the same verbs selected for the Verb-Noun collocation case study (§Table A-13 and Table A-14 in the Appendices). These verbs are considered part of the top 200 verbs used in the BNC as the reference corpus. They are used as the node words. Some of the collocations extracted are given in Table 4-18. Due to the large number of extracted Verb-Adverb collocations, the remaining collocations for the Arab EFLLs and the NBESs are provided in the appendices (§Table A-17 and Table A-18). In general, the Arab EFLLs demonstrated a similar use of Verb-Adverb collocations to the NBESs. It is clear that the reason for this is that they tend to use similar adverb collocates in Arabic.

Table 4-18 Verb-Adverb collocations in the Arab EFLL and NBES corpora

Arab EFLLs	LL score	NBESs	LL scores
be~~very	990.39	be ~~ very	2623.72
be ~~ always	27.31	be ~~ always	598.57
be ~~ quiet	25.73	be ~~ quite	717.73
be ~~ extremely	25.62	be ~~ extremely	487.68
have ~~ just	95.1	have ~~ recently	155.78
have ~~ ever	88.08	have ~~ significantly	86.49
have ~~ recently	30.8	have ~~ only	70.49
have ~~ already	14.91	have ~~ just	50.72
take ~~ away	9.39	take ~~ seriously	63.98
take ~~ immediately	7.34	take ~~ long	19.66
think ~~ twice	16.37	take ~~ directly	5.05
Mention~~ previously	5.78	think ~~ about	24.9
travel ~~ abroad	20.97	think ~~ differently	6
affect ~~ negatively	9.02	drive ~~ dangerously	9.11
arrive ~~ safely	5.93	drive ~~ forward	8.43
drive ~~ fast	19	drive ~~ carefully	7.74
work ~~ hard	32.86	work ~~ properly	15.22
work ~~ together	9.55	work~~ collaboratively	14.61
work ~~ effectively	6.41	grow ~~ rapidly	56.46
sleep ~~ well	5.37	reach ~~ approximately	6.2
rely ~~ totally	5.79	reach ~~ quickly	5.36
act ~~ aggressively	14.17	use ~~ exclusively	5.32
use ~~ wisely	12.85	use ~~ clinically	5.28
use ~~ extensively	10.08	use ~~ intensively	5.03
use ~~ excessively	7.74	use ~~ exclusively	5.32

The results presented start with the descriptive statistical data analysis, including general descriptive statistics about the Verb-Adverb collocations in the Arab EFLL and NBES corpora (§Table 4-19).

Table 4-19: Overall descriptive statistical results for the Verb-Adverb collocations

	No. of collocations	Maximum LL score	Mean	Standard deviation	Skewness	Kurtosis
Arab EFLLs	146	990.39	32.95	113.37	7.181	54.982
NBESs	651	17133.50	88.28	771.42	18.846	390.412

There were 146 Verb-Adverb collocations for the Arab EFLLs and 661 Verb-Adverb collocations for the NBESs. The number of collocations extracted for the NBESs was four times that of the Arab EFLLs. The highest LL score for the NBESs is also much higher for many collocations than for the Arab EFLLs with the maximum LL score reaching 17133.50 (M = 88.28, SD = 771.42), whereas the LL highest score was 990.39 for the Arab EFLLs (M = 32.95, SD =113.37). The minimum LL score was set to 5 for both groups.

Table 4-20: Percentage of Verbs of with threshold adverb *collocates* in Verb-Adverb Collocations

	No. of verbs that has an adverb collocate in the corpus	%
Arab EFLLs	53	26.5%
NBES	151	75.5%

Table 4-20 illustrates the number of verbs that had adverb collocates that emerged in the extraction phase from the Arab EFLL and NBES corpora. There were 53 verbs that had adverb collocates within the threshold, in the AEFLL Corpus accounting for 26.5% of the total 200 verbs searched for. However, there were 151 verbs that had adverb collocates within the threshold, in the NBESs Corpus accounting for 75.5% of the total verbs searched for. The number of collocates extracted for Arab EFLLs appears small compared to the NBESs, but that is because many of the collocations were below the frequency threshold used.

To carry out the most appropriate statistical significance test, the distribution of the data had to be determined. One indicator for the shape of the distribution is the results for skewness and kurtosis, which are given in (§Table 4-19). The

results indicate that the distribution was positively skewed, amounting to 7.18 for Arab EFLLs and 18.84 for NBESs. The results for kurtosis suggest that the distribution was heavy-tailed compared to the normal distribution, which was found to be 54.98 for Arab EFLLs and 390.41 for NBESs. The two histograms in Figure 4-9 show the distribution for the Arab EFLL and NBES data.

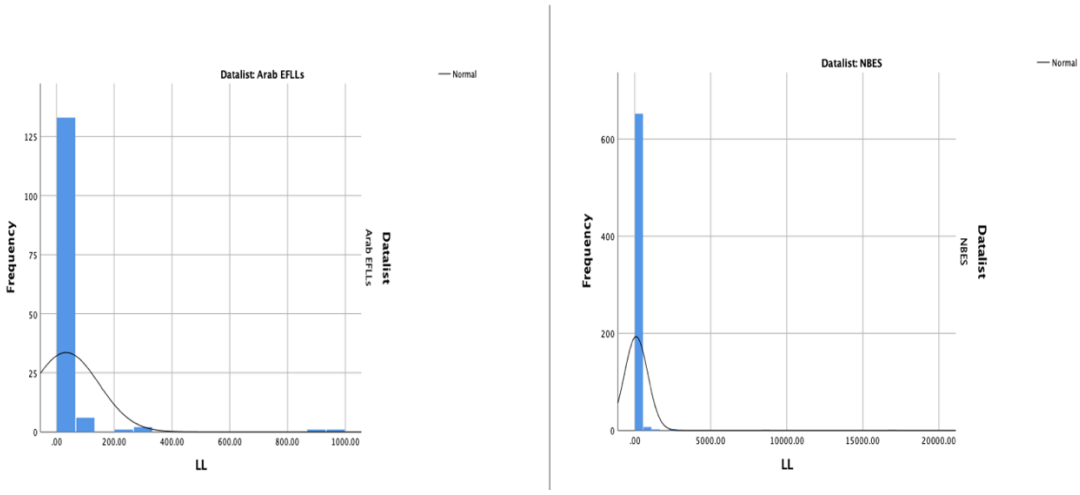


Figure 4-9: Histogram showing the distribution of Verb-Adverb collocations for Arab EFLL and NBES corpora

The histograms in Figure 4-9 illustrate that both data sets are positively skewed suggesting a non-normal distribution for both the Arab EFLLs and NBESs datasets although they have different shapes. Therefore, those two results confirm that this kind of distribution suggested a need for the non-parametric Mann-Whitney *U* test to be used.

The Mann-Whitney *U* test has four assumptions that must be met before applying the test (§3.4.3.2). The first assumption is that the Verb-Adverb collocation data fall into two independent groups; if so, the Mann-Whitney *U* test is the most suitable test to examine the differences in using Verb-Adverb collocations for the Arab EFLLs and the NBESs. The second assumption is that the data for the Verb-Adverb collocation have been observed and collected independently, meaning there is no relationship between the two groups (i.e., Arab EFLLs and NBESs). The third assumption is that the collocation data are extracted based on their LL

scores, and in the form of ordinal data in order to use the IntelliText web interface. The data distribution is the fourth and most fundamental assumption of the Mann-Whitney *U* test. As seen in Figure 4-9, the histogram shows that the distribution of the Verb-Adverb collocation data is positively skewed, which means that they are not normally distributed. The Mann-Whitney *U* test assumes a non-normal distribution, but the data distribution must be similar. Therefore, concluding the kind of data distribution based on the histograms alone is not enough; a homogeneity of variances test is needed, such as the Levene Statistic test. The Levene Statistic test was used to determine whether the analysis of variances and the distribution of scores around the median were considered equal (Davis, 2010). The homogeneity of variances test results are given in Table 4-21.

Table 4-21: Levene's Test of Homogeneity of Variances for Verb-Adverb collocations

Test of Homogeneity of Variances					
Verb-Adverb collocation LL	Based on Median and with adjusted *df*	Levene Statistic	*df*1	*df*2	*Sig*.
		0.747	1	656.149	0.388

Table 4-21 illustrates the SPSS output table for Levene's homogeneity of variances statistic results. The table shows the results based on the median and with adjusted *df*. Levene's statistic test showed that the assumption of homogeneity of variances was not violated for the Verb-Adverb collocation data ($F(1.656) = 0.747$, $p = 0.388$). Thus, the Mann-Whitney *U* test was used, and the results are illustrated in Figure 4-10.

Mann-Whitney Test

Ranks

	Verb-Adverb Collocation	N	Mean Rank	Sum of Ranks
LL	Arab EFLLs	146	385.21	56240.50
	NBES	651	402.09	261762.50
	Total	797		

Test Statistics[a]

	LL
Mann-Whitney U	45509.500
Wilcoxon W	56240.500
Z	-.801
Asymp. Sig. (2-tailed)	.423

a. Grouping Variable: Verb-Adverb Collocation

Figure 4-10: Mann-Whitney *U* test for the use of Verb-Adverb collocation for Arab EFLLs and NBESs

Figure 4-10 above includes two tables for the Mann-Whitney *U* test. The upper table indicates which data set has the higher median rank (*Mdn*[7]). The second table in Figure 4-10 illustrates the Mann-Whitney *U* statistic results. The results for the median rank for the Verb-Adverb collocation were quite similar, amounting to (*Mdn* = 385.21) for the Arab EFLLs and (*Mdn* = 402.09) for the NBESs. It can be concluded from the results of the Mann-Whitney *U* test that there were no statistically significant differences in the use of Verb-Adverb collocations between Arab EFLLs and NBESs (*U* = 45509, *p* = 0.423). Therefore, this result indicates that here we accept the null hypothesis for the use of the Verb-Adverb collocation amongst Arab EFLLs and NBESs.

4.2.1.3 Adjective-Noun Collocation Results

In this section, I illustrate and report the results of the frequency analysis of Adjective-Noun collocations. I also include a review of the statistical significance tests conducted to determine whether there was a difference between Arab EFLLs and NBESs for the use of Adjective-Noun collocations. The results are based on the LL scores of the collocations extracted for the Arab EFLLs and the NBESs (§Table A-21 and Table A-22 in the Appendix for the collocations

[7] The upper table in Figure 4-10 has a column for median rank, but it is automatically generated in SPSS as the mean rank.

extracted). Table 4-22 illustrates some of the Adjective-Noun collocations extracted from the Arab EFLL and NBES corpora. The results of these collocations indicate that both have similar collocations. The large number of collocations used by the Arab EFLLs could indicate that this collocation type is the easiest.

Table 4-22: Adjective-Noun collocations in the Arab EFLL and NBES corpora

Arab EFLLs	LL score	NBESs	LL scores
good ~~ way	49	good ~~ quality	202
good ~~ idea	15.58	good ~~ way	201.8
beautiful ~~ place	128.37	high ~~ rate	373.48
high ~~ rate	12.14	high ~~ temperature	298.61
high ~~ cost	9.64	high ~~ price	253.1
high ~~ price	7.48	high ~~ degree	196.09
negative ~~ way	9.42	major ~~ role	70.44
negative ~~ aspect	8.4	Equal ~~ opportunity	66.07
negative ~~ influence	8.4	Equal ~~ right	28.17
negative ~~ way	9.42	reasonable ~~ doubt	49.1
public ~~ school	52.99	reasonable ~~ doubt	49.1
public ~~ transport	32.77	public ~~ transport	262.71
large ~~ number	42.36	large ~~ number	562.04
large ~~ quantity	20.45	large ~~ quantity	95.91
Human ~~ life	10.88	human ~~ right	909.5
human ~~ race	10.47	human ~~ nature	177.18
bad ~~ habit	22.95	different ~~ aspect	103.52
major ~~ problem	22.18	low ~~ level	448.45
low ~~ population	7.49	low ~~ cost	311.08
basic ~~ need	12.21	basic ~~ need	27.16
traditional ~~ marriage	6.46	traditional ~~ view	66.14
future ~~ reference	6.68	major ~~ problem	96.62
easy ~~ access	16.45	major ~~ factor	86.43
specific ~~ type	15.66	huge ~~ amount	118.76
medical ~~ treatment	18.92	huge ~~ number	31.31

Using SPSS, the descriptive statistical results of the Adjective-Noun collocation are given in Table 4-23.

Table 4-23 Overall descriptive statistical results for the Adjective-Noun collocations by Arab EFLLs and NBESs

	No. of collocations	Maximum LL score	Mean	Standard deviation	Skewness	Kurtosis
Arab EFLLs	666	1417.08	21.94	65.0	15.98	324.14
NBESs	3258	1978.25	40.32	74.63	10.45	192.81

The table illustrates the general differences between the two corpora, providing the highest and lowest LL scores. The number of collocations for the Arab EFLLs is 666, and 3258 for the NBESs. The number of collocations is much higher for the NBESs than for the Arab EFLLs. However, this was also the first time that the total number of extracted collocations for the Arab EFLLs reached as high as 666 compared to the number of collocations extracted for the other three collocation types investigated in this thesis. The LL scores for the NBESs were higher than the Arab EFLLs for most of the collocations, with the maximum LL score being 1978.25 (M = 40.32, SD = 74.63), whereas for the Arab EFLLs the maximum LL score was 1417.08 (M = 21.94, SD =65.0). For this Adjective-Noun collocation, the LL scores were high for both the Arab EFLLs and the NBESs. The lowest LL score was set to 5 for both corpora.

Table 4-24: Percentage of Adjectives with threshold *noun collocates* in Adjective-Noun collocations

	No. of Adjectives that has a noun collocate in the corpus	%
Arab EFLLs	152	76%
NBES	192	96%

The collocation extraction phase has shown that Arab EFLLs tend to have more examples of Adjective-Noun collocations. Table 4-24 illustrates the number of adjectives that had a noun collocate in the Arab EFLL and NBES corpora. The Arab EFLLs had 152 adjectives with a noun collocate within the threshold, accounting for 76%, whereas the NBESs had 192 adjectives with a noun collocate within the threshold, accounting for 96% of the total adjectives. The percentage of Adjective-Noun collocations for Arab EFLLs was higher than for the other three collocation sets investigated in this study.

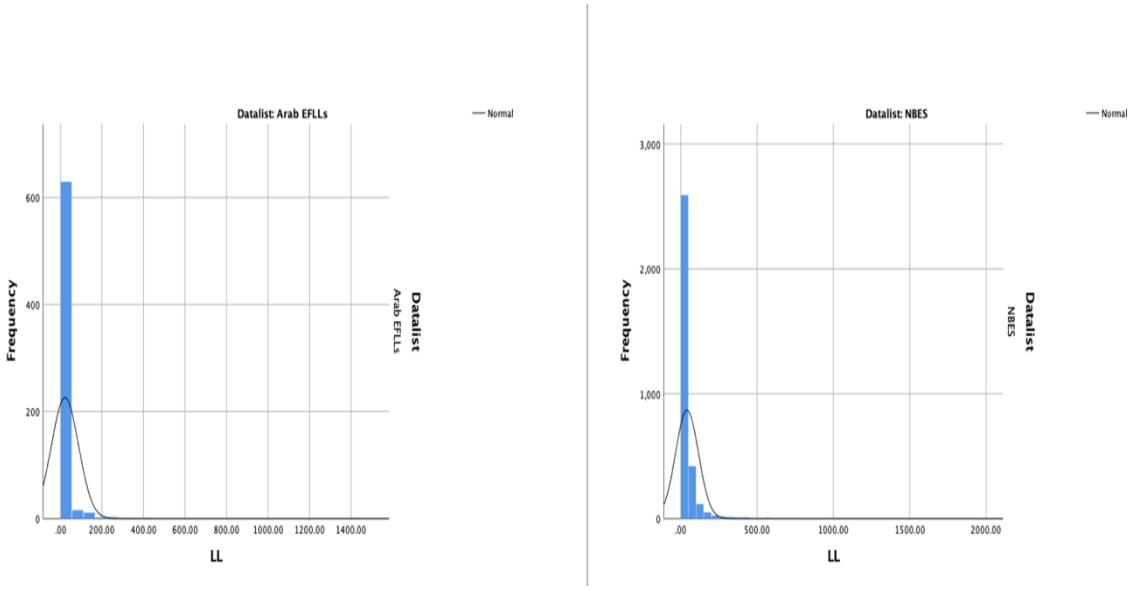


Figure 4-11: Histograms showing the distribution of the Adjective-Noun collocation for Arab EFLLs and NBESs

The histograms in Figure 4-11 show the distribution of Adjective-Noun collocations for the Arab EFLLs and NBESs. The skewness and kurtosis results are illustrated in (§Table 4-23), indicating the distribution was positively skewed, accounting for 15.98 for Arab EFLLs and 10.45 for NBESs. The results for kurtosis indicated that the distribution was heavy tailed compared to the normal distribution, which was found to be 324.14 for Arab EFLLs and 192.81 for NBESs. It can be concluded from the two graphs and the skewness and kurtosis results that the Adjective-Noun collocations for Arab EFLLs and NBESs do not have a normal distribution and are positively skewed, suggesting a need for a non-parametric test.

As with the previous collocations, the non-parametric Mann-Whitney *U* test was the most suitable test. However, the two histograms in Figure 4-11 differ in their distribution shape, whereas the essential assumption for conducting a Mann Whitney test is that their shapes should be similar. Therefore, I carried out Levene's homogeneity of variances test to determine whether the assumption of a similar distribution shape has not been violated.

Table 4-25 Levene's Test of Homogeneity of Variances for Adjective-Noun collocations

Test of Homogeneity of Variances					
Verb-Adverb collocation LL	Based on Median and with adjusted *df*	Levene Statistic	*df*1	*df*2	*Sig*.
		19.309	1	3893.506	0.000

According to the results in Table 4-25, the significance of Levene's statistic was (p = 0.000 > 0.05). This means that the assumption of an equal distribution test between the Aab EFLLs and the NBES was not satisfied (F (1.3893) = 19.309, p = 0.000). This implies that the non-parametric Mann-Whitney *U* test could not be used in a valid way. Therefore, an independent sample t-test with bootstrapping as the estimation technique had to be carried out instead.

Table 4-26: The Independent Sample t-Test of the Adjective-Noun Collocation

	No. of Collocations	**Mean**	**Std. Deviation**	**t-test**			
				F	***Sig.***	***t***	***df***
Arab EFLLs	665	21.9838	65.04914	26.413	0.000	-6.467	1054.152
NBES	3254	40.3626	74.66947				

An independent sample t-test was conducted to compare the LL scores of the Adjective-Noun collocation for Arab EFLLs and NBESs. Table 4-26 illustrates the SPSS output for the independent sample t-test using bootstrapping as the estimation technique. There was a significant difference in the LL scores of the Adjective-Noun collocation for Arab EFLLs (M = 21.9838, SD= 65.04914) and NBESs (M = 40.3626, SD= 74.66947); t (1054.152) = -6.467, p = 0.000. Therefore, based on the p-value of the independent t-test, we reject the null hypothesis and conclude that there is a significant difference between Arab EFLLs and NBESs in the use of Adjective-Noun collocations.

Summary

To conclude this section, I have created a table that summarises the overall statistical findings for the collocations investigated in this study (§Table 4-27).

Table 4-27: Overview of the statistical significance testing for the collocations.

Collocations	Arab EFLLs	NBESs	Statistical test	Statistical significance test Result
Adverb-Adjective Collocations	20	20	Kruskal-Wallis H test	Rejected H_0 at *p*-value >0.000.
Verb-Noun Collocations	145	393	t-test with bootstrapping as an estimation technique	Rejected H_0 at *p*-value >0.005.
Verb-(intermediate)-Noun collocations	31	31	Mann-Whitney *U* test	Rejected H_0 at *p*-value >0.000.
to + Verb-Noun collocation	13	14	Mann-Whitney *U* test	Failed to reject H_0 and accept H_A at *p*-value > 0.550. i.e., no statistically significant differences between Arab EFLLs and NBESs.
to + Verb-(intermediate)-Noun collocation	7	13	Mann-Whitney *U* test	Rejected H_0 at *p*-value >0.005.
Verb-Adverb Collocations	146	651	Mann-Whitney *U* test	Failed to reject H_0 and accept H_A at *p*-value > 0.423. i.e., no statistically significant differences between Arab EFLLs and NBESs.
Adjective-Noun Collocations	665	3254	t-test with bootstrapping as an estimation technique	Rejected H_0 at *p*-value >0.000.

The results of the comparison for the Adverb-Adjective collocations, previously presented in Alshammari (2021), reveal that there is a statistically significant difference between Arab EFLLs and NBESs. Overall, the results of the comparison between Arab EFLLs and NBESs were statistically significant for all the collocations except for two kinds of collocation for which the null hypothesis failed to be rejected: (1) for the *to* + Verb-Noun collocations, at p-values set at > 0.550; (2) for the Verb-Adverb collocations, at *p* >0.423. Interestingly, although

the infinitive is claimed to be an absent category in Arabic, the results for the *to* + Verb-Noun collocations indicated that there is no significant difference between Arab EFLLs and NBESs.

4.2.2 Measuring word choice complexity in collocation use through the CEFR

Having presented the statistical analysis of the use of lexical collocations, I now present the results for measuring word choice complexity within the collocations used by the Arab EFLLs and NBESs. In this section I use the Common European Framework of Reference for Languages (henceforth, CEFR[8]) to measure learners' language proficiency in using lexical collocations. The results are quantitative in nature including numbers of how their use of words within collocations are categorised based on the CEFR. The CEFR proficiency-graded word lists were used to identify how the word choice complexity within the Arab EFLLs' use of lexical collocations differed from that of the NBESs. The CEFR list comes from the KELLY project which combined several corpora (Kilgarriff et al., 2014). The results of KELLY are based on the competencies of learners' performance. These competencies reflect how well learners can use and understand English at different levels of proficiency. Using CEFR, allows me to only compare only the choice of single lexical words. The results for each collocation type will be presented in tables. The upper part of the tables will show the CEFR levels for the node words within the actual collocations. The bottom part of the table will show the CEFR levels for the collocates within each collocation. Below, I make the suggestion that the CEFR can be extended to measure collocations (§4.2.2.1). Figure 4-12 summarises the CEFR levels and what they mean:

CEFR level	CEFR category
A1 and A2	Basic User
B1 and B2	Independent User
C1 and C2	Proficient User

Figure 4-12 CEFR levels

[8]Some useful links about CEFR:
https://en.wikipedia.org/wiki/Common_European_Framework_of_Reference_for_Languages

Table 4-28: CEFR categorization of the Adverb-Adjective collocation

	CEFR	Arab EFLLs		NBESs	
		F	%	F	%
Adverbs within Adverb-Adjective collocations	A1	20	100	19	70.4
	A2	0	0	1	3.7
	B1	0	0	4	14.8
	B2	0	0	0	0
	C1	0	0	7	11.1
	C2	0	0	0	0
	UN[9]	0	0	0	0
	Total	20	100	27	100
Adjectives within Adverb-Adjective collocations	A1	12	60	16	59.3
	A2	6	30	6	22.2
	B1	1	5	5	18.5
	B2	1	5	0	0
	C1	0	0	0	0
	C2	0	0	0	0
	UN	0	0	0	0
	Total	20	100	27	100

I will begin by discussing the results for the Adverb-Adjective collocation illustrated in Table 4-28. Table 4-28 reveals that all the adverbs used by the Arab EFLLs were A1 level and 70% of the NBESs' choices were of the same level. The NBESs used some A2, B1 and C1 CEFR level adverbs. Therefore, part A (i.e., the node word, the adverb) of the Adverb-Adjective collocation for both groups seems to be mainly basic verbs classed under the lowest CEFR word level. As for the adjective choices within the Adverb-Adjective collocation, most are in the A1 CEFR group, accounting for 60%, for both Arab EFLLs and NBESs. Other vocabulary levels present in both groups are A2 and B1 adjectives. The Arab EFLLs had one adjective choice at the B2 CEFR level. The results indicate that both groups are similar to each other in terms of the complexity of their word choices for the adjectives within the Adverb-Adjective collocation. However, for the Arab EFLLs, the choice of adverbs within the Adverb-Adjective collocations is only at the A1 CEFR level, while the NBESs were using adverbs from other CEFR levels.

[9] Some words in the CEFR function in IntelliText were left unclassified (UN) according to CEFR.

Table 4-29: CEFR categorization the Verb-Noun Collocations

	CEFR	Arab EFLLs		NBESs	
		F	%	F	%
Verbs in Verb-Noun Collocations	A1	44	71	39	73.6
	A2	7	11.3	5	9.4
	B1	11	17.7	9	17
	B2	0	0	0	0
	C1	0	0	0	0
	C2	0	0	0	0
	UN	0	0	0	0
	Total	62	100	53	100
Nouns in Verb-Noun Collocations	A1	54	38	87	22.3
	A2	28	19.7	86	22.1
	B1	23	16.2	95	24.4
	B2	7	4.9	31	7.9
	C1	5	3.5	16	4.1
	C2	4	2.8	24	6.5
	UN	21	14.8	51	13.1
	Total	142	100	390	100

Table 4-29 illustrates the results for the complexity of word choice within the Verb-Noun collocations. The results for the verbs based on CEFR categorizations reveal that the verbs used come from three CEFR levels, A1, A2, and B1. The most frequent CEFR level for the verbs is A1, amounting to 71% for the Arab EFLLs and 73.6% for the NBESs. As for the noun collocates within the Verb-Noun collocations, the results suggest that the most frequent level for the noun choices was also A1 for Arab EFLLs and NBESs, respectively. There were some instances of C1-C2 level nouns, amounting to 6.3% for the Arab EFLLs and 10.6 for the NBESs. The results suggest that the Arab EFLLs are not all limited to basic vocabulary choices as it is clear that they use some vocabulary under the higher CEFR levels. Though the numbers are small, it could be just one or two higher-level students. However, it is not possible to verify this with the corpus data available.

Table 4-30 CEFR categorization for the Verb-Adverb collocations

	CEFR	Arab EFLLs		NBESs	
		F	%	F	%
Verbs in Verb-Adverb collocations	A1	38	71.7	91	62.3
	A2	6	11.3	19	13
	B1	7	13.2	35	24
	B2	2	3.8	1	0.7
	C1	0	0	0	0
	C2	0	0	0	0
	UN	0	0	0	0
	Total	53	100	146	100
Adverbs in Verb-Adverb collocations	A1	88	60.3	285	43.8
	A2	28	19.2	103	15.8
	B1	10	6.8	71	10.9
	B2	6	4.1	41	6.3
	C1	5	3.4	47	7.2
	C2	0	0	22	3.4
	UN	9	6.2	82	12.6
	Total	146	100	651	100

The results of the CEFR scores for the verbs used in the Verb-Adverb collocation are illustrated in Table 4-30. The verbs used by both Arab EFLLs and NBESs range between the A1 to B2 CEFR levels. Considering the number of verbs examined for the Verb-Adverb collocation, 83% of the Arab EFLLs' verbs were A1 and A2, i.e., basic English users. 71% of the verbs were classified under the category of A1 level. This is a high percentage as it equates to 38 out of the 53 total verbs used by the Arab EFLLs, showing that the Arab learners are using mainly basic verbs. The remaining verbs are at the B1-B2 level, amounting to 13.2% at B1 and 3.8% at B2, which are associated with users being at the CEFR "Independent" level, Overall, though, the results suggest the Arab EFLLs would be classified as basic English users. In comparison, while A1 verbs are also the most frequent for the NBESs, amounting to 62.3%, the NBESs had a higher percentage of B1 verbs, at 24%,.

The adverbs used in the Verb-Adverb collocation vary in CEFR level for both corpora (§Table 4-30). For the Arab EFLLs, the adverbs range between A1 and C1, with nine adverbs amounting to 6.2% not specified under CEFR. The Arab EFLLs' use of adverbs is mainly at basic English level according to the CEFR categorisation, in which 79.5% of the adverbs were at A1-A2 CEFR levels. A few

adverbs were categorised at B1 and B2 levels accounting for 6.8% and 4.1% of the total adverbs, respectively. There were some adverbs at C1 level, considered by CEFR as indicating proficient users accounting for 3.4% of the total adverbs used. However, the NBESs Corpus had higher percentages of B1-B2 and C1-C2 level adverbs, accounting for 17.2% and 10.6%, respectively (combining the scores for the relevant levels). The NBES had various adverbs considered to indicate intermediate to proficient users of English, unlike the Arab EFLLs' choice of adverbs that mainly showed them to be basic users of English. Therefore, based on the CEFR classifications for each component in the collocation, the Arab EFLLs' use of vocabulary within the Verb-Adverb collocation remains at the basic CEFR levels. A greater range of lexical choices were used by the NBESs, including vocabulary under the C2 CEFR level.

Next, I will discuss the complexity of the learners' vocabulary choices of within the Adjective-Noun collocation (§Table 4-31). The results of the CEFR scores for adjectives used in the Adjective-Noun collocation are illustrated in Table 4-31. The numbers of adjectives at the basic CEFR level for the Arab EFLLs and NBESs were similar, accounting for 61% and 57.7%, respectively. Also, the use of intermediate CEFR nouns was comparatively similar for Arab EFLLs and NBESs, accounting for 35.6% and 39.7%. As for what the CEFR categorise as advanced nouns used by proficient English users, there were only three nouns for the Arab EFLLs in the C1 and C2 CEFR levels.

Table 4-31: CEFR categorization for the Adjective-Noun collocations

	CEFR	Arab EFLLs		NBESs	
		F	%	F	%
Adjectives in Adjective-Noun collocations	A1	60	41.1	60	30.9
	A2	29	19.9	52	26.8
	B1	45	30.8	67	34.5
	B2	7	4.8	10	5.2
	C1	2	1.4	1	0.5
	C2	1	0.7	2	1
	UN	2	1.4	2	1
	Total	146	100	194	100
Nouns in Adjective-Noun collocations	A1	285	42.9	684	21
	A2	99	14.9	616	18.9
	B1	182	27.4	1117	34.3
	B2	31	4.7	244	7.5
	C1	16	2.4	111	3.4
	C2	14	2.1	149	4.6

	UN	38	5.7	334	10.3
	Total	665	100	3255	100

For the noun collocates within the Adjective-Noun collocation, the basic CEFR level is the highest for the Arab EFLLs, accounting for 57.8%. In comparison, the NBESs had more nouns classed at the intermediate English user level, accounting for 41.8%. Intermediate level nouns (i.e., CEFR B1-B2) were at the highest in the Adjective-Noun collocation for the Arab EFLLs compared to the other three collocation types investigated in this study, accounting for 32.1%. As for the advanced use of nouns, the Arab EFLLs had 4.5%, and the NBESs had 8% for nouns at the C1-C2 CEFR levels. Some nouns were not listed in the CEFR, accounting for 5.7% and 10.3% for the Arab EFLLs and NBESs respectively.

Overall, the CEFR categorisation for the Adjective-Noun collocation has shown that the Arab EFLLs had a greater use of intermediate to advanced noun collocates in this collocation type. The NBESs' lexical choices within the Adjective-Noun collocation included nouns at various CEFR levels. The highest CEFR level for noun collocates after A1 was B1 for both the Arab EFLLs and the NBESs.

4.2.2.1 CEFR System for Collocations (CEFR-Coll)

As stated above, one of the aims of this thesis was to examine the complexity of learners' word choices when using collocations. One possible way of achieving this was to measure their levels based on the CEFR. However, this scheme is meant to categorize the level of single words based on 6 CEFR levels (i.e., C2, C1, B2, B1, A2, A1). Therefore, the research conducted for this PhD will contribute an extension to the CEFR that would include a measure for collocations so that vocabulary profiling does not have to be restricted to measuring single words. This modification will create a new proficiency and complexity measure for collocations. I refer to this system as CEFR-Coll, with the '-Coll' standing for "collocations". CEFR-Coll is a system for assessing language proficiency through collocation-informed vocabulary profiling. Currently, it is impossible to measure the complexity of learners' lexical choices in the formation

of collocations because there is no framework designed for measuring collocation complexity. Therefore, this is an attempt to modify the existing CEFR so that it can categorize learners' English levels from collocations based on their choices of single, lexical words made within the collocations. It should include the single words that are not part of collocations. Single words will still be counted because they still contribute to profiling the vocabulary of the learner.

There have been previous attempts to classify words according to the CEFR levels, such as the English Profile project[10], which provides accurate information about the CEFR levels of words. This is a useful resource for language teachers and learners to understand which CEFR level language learners' words are at. The English Vocabulary Profile (EVP) tool assigns a CEFR level to certain collocations that have idiomatic meanings, where the combination of words creates a meaning that is not directly predictable from the individual constituent words. For instance, the verb 'make' is classified at the A1 CEFR level. However, if it is used to indicate "achieving a goal", such as in 'she made it to the airport', this suggests that the meaning is at the B1 CEFR level. The EVP also includes common phrases and collocations, such as the word 'central heating', which it assigns to CEFR level B1. The example of 'central heating' as a collocation will be used in my analysis to support the applicability of the CEFR-Coll method. I present the methodological framework I propose for CEFR-Coll in Figure 4-13 below.

The initial results of this method were positive. To return to the preceding example of 'central heating', which is categorised in the English Vocabulary Profile project as CEFR level B1 ("independent English language user"), it is not clear what criteria have been used to assign it to B1. Using an equation, I have established for the proposed CEFR-Coll, I will determine a more realistic score for the collocation. As single words, the term 'central' has a CEFR level of C1 (Ranked CEFR level=6), while the word 'heating' has a CEFR level of A2 (Ranked CEFR level=3). The indicative values for the two words in the collocation are therefore

[10] http://englishprofile.org/resources

6 and 3, respectively. To calculate the CEFR-Coll score, we take an 'adjusted average' of these two proficiency level values by first summing up them, then deducting 1 to take into account non-list words, before dividing by 2. In this case, the calculation is [(6 + 3 -1) □ 2], so 'central heating' gets a score of 4, which indicates that the collocation belongs to CEFR-Coll-Ind (Independent user) By taking into account the different proficiency levels of the constituent words, the proposed CEFR-Coll score more accurately represents the proficiency level of the collocation as a whole.

It must It must be mentioned that CEFR-Coll is a preliminary (work-in-progress) framework and needs further modification and testing to confirm its applicability to collocations with more than two words. To improve the quality of this suggested extension to the CEFR, there will need to be a collaboration with dictionary compilers to assess the possible collocations. Also, lexical complexity is an important indicator for a collocation. For example, *take, give, after or up* are very basic words, however, when these basic verbs are combined into phrasal verbs like *take after* or *give up* they become more lexically sophisticated, as the resulting collocations carry specific and nuanced meanings that go beyond the individual words. For example, in Sharoff et al. (2013) ranking was determined through LL scores of the individual components. Therefore, an alternative approach is ranking by LL scores, which is also a robust method to determine CEFR levels within collocations. This consideration will be incorporated into the ongoing work of CEFR-Coll for future improvements.

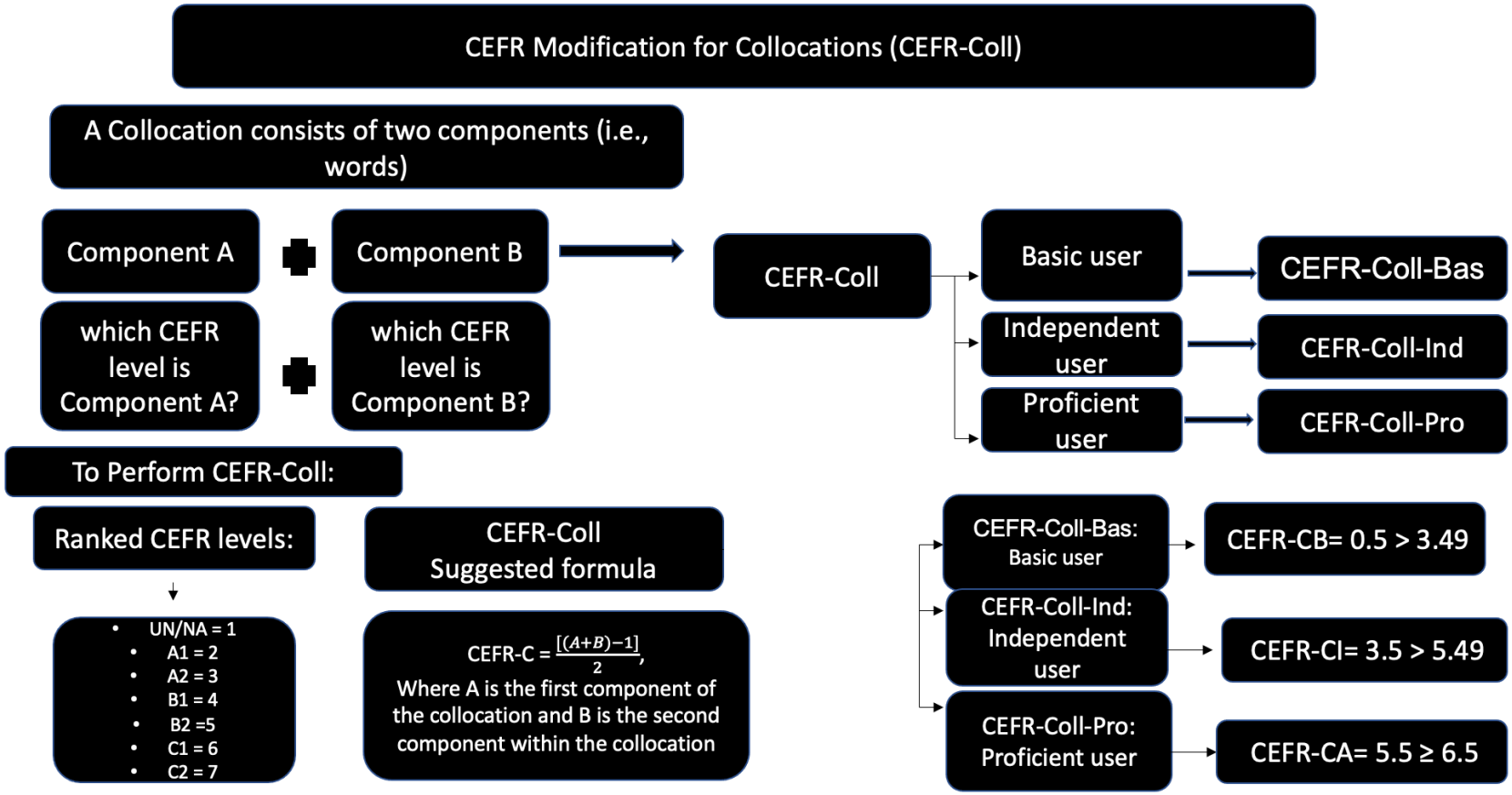


Figure 4-13: Suggested Methodological Framework for CEFR-Coll

4.2.3 Qualitative results: Contrastive Error-analysis (CEA)

In this section, I highlight the results of a contrastive error-analysis approach to the use of the four lexical collocations from the AEFLL Corpus. The results of the native speakers' judgment of the collocations' acceptability will be given. Then, I will illustrate the results of the contrastive-error analysis by identifying the most common errors in the collocations investigated.

4.2.3.1 Native-speaker Rater Judgment of the Collocations

In this sub-section, I will provide the raters' judgment of the collocations. Raters were used to increase the reliability of the error analysis (§3.5.2 for more information about the native raters). Due to the number of collocations extracted for each collocation type, it was hard to examine the acceptability of all of the collocations. Therefore, I used a sampling approach to determine the acceptability rates. The acceptability of collocations was determined using raters who are native British English speakers. The raters were a mix of male and female, and their ages were similar to those of the students whose essays are in the Arab EFLLs Corpus. This ensures that the raters are familiar with the collocations commonly used by that age group. The raters are also students of linguistics, which means they have a strong interest in the English language and are skilled at identifying language issues. As university students, they are a good match for the informants in terms of age and education level. I must mention that Covid-19 has an impact on collecting the responces from the native raters. There were many delays in getting the feedback and communicating with the two raters.

The raters were given a sample of 142 collocations, accounting for 25% of the Verb-Noun collocations, 10% of the Verb-Adverb collocations, and 5% of the Adjective-Noun collocations from the total collocations extracted for this study. I included all instances of Verb-(Intermediate)-Noun collocations, *to* + Verb-Noun collocations, and *to* + Verb-(Intermediate)-Noun collocations because their total numbers of collocations extracted were few in number amounting to 31, 13, and 7 for each collocation type respectively. The variance in the percentages is

attributable to several factors. First, those collocations that were not found in the two reference corpora were not large in number, reducing the number of collocations that had to be evaluated. Second, to avoid random ratings, it was not advisable to give a large quantity of data to the raters (Hughes, 2003; Langsford et al., 2018). Third, the number of Adjective-Noun collocations extracted was 664; therefore, 5% captured a substantial and sufficient amount of data to be evaluated.

Table 4-32: Overview of the native-speaker judges' response to the collocations investigated in this study

Collocations/Acceptability score	Overall number of collocations	Unacceptable (%)	Questionable (%)	Acceptable (%)
Verb-Noun collocations	36	11.11%	19.44	69.45%
Verb-(Intermediate)-Noun collocations	31	19.35%	9.6%	70.96%
to **+ Verb-Noun collocations**	13	16%	25%	83.3%
to **+ Verb-(Intermediate)-Noun collocations**	7	0%	42.85%	57.14
Verb-Adverb collocations	12	41%	8.3%	50.7%
Adjective-Noun collocations	43	16%	34.88%	49.12%
Total	142			

Table 4-32 illustrates the results of the raters' acceptability judgments. The first noteworthy aspect is the generally high percentages of acceptable collocations, with the *to* + Verb-Noun collocations at 83.3%, followed by the Verb-(Intermediate)-Noun collocations at 70.96%, and Verb-Noun collocations at 69.45%, while the Adjective-Noun collocations had the lowest percentage at 49.12%. The highest rate of unacceptable collocations was 41% for the Verb-Adverb collocations. Collocations that were viewed as questionable were high in the *to* + Verb-(Intermediate)-Noun group at 42.85% and the Adjective-Noun group at 34.88%.

The highest rate of unacceptable collocations was found to be in Verb-Adverb collocations, including the following: ‘see recently’, ‘mention previously’, ‘do necessarily’, ‘produce environmentally’, and ‘*drive speedly’. I must discuss the fact that the collocations ‘see recently’ and ‘mention previously’ were judged as unacceptable though, given the right context, other native speakers might see them as acceptable. Therefore, this indicates an important issue about the expected outcomes of the rater judgment. In this study, my expectation was that the native raters would perceive an incorrect collocation as correct when presented without context. However, the results showed the opposite effect, where a correct collocation was instead considered incorrect without the context. This finding highlights the importance of context in determining the acceptability of collocations for native speakers. Some examples of unacceptable Verb-Noun collocations include the following ‘want humanity’, ‘keep silence’, ‘moving castle’, ‘make revision’, ‘make exercise’. Adjective-Noun collocations that were not regarded as acceptable include the following: ‘most urban dwellers’, ‘personal computer revolution’, ‘right for self-expression’, ‘violent media’, ‘easy victims’, ‘smart device detriment’.

Besides carrying out the survey, the English native-speaker judges were asked to provide comments explaining why they selected some collocations as questionable. Their responses were either that the collocations did not sound familiar, or their meaning depended on the context and “dialects used if they are acceptable”. As to why they had selected some collocations as not acceptable, the English native-speaker judges claimed that they regarded these collocations as not acceptable because the collocations were “not very common phrases used, so they do not sound acceptable when saying them out loud” or “they do not sound too familiar”. Therefore, although some context was giving, providing more context regarding how collocations were used would improve the acceptability judgment results. This was not possible, however, due to the large amount of data. Therefore, the collocations that were judged as being unacceptable or questionable were considered again in the contrastive-error analysis section.

4.2.3.2 The Contrastive-Error analysis of the Lexical Collocations

In Section (§4.2.1), I discussed the statistical results of the frequency-based analysis for the four collocation types. In this sub-section (§4.2.3.2), I will present the contrastive error analysis and highlight the most common errors made by Arab EFLLs in their use of the four lexical collocations. The particular method of contrastive error analysis (henceforth CEA) deployed was adapted and modified for the purpose of this thesis from Lado (1957) and Gass and Selinker (2008) and was explained in (§Chapter 2 and section 2.2.2). Table 4-33 summarises the different types of error in the use of collocations by Arab EFLLs. These errors are classified into three categories: interlingual errors, intralingual errors and overlap errors (§2.2.1). The last are errors that are caused by the rules of both L1 and L2 being involved in the same collocation.

Table 4-33: Contrastive Error analysis types in the AEFLL Corpus

CEA/ Types of errors	Interlingual errors %	Intralingual errors %	Overlap errors %
Adverb-Adjective collocations	53%	37%	6%
Verb-Noun collocations	26.2%	38%	8.3%
Verb-Adverb collocations	33.3%	16.66%	0
Adjective-Noun collocations	2.05%	0%	6.9%

The first category is “interlingual errors” which are caused by L1 interference on the use of the target language. The interlingual errors included grammatical types of error underlying the use of the lexical collocations such as errors in tense, the sentential position of the collocations and word-order errors. The sentential position of the collocations explains the position of the collocations in the sentence. The second category is “intralingual errors”, caused by the rules if the target language including the use of subject-verb agreement errors, wrong lexical items within the collocations and confusion between similar words. The third type is the overlap error which consists of errors caused by the interference of the rules of the L1 and L2 on the use of collocations (§2.2.1 for more explanation of these types of errors). The overlap errors include those that are difficult to classify

under the interlingual or the intralingual errors. In the following sub-sections, I will illustrate the CEA of the four collocations investigated in this study.

4.2.3.2.1 CEA Results for the Adverb-Adjective Collocations

In this section, I will present the CEA results for the Adverb-Adjective collocations. The results for the Adverb-Adjective collocations were initially presented in Alshammari (2021). The total number of collocations found for the Arab EFLLs was 20. Table 4-33 illustrated the overall percentages for the CEA of the use of the Adverb-Adjective collocation. The results suggest that the Arab EFLLs had a high rate of interlingual errors amounting to 53%. The accuracy rate for the Adverb-Adjective collocations amounts to 4% of the total collocations. Interlingual errors were observed to occur in word order issues and the placement of collocations within sentences. The learners tend to place the Adverb-Adjective either at the beginning of a sentence or end a sentence with a collocation such as '*He brother is drive the small car is very happy'. There are many fragmentary sentences that illustrate the use of Adverb-Adjectives collocations, such as[1]:

- *We saw the kaaba we had very happy — when we saw the Kaaba, we were very happy
- *I am very nice wonderful — I am very nice and wonderful.
- *my holiday in Egypt. In started I was very happy— I was very happy when my holiday started in Egypt.

The second example above shows that the learner is describing him/herself as 'nice' which cannot be used in English because it sounds as moral judgment. The Arab EFLLs either avoid using a noun head determiner following the collocations or they tend to place the noun determiner in an inaccurate position in the sentence. For example:

- **Food its very good – it is very good food.*
- **In the dessert the weather the very good –The weather is very good in the dessert.*
- **I saw in France area very beautiful – I saw in France a very beautiful area*
- **It is a very green area and very big – it is a very big green area.*
- **Weather very hot. – very hot weather.*

[1] The examples on the right are corrections of the sentences found in the AEFLL Corpus.

Intralingual errors can be observed in subject-verb-agreement and wrong word order in what follows or precedes the collocations, amounting to 37% of the total collocations. The learners tend to place the modifier before the use of the Adverb-Adjective collocation. The following are some examples:

- **We was very happy.— we were very happy.*
- **We was so afred and very sad — we were so afraid and very sad.*
- **But it have some films are very good.— but it has some very good films.*
- **There is a swimming ball its very big — there is a very big swimming pool.*
- **The holiday it is very bad – it is a very bad holiday.*
- **Very wonderful my summer holiday. — my summer holiday was very wonderful.*

Also, the overuse of 'was' with the subject 'we' is a common issue that Arab EFLLs face in English. It is more related to subject-verb-agreement. This could be an interference of L1 mainly because of the use of كان(*kaan*) 'to be' in Arabic which is a 'temporal auxiliary verb' that serves to situate the action to a certain tense (Schulz, 2004, p.176). Therefore, this error could be attributed to both L1 and L2 which I refer to as an overlap error (it will be discussed in the following paragraph).

The third type of error in the use of the Adverb-Adjective collocations, overlap errors, amounted to 6% of the total collocations. They can be seen in the sentence structure surrounding the collocation itself, such as in the inaccurate use of relative clauses. Although these overlap errors are not related to the collocations themselves, I believe it is worth highlighting them because they are commonly made by Arab EFLLs. For example, there is an error in saying '*I come back from Alain I was very sad' which should be said as 'when I came back from Alain, I was very sad'. Another example is '*I go with my family, and I will be very happy' which should be said as 'when I go with my family, I will be very happy'. Another overlap error is seen in the use of double verbs in a sentence and the use of the verb 'to be'. For example:

- **I am is very happy— I am very happy.*
- **I am be very happy — I am very happy.*
- **I'm was very happy with my family — I was very happy.*

In this section, I have discussed the CEA results for the Adverb-Adjective collocations. In the following section, I will present the CEA for the Verb-Noun collocations.

4.2.3.2.2 CEA Results for the Verb-Noun Collocations

In this section, an overview of the CEA of the three subtypes of Verb-Noun collocation will be discussed: (1) Verb-(I)-Noun collocations with an intermediate (e.g. 'book an appointment'), (2) *to* + Verb-Noun collocations without an intermediate word in the combination (e.g. 'to improve quality') and (3) *to* + Verb-(I)-Noun collocations (e.g. 'to book an appointment'), where an intermediate word is included.

Overall, the results of the CEA for the use of Verb-Noun collocations revealed interlingual, intralingual, and overlap errors totalling 26.2%, 38.0%, and 8%, respectively. The highest error type among the Verb-Noun collocations was the intralingual error (§Table 4-33). 27.5% of the remaining collocations were deemed acceptable.

Table 4-34 shows the LL scores for the Verb-(I)-Noun collocation for the Arab EFLLs and NBESs. The results indicate that the LL scores are relatively low for both the Arab EFLL and the NBES data sets, which was not an anticipated result. The expected result was that NBESs would have higher LL scores than the Arab EFLLs for the use of this combination. However, the LL scores for both datasets are to some extent the same. Generally, the collocations are acceptable as separate combinations apart from the collocation 'take a gap' for the Arab EFLLs. The collocation 'take a gap' was also found to be correct as it was followed by the word *year* as a noun modifier as in 'taking a gap year'. This is not a common collocation for the NBESs who had only one instance of this collocation (e.g. 'I'm really glad I did take a gap year'). The remaining Arab EFLL collocations include some orthographic issues such as spelling mistakes such as **anti* instead of *auntie* and some capitalisation issues for proper nouns such as **paris.* Also, the learners tend to split the word 'football' into two words. These orthographic errors do not interfere with the intended meaning of the collocation.

Table 4-34: The LL ratios of Verb-Noun collocations allowing intermediates[2]

Arab EFFLs	LL	NBESs	LL
Do plastic surgery	6.04	take into account	61.39
Be no doubt	2.58	Do it explode	8.23
*See my anti	2	Be an ex-smoker	8.23
Go to Mecca	1.93	Be a non-smoker	6.94
Take a gap	1.76	Contact by telephone	5.66
*Go to paris	1.64	Regulate the productive	5.53
Watch T.V	1.52	Experience the normalization	5.14
Be a comedy	1.31	Be too Blunt	4.95
Be in favor	1.19	Be an isomorphism	4.66
Go to Salalah	0.88	Flout the maxim	4.61
Spend their leisure	0.87	Use political ideas	3.33
Boycott the election	0.79	Buy in yarn	3.26
Be gum infection	0.73	Beware the Cat	3.19
Burn fossil fuel	0.42	Be rigid designators	2.83
Go to Alain	0.41	Lose their birthright	2.83
Go in Hatta	0.39	Taking rights seriously	2.7
Have the ability	0.38	Be lose antemortem	2.34
Be with migration	0.38	Be a bijection	2.33
Go to Salalah	0.31	Represent the unrepresentable	2.33
Go to London	0.26	Calculate the molarity	2.14
Lead to isolation	0.25	Mark a watershed	1.94
Use the phrase	0.21	Bear the brunt	1.93
Take into consideration	0.19	Use in conjunction	1.72
Be a threat	0.13	Be a subfield	1.62
Distribute the burden	0.08	Be an affront	1.37
Go to England	0.07	Concern Human Understanding	1.31
Keep in touch	0.03	*Make the biocomposite	1.16
Take a taxi	0.03	Go into liquidation	1.12
Play foot ball	0.02	Be a subgroup	1.11
Get high mark	0.01	Be the quotient	1.02
Stay in contact	0.01	Be time consuming	0.97

The collocations that were rated as unacceptable include ‘be no doubt’, ‘be a comedy’, ‘spend their leisure’, ‘is gum infection’, ‘distribute the burden’, ‘get high mark’. Some of these collocations required an intermediate to make it correct

[2] Forms considered orthographic errors are followed by an asterisk.

such as the definite article ‘a*’* in ‘is a gum infection’. The British native-speaker raters judged 9.6% of the remaining Verb-(I)-Noun collocations as questionable including ‘do plastic surgery’, ‘*see my anti’, ‘take a gap’. As for the *to* + Verb-Noun collocations, there was a collocation that was rated as questionable (‘to do plastic surgery’). While the collocation ‘do plastic surgery’ may be used by native speakers of English, the more common verb used is the verb *perform* (§more clarification is given in Table 4-38). Additionally, the collocation **see my anti* is inaccurate due to a spelling issue. Lastly, the collocation ‘take a gap’ is part of a longer phrase such as ‘take a gap year’. I must clarify the reason for the high percentages of Verb-(I)-Noun collocations being judged as questionable and unacceptable. The percentages seem high compared to the Verb-Noun collocation results because the total number of the Verb-(I)-Noun collocations was 31 collocations in total.

Table 4-35: List of the *to* + Verb-Noun Collocation with LL scores for Arab EFLLs and NBESs

Arab EFFLs	LL	NBESs	LL
To play football	13.44	To take advantage	7.74
To have fun	5.94	To tackle truancy	6.48
To keep fit	4.79	To keep track	4.13
To face interaction	3.74	To maintain homeostasis	3.64
To take care	3.25	To survive anoxia	3.03
To facilitate communication	3	To do something	2.58
To quit smoking	2.6	To do anything	2.51
To do everything	1.56	To take place	2.51
To lose weight	1.4	To aid self-awareness	2.43
To take action	1.18	To raise awareness	2.43
To do anything	1.08	To commit suicide	2.22
To do something	0.99	To detect methylation	2.15
To stop smoking	0.76	To grant amnesty	2.07
		To avoid misunderstanding	1.4

Table 4-35 displays the results for the use of the *to* + Verb-Noun collocations in both corpora. The use of this collocation combination is accurate and does not show any errors for both data sets. The LL scores for the use of the *to* + Verb-

Noun collocation are relatively low for both and thus there is no difference between Arab EFLLs and NBESs. The results for the collocations indicate very specific topics discussed in the respective corpus such as the collocations with *anoxia* or *methylation* (see the examples in Table 4-35). This can lead to issues with robustness of lexical findings because these two collocations are highly context-dependent (Sharoff, 2020).

Table 4-36: List of the *to* + Verb-(intermediates)-Noun Collocation with their LL ratio for Arab EFLLs and NBES

Arab EFFLs	LL	NBESs	LL
To do plastic surgery	3.09	To take into account	17.06
To travel to Egypt	2.11	To calculate the molarity	8.53
To save the Earth	1.17	*To make the biocomposite	5.32
To decrease the number	0.84	To be a bat	4.59
To watch a movie	0.35	To operate the crane	3.55
To resolve the issue	0.2	To be a flagship	3.45
To resolve this issue	0.04	To achieve commercial sterility	3.29
		To commit for contempt	2.21
		To live under obedience	1.93
		To pay its creditor	1.81
		To kill in self-defence	1.56
		To keep a diary	1.33
		To work in partnership	1.25

Table 4-36 illustrates the results for the *to* + Verb-(intermediates)-Noun collocation. This table differs from Table 4-35 in that it includes an intermediate word within the collocations, such as the article '*a*' in the collocation *to make a payment.* The results illustrate a limited use of Verb-Noun collocations in the infinitive form *to*, which is apparent in the limited number of collocations extracted in both corpora and in the LL score being relatively low for the extracted collocations.

Based on the collocations listed in the tables for the three collocations in this section (§4.2.3), it is evident that despite the limited use and low LL scores for the Verb-Noun collocation when it is in the infinitive form *to*, the results reveal that the use of this structure by the Arab EFLLs is accurate in nearly all the cases. Arabic does not have an infinitive structure, which may create difficulties for the Arab EFLLs when using the infinitive *to* with Verb-Noun collocations. A sample of 100 examples of the use of (*to* + Verb-Noun collocations) in their contexts was collected from the AEFLL Corpus, through the concordance search output (§the concordance output for the sample is given in Figure A-6 in the appendix). The results suggest that there were some interlingual errors in using the infinitive *to* with the Verb-Noun collocations accounting for 16% of the sample which are classified into two categories: wrong verb tenses and missing intermediates when needed within the collocations.

The concordance lines show that Arab EFLLs frequently use the wrong verb tense in collocations with infinitives. The verb following the *to* infinitive should be in the base form (e.g. *see, go,* etc.). The students tend to use the verbs in the past tense with the infinitive *to* accounting for 6% of the sample. The following are examples of verb tense errors in the collocations:

- **I wish to visited again – to visit*
- **I had to asked Hanmda – to ask*
- **We stayed in the coffee shop to drank tea – to drink*
- **Many students want to got grade A – to get*
- **I would like to go again to ate their food – to eat*

There are some spelling issues in the use of the verb ‘take’ ‘*To tack pictures’ which should be as ‘To take pictures’. This spelling issue is observed in a couple of essays from different learners in the corpora.

The second type of interlingual error is the absence of intermediates when these are obligatory within the collocations. When reviewing missing intermediates, however, only 4% of the learners’ collocations in the sample were missing the intermediate needed to produce an accurate and complete sentence. The examples listed below show that the learners are not using articles when needed

in a collocation, and some other words are missing, such as the word ‘to’ in the third example below (i.e., ‘*to go school’). The examples below show learners’ use on the left with an asterisk, and the proper use on the right:

- *+To reduced sense – To a reduced sense deprived from your senses.*
- *To go scool – To go to school*
- *To be teacher of math – To be a teacher of math*

Having discussed the results for the Verb-(intermediate)-Noun collocations, *to* + Verb-Noun collocations and *to* + Verb-(intermediate)-Noun collocations, I now present the overall CEA results for the Verb-Noun collocations. The first type is interlingual errors which are seen in the use of the incorrect word form of the noun in Verb-Noun collocations such as ‘*in the classe every one keeps silence’. The correct form for saying this collocation is using the noun ‘silent’ or better used ‘quite’ as in ‘keep quite’. The collocation keeps silence is grammatically correct but a common expression would include the noun ‘quite’. Also, ‘remain silent’ is more formal than ‘keep quite’. Another kind of interlingual error is in the verb tense in the collocations:

- *My father *make accident.*
- *Sudenly, came a faster car and *make accident* with my father car.
- *When she *do accident*, she started cnray and the police man couldn't ask her any questions.
- *Some one forgive them if they make a mistakes.

Interlingual errors were also seen in the choice of lexical items within collocations, which can be influenced by the L1. The Arab EFLLs are probably speaking about the act of doing exercise in the following examples:

- *I play sports every morning*
- *I advise play the sport*
- *and go to the park can play sport*

Therefore, for them, the use of *play* with noun *sports* as a collocation is acceptable but the native-speaker raters classified this as "questionable". The contexts indicate that learners meant the routine activity of physical exercise. The BNC has examples of ‘play sport’ and also ‘play sports’ (with an 's'). It is also in

the Collins English Dictionary, though it does not refer to general exercise in all those cases. This could be a case of L1 interference because this collocation is acceptable in Arabic, where it is expressed as العَبُ الرِياضةً كُلُ صَباحٍ (*al*ʕa*bu al-rriyāḍata kul ṣabaāḥ*) 'I play sports every morning'. There is a clear L1 influence in the insertion of the article '*the'* when it is not needed as in '*play the sport'. The omission and insertion of certain elements within the collocations is an overlap error attributable to both interlingual and intralingual factors. This includes the omission of possible intermediates within the collocations such as articles and the use of *to.* The following are some examples:

- **To use alot,* to use agood, *acar, *adays*
- **To go school*
- **To be teacher of math.*
- **To reduced sense*
- **To use computer*
- **My father make accident.*
- **We should forgive them when they do mistake.*

The Verb-Noun collocation list for NBESs indicates that they use a wide range of vocabulary, which is clear from the variety of noun collocates for every single verb found in the corpus (§Table A-15 and Table A-16). The Arab EFLLs, on the other hand, have a limited vocabulary range and as a result, there are fewer instances for the noun collocates for every verb. For example, there are two collocations for the verb *give* in the AEFLL Corpus ('give birth' and 'give student'). In comparison, the NBESs Corpus has 23 noun collocates such as 'give rise', 'give birth', 'give formula', 'give information', 'give evidence', etc. It must be mentioned that there are some combinations that are not considered as collocations in the previous example such as 'give student' used by Arab EFLLs and 'give formula' by the NBES; those two collocations are part of a bigger phrase as they require a noun as in 'give the students time' and 'give formula milk'. Another example is the noun collocates for the verb '*make'* and '*reduce'* listed in Table 4-37 which shows the difference in vocabulary range between the Arab EFLLs and NBESs:

Table 4-37: Examples of Noun collocates in the Verb-Noun collocation in both corpora

Verb	Noun collocates in AEFLL Corpus	Noun collocates in NBESs Corpus
MAKE	People—sense—decision—*exrcise—life—revision	Sense—decision—use—mistake—profit—money—interference-process—comparison—judgment—generalisation—assumption—people—reservation—friend—progress—fun—purchase—concession—adjustment—*improvement
REDUCE	Smoking—greenhouse	Cost—sugar—risk—poverty—mortality—unemployment—inequality—expenditure—disturbance—nitrate—traffic—blood—health—congestion—pain—energy—plasma—dose—drag—stress—error—agent—defect—yield—uncertainty

The Arab EFLL collocation list contains six noun collocates for the verbs *MAKE* and two for the verb *'*reduce'. In contrast, the NBESs had a wide variety of noun collocates for the same verbs 'make' and 'reduce'. The numbers of nouns that collocate with the verbs 'make' and 'reduce' are 21 and 25 respectively. This discrepancy may be due to the range and extent of the NBES data. The verb 'eat' was the only verb that had eight noun collocates for the Arab EFLLs. The remaining verbs searched for in the AEFLL Corpus had fewer nouns collocating with them. In addition, there were some verbs in the AEFLL Corpus which have either one or two noun collocates, representing 17.2% and 13.1%, respectively. Consequently, this distinction also explains why the Verb-Noun collocations extracted for NBESs contain more nouns than those for Arab EFLLs.

There are some differences in the use of lexical items within collocations. The Arab EFLLs tend to use a limited choice of noun collocates. For example, the noun collocates for the verb 'give' in the AEFLL Corpus were 'give student', 'give human qualities', 'give a chance' and 'give birth'. The NBESs, on the other hand, use other noun collocates with the verb give such as 'give rise', 'give information', 'give an insight', 'give credit', 'give a clue', 'give priority'.

The results indicate that the majority of collocations are accurate as separate entities for both Arab EFLLs and NBESs, although there are exceptions on the

Arab EFLLs list, such as 'be gum' which has no meaning. The use of the collocation 'be gum' is peculiar in both Arabic and English. Figure 4-14 is a screenshot of the concordance lines for the collocation *BE* + *gum*.

left	match	right
people who do plastic surgery could	have infections if	they have poor surgery procedures
results. Mour serious results	are gum infection	and damage to bones and tissues
resultes. More serious results	are gum infection	and damage How you perfect holiday
results. More serious results	are gum infection	and damage to bones and tissues
results. More e serious results	are gun infection	and damage to bones and tissues

Figure 4-14: A screenshot is taken from IntelliText web interface for *'BE + gum'* in the AEFLL Corpus.

The concordance lines in Figure 4-14 show that the collocation verb (*be)* + noun *(gum)*. The verb (*be*) stands as the lemma form of the verb *be* (i.e., *be, is* and *are*), so the verb in the collocation output stands for all the derivatives of the verb *be,* creating meaningful utterances. It is important to know that IntelliText users can see the actual form when they click on the collocations to see how they are used in the concordance lines. Therefore, this is a display issue and should not be treated as a faulty collocation.

There are some Verb-Noun collocations in the Arab EFLL collocation list that are not considered as collocations, yet they form part of a bigger phrase such as 'make people' and 'make life'. Also, the second example in Table 4-37 *reduce greenhouse* is not a complete collocation in itself; the collocation is part of 'reduce gases' as in 'reduce greenhouse gases'. The collocation 'add music' is another example that has been judged as unacceptable by the native-speaker raters in the AEFLL Corpus. However, the concordance search for this collocation indicates that it is followed by a noun modifier as in 'add music major[3]', suggesting that it is part of a bigger phrase. The number of combinations that form part of a bigger phrase in the extracted Verb-Noun collocation list for the Arab EFLLs is 13, accounting for 8.9% of the total number of collocations such as 'give student', 'add music', 'return police', 'cause health', 'read Holy', 'eat junk', 'use cell', 'cause air', 'cause heart', and 'play foot'. In contrast, for the NBESs

[3] From the context, it is clear the students are discussing Major and Minor course options.

there are 9 instances that form part of a bigger phrase, such as ‘give people’ and ‘look tonight’, accounting for only 2%, which is very low compared to the Arab EFLLs.

The choice of collocates is another issue within the Arab EFLLs, for example the verb *do* is more appropriate than the verb ‘make’ in the collocation ‘make exercise’. In the BNC, there is no use of ‘make exercise’ yet there are 36 instances of the verb *do* with the noun *exercise*. Also, this is evident in the BAWE corpus which contains two instances of *make exercise* but nine other instances of *do exercise*. The collocation is utilised differently in the two occurrences of ‘make exercise’ in the NBESs Corpus than it is in the corpus of Arab EFLLs; the concordance output below explains the difference:

- *The early steps, screening and scoping, are the most important, as they involve the most complex decision making exercises.*
- *Finally, an example of a decision making exercise was analysed,*

The two examples suggest that these two NBES instances were used to refer to a type of task for making decisions. However, the Arab EFLLs use the collocation to refer to physical activity for exercise. Some examples of the collocation search in the concordance are the following:

- **However, most of them don’t do exercise.*
- **But know adays only some people do exercises.*

The collocation ‘make revision’ is another example that may appear acceptable to Arab EFLLs when employed in close proximity but did not sound acceptable to the native English speaker judges. The collocation ‘make revision’ is not used by native speakers, as neither the BAWE corpus nor the BNC contain any examples. Even the plural form of the collocation (i.e., ‘make revisions’) does not appear in either corpus, though the collocation with the plural is acceptable in English. The following lines demonstrate how two learners in the AEFLL Corpus employed the collocation:

- **She didn’t study day by day and make revision.*

- **Study day by day and make revision ask any question that you didn't know about it.*

The collocation *make revision* is used as a complete collocation by the Arab EFLLs without having a word that follows to create a meaningful collocation. However, in the second reference corpus (I-EN), there were two instances of this collocation and six instances when used in the plural form. The collocations are usually followed by noun modifiers or prepositions (*of* or *to*) and sometimes preceded by *to*. The following are examples of the collocation as it occurs in the I-EN:

- *We will make revisions of the current method…*
- *I have avoided the temptation to make revisions.*

Another example that was rated as unacceptable by the native-speaker judges is 'do plastic surgery'. The following are two out of the 22 instances of 'plastic surgery' in the AEFLL Corpus:

- *Nowadays, everyone wants to do plastic surgery in order to hide even the smallest physical deficiencies.*
- *They might need to do plastic surgery.*
- *Therefore, you need to do plastic surgery.*

However, the NBESs Corpus showed only two instances of the collocation 'plastic surgery' including the following:

- *Indeed, the media fascination with cases of plastic surgery going wrong or simply too far.*
- *She is being taken to children 's hospital, then a plan of action will be formulated by the plastic surgery team.*

The collocation 'plastic surgery' collocates frequently with the verb *do* in most of the cases in the AEFLL Corpus amounting to 90% of the total occurrences. However, based on a sample of 100 lines of *plastic surgery* in the BNC, the most frequently occurring verbs are given in Table 4-38.

Table 4-38: Verbs collocating with 'plastic surgery' in the BNC

Verb	Percentage	BNC Examples
have	12.8%	I had to *have a plastic surgery*.
need	7.14%	In view of the fact she was a young girl and the lacerations were right across the cheek I decided she *needed plastic surgery.*
undergo	5.7%	I will tell the tale of a branch who loaned a customer a sum of money so he could *undergo plastic surgery*.
get	1.4%	You 'll be *getting plastic surgery.*
require	1.4%	Tordo may *require plastic surgery.*

These errors are attributable to the learners' L1 interference in the target language. L1 influence is seen in inappropriate use of the nouns 'mistake' and 'accident' to collocate with the verbs 'make' and 'do'. The Arab EFLLs tend to use the collocations '*make/did accident'* and 'make /did a mistake'. The Arabic language has many alternatives for those two verbs, such as the following:

- I had an accident — حَصَلَ لِي حَادِثٌ (*ḥaṣal lī ḥādiṯ*)
- I did an accident — تَعْرضتُ لِحَادِثٍ (*tʕaraḍtu liḥādiṯ*)
- I made an accident[4] — عَمِلتُ حَادِثٍ (*ʕmylt ḥādiṯ*)

Based on the reference corpora, the verb most frequently used with the noun 'mistake' is 'make', occurring 93 times in the BNC and 133 times in the I-EN corpus. The noun 'accident*'* collocating with the verb 'to have' occurs 3 times in the NBESs Corpus, 104 times in the BNC corpus, and 35 times in I-EN corpus.

The Arab EFLLs use the verb 'read' with the noun 'programme' which indicates an influence of the L1 on the choice of the right hand collocate. There is no occurrence of this collocation in the NBESs Corpus nor the BNC or I-EN corpora. This collocation occurs in Arabic which could indicate that the learners are depending on their L1 when using this collocation. There was no occurrence of

[4] All the given examples in Arabic are taken from the ArTenTen12 (Arabic Web 2012 sample 115m). This can be accessed via Sketch Engine: https://www.sketchengine.eu/artenten-arabic-corpus/). ArTenTen12 is a reference corpus for Arabic language comprised of texts gathered from the Internet.

this collocation in the ARTENTEN12 corpus either. However, as a speaker of Arabic, I can affirm that the collocation ‘read programmes’ exists in Arabic. The following are examples taken from Google searches and Twitter that prove that this collocation is used in Arabic:

- *هل قرات البرنامج الحكومي في عهد حكومه (hl qrāt al-barnāmj al-ḥakūmī fī ʿahd ḥakūma)*
  ‘Did you read the government program during the government's era?’
- *قريت برنامج (qarīt barnāmj) ‘ I read the programme.’*

Some collocations may not be lexically acceptable to the native judges but can be understood by a native English speaker. These collocations include but are not limited to ‘love marriage’, ‘read Quran’, ‘ask Allah’ in which the last two collocations relate to the religious background of the learners. The collocation ‘love marriage’ is a Noun-Noun collocation, indicating a marriage that is based on mutual love rather than a marriage arranged by families. It is not commonly used in English but the meaning is likely to be understood by many native speakers. Though these collocations do not occur in the BNC as the reference corpus, the collocations can still be used in English. Therefore, this indicates that the acceptability of collocations may vary depending on the rater’s cultural background or breadth of knowledge.

It should be noted that the main verbs that have a lot of noun collocates within both datasets can be classified as delexicalized verbs. Delexicalized verbs are verbs that are attached to a noun phrase to perform an action that is indicated by the noun. The meaning in this structure relies on the noun itself; the delexical structure modifies the meaning of the verb. For instance, ‘have breakfast’ contains the verb *have*, which signifies possession, but the meaning is altered when employed in a delexical form. Some examples of delexicalized verbs are ‘have’, ‘make’, ‘do’, ‘take’, ‘go’, etc. Both corpora show use of delexical structures in Verb-Noun collocations with 20 and 75 collocations with delexical structures used by the Arab EFLLs and NBESs respectively.

Table 4-39 Examples of delexical verbs used in the Verb-Noun collocation[5].

Delexical verbs	AEFLL Corpus	NBESs Corpus
Have	Have fun	Have access - have difficulty-have implication - have child-have trouble - have formula-have fun- have resource
Make	Make exercise- Make sense make decision-Make revision	Make sense- make decision- make use- make mistake- make profit-make reference - make money- make interference
Do	Do something- do anything	Do something - do anything- do mathematics
Go	Go shopping	Go bust- go shopping- go clang- *go strawg- go ping (§Table A-14 for more examples)
Give	Give birth	Give rise- give birth- give formula- give information- give advice- give evidence – give insight- give credit

Table 4-39 lists the delexical verbs in both corpora. The first column shows the use within the AEFLL Corpus and the second the NBESs Corpus. The Arab EFLLs do not use delexical verbs to the same extent as the NBESs which is evident in the number of collocation examples in Table 4-39 for both learners. This highlights a significant difference in the use of the delexical verbs between the two. Additionally, Arab EFLLs find the delexical structure difficult as sometimes they do not know how certain words can collocate (Shammas, 2013; Suleiman, 2022).

In this section, I have discussed the CEA findings for the Verb-Noun collocations. The discussion of these findings will be given in Chapter 5. In the following section, I move on to provide the results for the Verb-Adverb collocations.

4.2.3.2.3 CEA Results for the Verb-Adverb Collocations

In this section, I present the CEA results for the Verb-Adverb collocations. I will first highlight the results based on the native speakers' judgements of the data. Then, I will present the main CEA findings for the use of the Verb-Adverb collocation tabulated in (§Table A-17 and Table A-18 in the appendices).

[5] (§ Table A-15 in the appendix for more examples)

There were few errors in the use of Verb-Adverb collocations by Arab EFLLs. Thus, only a small number of collocations were provided for the native-speaker raters to assess their acceptability in the sample. Based on the native speakers' judgement, there was a total of five errors, four of which were classified as interlingual errors and one of which was attributed to an intralingual cause. Some of the Verb-Adverb collocations that were rated as unacceptable and questionable were: '*see recently', '*mention previously', '*do necessarily', '*produce environmentally' and '*drive speedly'. The three collocations '**see recently*, **do necessarily* and **mention previously'* were used accurately in the concordance lines. These three collocations were used as 'seen/saw recently', '*does not necessarily' and 'mentioned previously'. Therefore, these three were rated as unacceptable because of an author typo error when the survey was conducted and should not be regarded as errors. The only error is in the placement of the negative 'not' with the verb 'do' as the Arab EFLLs tend to split off the '*n't'* in the abbreviated form which should be attached to the main verb[6]. This kind of error can be classified as an intralingual error. The following are two examples explaining the error:

- **Caffeine impairs insulin action, but <u>does n't necessarily</u> affect blood sugar*
- **Being healthy and fit <u>does n't necessarily</u> mean to be stick thin and starve yourself*.

The fourth unacceptable collocation is 'produce environmentally'. The following example shows the collocation in context as part of a bigger phrase.

- *Offer them tax policies to manufacturer that <u>produce environmentally friendly bags</u>.*

Another Verb-Adverb collocation that was unacceptable to the raters was '*drive speedly'. This collocation is incorrect, and this was also evident in the absence of this collocation in either the NBESs Corpus or the BNC as reference corpus. The following two points are two examples of the collocation taken from the AEFLL Corpus:

- **They <u>drive speedly</u> weather the roads are impety.*
- **In addition, they <u>drive less speedly</u> than male but not slow.*

---

6 The splitting of the negative *not* with the verb *do* could be a cause of result of a corpus tagging error. However, as a teacher this was a common error for my students.

The Arab EFLLs used 'speedly*'* as the adverb expressing *fastness.* This unacceptable use could be attributed to their L1, and thus be classed as an interlingual factor. The correct expression would be *DRIVE fast/ speedily.* The Arab EFLLs tend to have a word order issue in these collocations. The collocation *driving fast* is also used by the Arab EFLLs. A grammatical error in this collocation was observed in the incorrect spelling of the adverb 'speedily'. The error in 'drive speedly' could be interlingual, caused by the differentiation from Arabic to English, but could also be intralingual, as the learners add '-ly' incorrectly to the noun or verb form 'speed' to express extreme speed. The Arab EFLLs use of *speedily* instead of the basic form *fast* could be explained by the dependency on -*ly* to express the intended meaning. This is a common error made by Arab EFLLs because there is no direct equivalent to '-*ly'* in Arabic, which could mean that Arab EFLLs tend to use them more frequently than necessary. This is in line with previous studies such as that of Al-Shormani and Al-Sohbani (2012, p.136). For example, Arab EFLLs may say 'you look sadly today' or 'they have deeply feeling in their hearts*'*. The use of the adverbial suffix '-*ly'* when not necessary can be considered a developmental error caused by overgeneralisation of the rules of the L2, the language being learnt.

The third collocation was '*break fast'. The collocation is not correct as the students were talking about the first meal of the day, which is breakfast. The Arab EFLLs tend to split compound words into two. After examining the NBESs Corpus and the two reference corpora, the word *breakfast* as one word is used far more than splitting the compound word into two. The following is a table that shows the occurrences of both in these three corpora:

Table 4-40: The total frequencies of the use of *break fast vs. breakfast in the English native speaker corpora

Corpus	Break fast		Breakfast	
	F	NF	F	NF
NBESs Corpus	0	0	35	4.537
BNC	7	0.063	3,176	33.808
I-EN corpus	2	0.11	4,084	21.480

When the two words are used separately, to **break fast,* they are now mainly used to mean the breaking of a religious fast which explains why Arab EFLLs have this confusion about splitting the two words. The following are examples taken from the AEFLL Corpus:

- **After that, I went to get some break fast*
- **The next day we got up early and ate the break fast.*

Another aspect of writing '*break fast' in two words also relates to Arabic as the language does not have compound words. So, using compounds in translation from English to Arabic is difficult for Arab EFLLs because of this linguistic difference between the Arabic and English languages (Mheel et al., 2018). Therefore, this error can be explained as caused by an absent category in the Arab EFLLs' L1. There is another contextual interlingual error where the learners use the article 'the' with 'breakfast' as in the second example above.

Other interlingual errors within the Verb-Adverb collocation can be seen in the tenses of verbs. For example, the verbs 'spend' and 'see' are used in the present to express an event in the past, such as:

- **In the last summer holiday I spend a very perfect holiday*
- **In this place me and my family spend very nice time*
- **The best film I have see recently, called "Uars 9 ours".*

There is also a noteworthy point to be made regarding possible intermediate words within the Verb-Adverb collocation, such as an adverb following the verb. The Arab EFLLs tend to rely on the adverb 'very' which is caused by L1 influence. The Arab EFLLs use of 'very' within these collocations is correct. The following are two examples taken from the AEFLL Corpus and the NBESs Corpus:

- *Because he failed to drive more carefully (NBESs)*
- *Some women who drive very carefully (Arab EFLLs)*

For the Arab EFLLs, the possible intermediate word was the adverb ‘very’. However, in comparison, the NBESs used the word ‘more’ rather than ‘very’ in most of the cases to express the same meaning as the Arab EFLLs. Although there is a semantic differences between the two phrases in the examples given, there is a much higher use of the adverb ‘very’ in the AEFLL Corpus than in the NBES which emerges clearly in the number of total occurrences within the corpus (§Table 4-41). The number of uses of the adverb ‘very’ was 3,175.27 ipmw in the AEFLL Corpus and 764.397 ipmw in the NBESs Corpus. This supports the argument proposed in Alshammari (2021) regarding the Arab EFLLs’ overuse of the word *very* to express emphasis.

Table 4-41: Normalised frequencies of the adverb ‘very’

Adverb	AEFLL Corpus NF	NBESs Corpus NF
Very	3,175.27	764.397

The overall results indicate that there were far fewer errors in the Arab EFLLs’ use of Verb-Adverb collocations than in the other collocation types. This result was not expected but it supports the result of the statistical significance testing of this collocation which accepted the null hypothesis that there were no statistical significance differences between Arab EFLLs and NBESs in the use of the Verb-Adverb collocation.

4.2.3.2.4 CEA Results for the Adjective-Noun Collocations

In this final section of the CEA, I will illustrate the results for the Adjective-Noun collocation used by Arab EFLLs. The extracted Adjective-Noun collocations are listed in Table A-21 and A-22 in the Appendix. The total number of the Adjective-Noun collocations extracted from the AEFLL Corpus was 666, the highest number among the four types of collocations investigated in this study.

The first step in the CEA was to identify possible errors within the Adjective-Noun collocations. Based on Table 4-33, the Adjective-Noun collocations had few errors, amounting to 3.05% for interlingual errors and 6.9% for overlap errors.

Interlingual errors were most prevalent in the choice of lexical collocate, due to the influence of the L1. Other adjectives in these collocations might be more appropriate than those used. The first example is the collocation 'huge headache*'*. There is a lack of the article 'a' in the first example which is caused by interlingual error. The AEFLL Corpus indicates that the learners employ the collocation to allude to its literal meaning. However, the native speakers typically use this collocation figuratively. The following two example are from the AEFLL Corpus:

- **I woke up one day with huge headache and backache which stopped me from moving my body as the pain was to an extreme point.*
- **In the other hand, over sitting on computer would cause a huge headache.*

There was no occurrence of this collocation in the NBESs Corpus, and only one occurrence of this collocation in the BNC:

- *The French referendum on closer European ties left a huge headache for Premier John Major, currently EC president and still reeling from last week's pound crisis.*

The learners intended to explain how severe their *headaches* were by using the wrong collocate, the adjective ***huge***. An example of the collocation in an Arabic corpus was the use of the adjective 'severe/strong' with the noun 'headache' صُداعٍ شَدِيدٍ (*ṣdāʕ šdīd)* which occurred 1,431 times on the Arabic Web 2022 (arTenTen12). This example could be an error attributed to an interlingual factor where the learners mistakenly used the incorrect collocate to express the intended meaning. Following the example found in the BNC, I searched the web, mainly Google Scholar, to prove my argument that 'huge headache' is used figuratively in English. The phrase is used to express how a thing might cause difficulties to someone, and not to express the extent or severity of a literal

headache. Google Scholar showed 895 instances in which it is used figuratively, and the following are some examples:

- *As expected, the act has turned out to be a huge headache to enforce and administer*
- *I may not have symptoms, but COVID-19 is a huge headache.*
- *the handling of these assets alone is already a huge headache for management.*

Another collocation that was deemed unacceptable by the native-speaker judges was the collocation 'easy victims' as in 'They are easy victims of cigarettes'. Though, the collocation 'easy victim' is correct in itself, the error here is mainly in its use in this context. This collocation did not occur in the NBESs Corpus. However, there are 3 instances of this collocation in the I-EN corpus and 5 instances in BNC including:

- *A pillar of grey smoke rose far off to the west from some other hill, its spread lazy, an easy victim to the wind indicating it belonged to a fire already dead. (I-EN)*
- *they saw the land of ethopia as an easy victim for there own benefactor.(I-EN)*
- *he was as simple-hearted and honest as the day was long, and so she was an easy victim.(I-EN)*
- *Both of them were easy victims: where she was slow and tongue-tied. (BNC)*
- *They come unprepared, easy victims tounscrupulous employers. (BNC)*

Another interlingual error is the choice of the noun collocate 'shops' in 'fast food shops'. The noun 'restaurant' is more formally appropriate than 'shop' but a possible explanation is the L1 interference in which the word 'shop' is used in Arabic as in مَحِلُ الوَجَبَاتِ السَرِيعةِ (*mḥl al-wjbāt al-ssarīʕti)*. If the emphasis of meaning is on taking away food rather than sitting down inside to eat, then the collocation 'fast food shops would be acceptable. However, the following examples show that the use of 'restaurant*'* would be better since they mean the place to eat food:

- **why fast food shops such as Burger King.*
- **of taking them to any fast food shop, take them to any healthy food shop so that they can enjoy the food …*

Additionally, the results of an IntelliText frequency comparison between 'fast food shop(s)*'* and 'fast food restaurant(s)*'* reveal that there was no occurrence of the first collocation in the BNC, NBES, or the I-EN corpora, and raw scores of 12, 108 and 1, respectively for 'fast food restaurant(s)*'*. Therefore, this would suggest that the noun 'restaurant*'* is the commonly used form. In the BNC*, fast food* frequently occurs with 'restaurants and 'market'. A 100-line sample revealed that *fast food* collocating with the noun '*restaurant'* amounted to 7%, and with the noun *market* 10%. The following are some examples:

- *The report forecasts that the UK fast food market is set to continue growing.*
- *Some fast food restaurants will be obliged to ..*

The difference between the use of the two nouns must be explained. If the noun *restaurant* is used, then it refers to a venue where customers go and eat prepared meals. However, when the noun *market* is used, then it is used to refer to the fast food sector within the economy.

Overlap errors in the collocations include errors attributable to both interlingual and intralingual factors. The first example is the collocation 'Grand parents', which should be written as one word, but which Arab EFLLs tend to split into two words. The error in this collocation could be attributed to interlingual factors as Arabic does not have compound words.

There are some collocations that were judged by the raters to be unacceptable, but where the contexts of these collocations indicated that they were accurate. Therefore, they should not be regarded as collocational errors. These collocations include the following:

- *Smart driving* – 'Reduce your Carbon Footprint with smart driving'
- *Smart detriment* – as in 'smart device detriment'
- *Enough restaurants* – 'there are not enough restaurants.'
- *Live life* – 'You live life the way you are told to' and 'It is about living life for you'.

The collocation *smart detriment* was given to the raters without the intermediate noun *device*, making them understandably view it as unacceptable. It was not expected that the native raters would view the collocations *smart driving* and *enough restaurants* as unacceptable because they do occur and are used in English. The raters explained at the end that further context would be required to judge the acceptability of the collocations.

In this section, I have illustrated the Adjective-Noun collocation errors with some examples analysed qualitatively through a concordance search of the AEFLL Corpus. I have also explained and classified those errors. Some errors in omitting certain elements within a collocation were attributed to L1 interference (i.e., interlingual errors). Some errors can be attributed to more general learning or developmental processes (e.g. overgeneralisation of English rules, incomplete application of language rules or ignorance of language rules). Overall, the Adjective-Noun collocation displayed the least number of collocational errors.

4.3 Conclusion

In this chapter, I first presented a comparison between BE and AE in the use of collocations, which functioned as an initial trial of the methodology. Then, I presented the results of the comparative corpus-based study of the use of four lexical collocation types by Arab EFLLs and NBESs in two sections. The first section was a corpus-based statistical analysis of the frequencies of collocation use of the two groups. The complexity level of the word choices made within the collocations was also analysed through the CEFR categorisation of words into levels of proficiency, and as part of that discussion, I briefly introduced a proposal for the CEFR vocabulary profiling to be done on the basis of collocations instead of single words—a collocation-informed vocabulary profiling system which I call the CEFR-C. The second section presented a contrastive error analysis of the Arab EFLLs' use of collocations, which offered a more nuanced view of the learners' collocations through a close reading of the textual contexts in which the collocations appeared. In the next and last chapter, Chapter 5, these findings and their implications will be thoroughly discussed.

Chapter 5 Discussion

Introduction

In Chapter 4, I presented and analysed the results of this study that examined the use of four sets of lexical collocations, the *Adverb-Adjective*, *Verb-Noun*, *Verb-Adverb*, and *Adjective-Noun* collocations, by Arab EFLLs and NBESs. The analysis was conducted using a corpus-based frequency approach and a contrastive error-analysis approach.

The purpose of this chapter is to provide a comprehensive overview of the findings and discuss their relevance to the research questions that guided this study. Specifically, this chapter aims to answer whether this particular corpus-based frequency analysis of the use of lexical collocations has revealed any statistically significant differences between Arab EFLLs and NBESs, and to investigate the factors that may contribute to these differences through the implementation of CEA as detailed in the last chapter.

5.1 Answers to the research questions

In the upcoming sections, I will present a detailed discussion of the research questions and the main findings of this study and will discuss how these findings contribute to and extend the existing knowledge base in the field.

5.1.1 First and Second Research Questions (RQ 1)

The research questions (RQs) were addressed previously in the first chapter (§1.6) and as a reminder, the following is the first research question:

> ***RQ1:*** *Can hypothesis testing of relevant corpus data help us to identify the significant differences between Arab EFLLs and NBESs in their use of the four lexical collocation types chosen for this study?*

The aim of the first research question was to discuss the differences between the Arab EFLLs and NBESs in their use of four chosen lexical collocations. The intention of the first research question was to discuss the results of the corpus-based frequency findings through the use of sophisticated statistical hypothesis testing.

Therefore, to discuss the first part of the first research question, the following hypothesis was formulated:

The null hypothesis ($H1_0$): There are no statistically significant differences in frequency between the lexical collocations used of by Arab EFLLs and NBES.

Alternative to the null hypothesis ($H1_A$): There is a statistically significant difference between the lexical collocation use of Arab EFLLs and NBESs.

In the results chapter, the four collocations were analysed, and the corpus-based frequency results indicated that the Arab EFLLs used fewer collocations than NBESs (§4.2.1). This was evident from the numbers of extracted collocations in both corpora. Adverb-Adjective collocations were found to be the least used by Arab EFLLs, with only twenty collocations found in the corpus available for investigation. One possible explanation for this is the lack of Adverb-Adjective collocations in Arabic, as noted by previous scholars (Husamaddin, 1985; Ghazala, 1993; Mustafa, 2010; Abd Ai-Qadir, 2015). Abdul Ridha and Al-Riyahi (2011) had found that Arab EFLLs tend to avoid using Adverb-Adjective collocations, opting for longer sentences instead, although they did not provide any examples to illustrate this preference. This finding is consistent with a study of Korean EFLLs, who reportedly also used few Adverb-Adjective collocations (Lee, 2006).

The findings of my study are consistent with previous research on Arab EFLLs in general (Laufer and Waldman, 2011; Men, 2015; Farooqui, 2016) which found that Arab EFLLs use fewer collocations. Interestingly, Adjective-Noun collocations were found to be commonly used by both Arab EFLLs and NBESs, with 665 Adjective-Noun collocations extracted from the AEFLL Corpus. This supports the findings of Siyanova and Schmitt (2008), who found that Russian EFLLs tend to to use most of the Adjective-Noun collocations appropriately. However, this contradicts the findings of El-Dakhs (2015) who found that Arab EFLLs were less confident when using Adjective-Noun collocations compared to Verb-Noun collocations. It is important to note that El-Dakhs' study investigated the competence of collocation use among Arab EFLLs, while the current study focuses on their frequency of use. Nevertheless, the results suggest that Arab EFLLs tend to use Adjective-Noun collocations frequently, which is in line with other studies that have found that Arab EFLLs tend to frequently use this type of collocation regardless of their confidence in their accuracy (El-Dakhs, 2015; Dukali, 2018).

In this study, both Arab EFLLs and NBESs showed accurate use of Verb-Noun collocations in the infinitive form. Although it was expected that NBESs would have more examples and higher LL scores, the analysis revealed that both data sets showed limited use of this form.

Regarding the third set, the frequency analysis of the Verb-Adverb collocation showed that only a few verbs collocated with adverbs, with only 53 instances for Arab EFLLs and 151 for NBES. This suggests that Arab EFLLs struggle with using this type of collocation, their examples amounting to only 35% of the total NBES use. The difference between the two corpora can be attributed to the differences in topics and (sub)-genres in the two corpora. The Arab EFLLs corpus encompasses everyday subjects and is also smaller in terms of the word count compared to the NBES corpus. The results of hypothesis testing, however, showed no significant difference between the Arab EFFLs and the NBESs in their use of the Verb-Adverb collocation, which was in contrast to the other three collocation types studied.

The frequencies of most of the collocations were found to be significantly different between Arab EFLLs and NBESs, leading to accepting the null hypothesis. However, Verb-Adverb collocations and infinitive *to* + Verb-Noun collocations were interestingly found to be not significantly different. The results for the infinitive *to* + Verb-Noun collocations were not expected to be the same for both Arab EFLLs and NBESs; the LL scores were relatively low and there were not many occurrences of this combination within the two learners' corpora. Hence, the results at the moment may be said to be inconclusive. Generally, the Arab EFLLs seem to be limited in their use of the Verb-Noun and *to*+ Verb-Noun lexical collocations. This suggests that Arab EFLLs need to be familiarised with how to employ words in a sentence and how words often co-oocur and combine with others. From my experience as a student and then as a lecturer of writing courses for university-level students, the teaching of vocabulary in Arab countries is mainly focused on translation and not on how words should be used in real sentence contexts.

Based on the statistical significance results summarized in §Table 4-27, it is worth noting that the linguistic data in this study appear to favour non-parametric tests. While parametric tests are commonly used and considered to be the most powerful in social science research (Saldanha, 2009), they were not applicable to the linguistic data used in this thesis. This finding is consistent with the views of other scholars who argue for the use of non-parametric tests for linguistic data, such as Deane (2005), who notes that the great particularity of the relations between linguistic units leads to highly skewed data.

However, there are some scholars, such as Brezina (2018a), who claim that linguistic results obtained through parametric tests are still robust even if the normality assumption is not met. In my opinion, if the researcher can confirm the assumption of normality for the linguistic data, then using parametric tests would yield more robust results than simply applying non-parametric tests.

In summary, the corpus-based comparison of frequencies study revealed that there are generally significant differences in the use of lexical collocations between Arab EFLLs and NBESs. It was challenging to find literature supporting the results due to differences in methodology. Previous scholars did not investigate collocations using the statistical approach used in this study. Thus, these findings are original and could serve as the basis for further research to either confirm or oppose the results. To gain a deeper understanding of these differences and their underlying causes, the study further investigated the Arab EFLLs' use of lexical collocations through CEA. This will be discussed in the next section.

5.1.2 Second Research Questions (RQ 2)

The following is the second research questions, that was addressed previously in the first chapter, to determine the differences in word choice within the four collocations in terms of vocabulary difficulty (based on the current CEFR classification of words):

> ***RQ2:*** *Using the CEFR classification of words into typical user proficiency levels, how can analysing word choices within language learners' collocations assist in understanding their language proficiency?*

As for RQ2, there were several outcomes of the exploration using CEFR vocabulary levels for profiling the complexity of word choices within collocations. The results of the study revealed that both Arab EFLLs and NBESs (as represented by BAWE, at least) exhibit fairly similar levels of complexity in word choices for the Adverb-Adjective collocation (i.e., mainly basic English uses at A1 to A2 CEFR levels), but some Arab EFLLs displayed higher-level vocabulary in the Verb-Noun collocation (i.e., reaching the CEFR proficient levels of C1-C2). However, their use of vocabulary in the Verb-Adverb collocation remained at basic CEFR levels, while NBESs display a greater range of lexical choices, including vocabulary at the C2 level. The study also indicated that the Arab EFLLs

showed a greater use of intermediate to advanced noun collocates in the Adjective-Noun collocation, while the NBESs used nouns at various CEFR levels.

Overall, the analysis of the data revealed that the vocabulary choices of Arab EFLLs are from the advanced level for some collocations, although their overall CEFR classification remained at the basic-intermediate level. Therefore, it cannot be said that their use of words can be classified under a single CEFR category. However, a limitation of this finding is that the advanced level was only reached in the choice of collocates, and not in the choice of the node words of the collocations. This suggests that while the Arab EFLLs may display some proficiency in using collocations, their overall language proficiency level may still be limited.

5.1.3 Third Research Question (RQ 3)

In the preceding section (§5.1.1 and 5.1.2), the first two research questions for this PhD thesis was addressed. This section will provide an answer to the third research question, taking into account the findings presented in the results chapter4.2.3 (§4.2.3). As a reminder, RQ 3 was formulated as follows:

> ***RQ3:*** *What parts of the lexical collocation errors made by Arab EFLLs can be explained by the influence of Arabic as an L1 (interlingual errors)?*

This question aimed to explore the impact of Arabic, as a first language, on the use of four specific lexical collocations in English by Arab learners of English. Specifically, this inquiry sought to understand the linguistic differences between English and Arabic that may lead Arab EFL learners to use faulty expressions. Two main areas of investigation were identified: (i) the influence of Arabic as an L1 with a focus on the use of the four lexical collocations in English, and (ii) the potential causes in both languages of lexical collocation errors in AEFLL writing. Overall, this research question aims to provide insight into how the influence of a learner's L1 can impact their acquisition and use of lexical collocations in English,

with a particular emphasis on AEFLLs.

The first notable difference is in the numbers of collocations extracted. In these the Arab EFLLs' are always fewer than the NBESs' for the four collocation types, which is in line with previous studies that compared various native speakers with EFLs, such as Howarth (1998) and Granger (1998b). The collocation list for NBESs displays a diverse range of vocabulary usage, whereas the Arab EFLLs collocation list features a more generic vocabulary usage. Generic vocabulary can be defined as general words that are commonly used in a language but may not be appropriate for use in all contexts (Durrant, 2013), and this could be influenced by factors such as language exposure and proficiency (Ariffin and Abdi, 2020). This result does not imply that Arab EFLLs are restricted in their lexical choices, but rather that they may have less familiarity with certain words than native speakers (Mahmoud, 2005; Alzi'abi, 2017; Al-Jarf, 2022). Also, this must not be taken as the whole truth because it is important to recognise that there are genre differences between the two corpora. While the NBESs Corpus includes only academic genres, the AEFLL Corpus, although academic in origin, also contains various everyday topics. For example, the second highest type of essays in the Arab EFLLs Corpus was narrative essays (§Table 3-1). Therefore, the use of generic vocabulary would be expected more in the AEFLL Corpus.

I anticipated a significant number of errors in the data collected from Arab EFLLs; however, surprisingly, the collocations as separate entities were mostly accurate. However, errors were identified in how these collocations were used in context, the choice of the lexical items within the collocations, a finding which is in line with Almahameed and Al-Shaikhli (2017) and Youcef (2022). Given that the participants were final year high-school students and undergraduates, I had expected to find more errors, especially considering that previous studies investigating collocations have mainly used postgraduates (Shammas, 2013; Farooqui, 2016), and EFLLs commonly use proof-readers before submitting their assignments (Harwood et al., 2009; Al-Inbari and Al-Wasy, 2022).

The errors in collocations identified in this study were classified into three categories: interlingual, intralingual, and overlap errors (§2.2.1). Interlingual errors were observed in the placement of collocations within sentences, and in the verb tense issues identified in the Verb-Noun and Verb-Adverb collocations. Intralingual errors were found in fragmented sentences, subject-verb agreement, and missing obligatory intermediates or modifiers. Overlap errors occurred when the context indicated both types of errors, making it challenging to determine whether the cause of the errors was due to L1 interference or L2 rules such as articles. Arabic does not have articles which mean interlingual issue; but the learning of articles can be developmental/learning issues which mean intralingual error. Hence, article usage can be classified as an overlap error.

In analysing the errors related to collocations, it was observed that the majority of them are contextual, occurring within the surrounding text. This is in line with (Al-Shormani and Al-Sohbani, 2012). These errors include wrong verb tenses, reliance on certain vocabulary, incorrect word order, subject-verb-agreement and fragmentary sentences, and missing articles. The primary reason for these errors is often attributed to the interference of the learners' first language (L1) or the rules of the second language (L2) or both. In particular, learners may struggle to identify the appropriate placement of collocations within sentences or struggle with verb tense issues, as observed in the interlingual errors. Intralingual errors, on the other hand, involve issues such as fragmented sentences or subject-verb agreement errors. Overlap errors are particularly challenging because they involve both types of error. Overall, understanding the various types of errors related to collocations and their causes is critical to improving language learners' proficiency in using collocations correctly.

I will now discuss the issue of verb tense and how Arab EFL learners use it in the formation of collocations. In the results for the Verb-Noun collocations, the initial analysis of the Verb-Noun collocation lists (§Table A-15 and Table A-16) showed that many of the collocations were meaningful on their own, decontextualised. The grammatical patterns that were investigated, including the use of all verbs and the infinitive form, revealed that syntactically, most of the collocations followed the basic Verb-Object (VO) syntactic pattern. Overall, the Verb-Noun

collocation comparison suggested that Arab EFLLs were limited in their collocation choices when compared to NBESs. Also, it is interesting to find that Arab EFLLs accurately used most of the infinitive Verb-Noun collocations with the exception of incorrectly producing many verbs in the past tense following the infinitive *to*.

Previous studies have shown that incorrect verb tense is the most common error among Arab EFLLs (Sawalmeh, 2013; Almahameed and Al-Shaikhli, 2017; Nuruzzaman et al., 2018; Arabet, 2020). Wrong verb tense use was found by other scholars who looked at Arab EFLLs' use of Verb-Noun collocations, such as Dukali (2018) and Alangari (2019). This finding is not limited to Arab EFLLs as it has also been found in Malaysian EFLLs (Hong et al., 2011). Sawalmeh (2013) identified verb tense errors, such as the use of the present tense to describe a past event (e.g. 'three children die in the accident last year') in the Malaysian data, while Nuruzzaman et al. (2018) found verb tense errors, such as the use of the present tense to describe a past action (e.g. 'Last week, I visit Riyadh'). Similarly, Arabet (2020) found that Arab EFLLs tend to make tense errors in Verb-Noun collocations, particularly with the infinitive *to*. Arab EFLLs often use the past tense for the verb following the infinitive (e.g. '*I wish to vistied again'/ '*We stayed in the coffee shop to drank tea'), which has generally been attributed to the lack of this structure in Arabic (Schulz, 2004; Kniaź and Zawrotna, 2021).

Additionally, some spelling mistakes were observed, mainly in nouns, while nearly all verbs were spelled accurately. It is worth noting that this inaccurate use of the infinitive form is not limited to Arab EFLLs (Gizatova et al., 2020), due to differences in the use of the infinitive in other languages. In many languages, a single word is used to represent the infinitive that incorporates a characteristic inflectional ending such as 'manger' in French for 'to eat' and 'portare' in Italian for 'to carry'. Other languages that use a characteristic inflectional ending for the infinitive include Latin and German. This supports the argument that the use of infinitive verbs is challenging for EFLLs from various L1s, including Arab EFLLs (Hafiz et al., 2018; Khatter, 2019), as well as Indonesian EFLLs (Sari and Gulö, 2019), and Iranian EFLLs (Khorshid and Alishah, 2021). These findings shed light

on the complexities of collocations and the challenges that learners face when using them in a second language context.

Having discussed these verb tense issues, I will now discuss the choice of lexical items and the reliance on basic or generic vocabulary. Choice of collocate is the main issue that I will focus on, and many studies confirm that there is still a need for ongoing research to investigate this issue. For example, Al-Shormani and Al-Sohbani (2012) found that Yemeni Arab EFLLs encounter problems with the choice of collocates that could have a direct equivalent in some dialects of Arabic. Many other studies also found errors in the choice of lexical collocates (Farghal and Al-Hamly, 2007). The lexical choices Arab EFLLs make are limited, particularly in the Adverb-Adjective collocation, where the use of adverbs other than 'very' and 'really' (simple adverbs) is infrequent, as indicated by a previous corpus analysis (Alshammari, 2021, p.50). Although the Arab EFLLs may use other adverbs, they were not included in the current research due to their low frequency. A sample analysis of 50 concordance lines from the AEFLL Corpus showed that the adverb 'very' was the most commonly used adverb (52% of all tokens of adverbs). This is likely due to L1 influence, as the adverb *very* جدّاً (jidā) is commonly used in Arabic (Alanazi, 2017, p.23). The NBESs were similar to the Arab EFLLs in 'very' being the most frequent adverb used in the Adverb-Adjective collocation. However, overall, Adverb-Adjective collocations were less frequent in the Arab EFLLs compared to the NBESs, suggesting that it may be a challenging area for Arab EFLLs. This is in line with previous studies that found that Adverb-Adjective collocations were the least commonly used by EFLLs (Demir, 2017) and it seems to be the most difficult collocation pattern (Martyńska, 2004; Shooshtari and Karami, 2013; Boonraksa and Naisena, 2022).

Adverbs seem to be a particularly difficult area for Arab EFLLs as they tend to make many types of error when trying to choose accurate adverbs. It is worth noting that dependency on the adverb 'very' is also present in other EFLLs with different L1s (Lee, 2006, p.11).

The Verb-Noun collocation study indicated that Arab EFLLs are limited in their verb choices. This was evident in the use of certain verbs that are said as such in Arabic. For example, ‘make exercise’ and ‘make revision’ are examples of Arabic interference. The collocation ‘make revision’ is used in contexts when the learners meant ‘study’. This is a direct transfer from Arabic in which ‘revise’ could be used to mean ‘study’. These two examples were also regarded as unacceptable collocations by the two native raters. Dependency on particular vocabulary is common among EFLLs from other L1s, even with German advanced EFLLs (Lennon, 1996).

The Arab EFLLs tended to use Adverb-Adjective collocations in incomplete or fragmentary sentences which can be attributed to incorrect placement of the collocations. Most of the incorrect collocations were placed at the end of the sentence, leaving the meaning incomplete. The use of double verbs could be influenced by the learners’ L1, as they may be using the verb ‘*to be’* to correspond to the verb كان (*kaān*) ‘to be’ in Arabic (e.g. ‘He brother is drive the small car is very happy’). Therefore, there are frequently double verbs, which is clear when the verb ‘is’ follows the subjects ‘He brother’. The collocation ‘very happy’ is placed at the end of a sentence, leading to inaccurate use, as in standard English it should be ‘His brother is very happy to drive the small car’. Double verbs and fragmented sentences have also been observed in the written compositions of Arab EFLLs (Almahameed and Al-Shaikhli, 2017, p.163).

The Arab EFLLs’ collocation data lists indicated that they tend to not use the English articles. The data have shown that the use of articles could be quite challenging in some of the combinations such as the Adverb-Adjective collocation and the Verb-Noun collocation. The error of missing out an obligatory intermediate article within a collocation, found particularly in the infinitive + Verb-Noun collocation, is in line with Crompton (2011), Al-Khasawneh (2014) and Al Jawad and Mansour (2021), who found that not using an obligatory article is the most frequent error usually made by Arab EFLLs. Also, the unnecessary use of articles is a standard error produced by Arab learners (Al-Khasawneh, 2014, p.93), including the adding of or the insertion of articles into English words. The main reason for this error goes back to the differences in the way the article

system operates in Arabic, where the concepts of definite and indefinite reference are represented differently compared to English (Thyab, 2016, p.2). The indefinite in Arabic is indicated by the absence of an article, so *a book* is said as كِتَاب *(kitab)* 'book' without the use of an article equivalent to 'a' (Al-Qadi, 2017, p.73). The error analysis findings regarding the misuse of articles in English support the findings of other scholars who looked EFLLs from other L1 backgrounds, such as Thai (Kampookaew, 2020) and Urdu Pakistani EFLLs, (Sarfraz, 2011).

Regarding the Adjective-Noun collocation, NBESs used many compound nouns whereas the Arab EFLLs did not for this type of collocation. The numbers of collocations extracted for both groups were relatively large, which accords with previous studies that found that the Adjective-Noun collocation type was the most frequent one used by EFLLs. In Balıkçı (2011) study, for example, 652 out of 937 collocations were found to be Adjective-Noun. Since the Adjective-Noun collocation was the most commonly used one by the Arab EFLLs, it suggests that they had no problem with this collocation type, and this is supported by the fact that the collocations contained very few errors. Adjective-Noun collocations having only a few errors was a finding of other scholars who studied Arab EFLLs (Mahmoud, 2005), and Iranian EFLLs (Ganji, 2012; Shooshtari and Karami, 2013), but not EFLLs from Thailand (Yumanee and Phoocharoensil, 2013). However, one issues was that learners tended to use the literal meaning instead of the figurative meaning of the Adjective-Noun collocations, which is the opposite of what native users of English do (Liu, 1999; Tang, 2004.). This could be due to influence from Arabic, as learners tend to use some collocations as they are used in Arabic, creating unacceptable collocations. For example, the Arab EFLLs used 'easy victim' and 'huge headache' in their literal meanings as in Arabic, as Ahmed (2008, p.61) has also found.

Other contextual errors caused by L1 interference include word order errors in the positioning or arrangement of collocations within sentences. These errors are inter-related, combining to form fragmented sentences. It has been observed that the most frequent cause of fragmented sentences was the placement of collocations at the end of a sentence, which occurred in all the four lexical collocations investigated in this study (e.g. '*We saw the kaaba we had very

happy' instead of saying 'when we saw the Kaaba, we had a very happy feeling'). Another example is in the positioning of a collocation in a sentence (e.g. '*area very beautiful' instead of 'very beautiful area'). It is this kind of error that led to Arab EFLLs having fragmented sentences because of where the adverbs were placed in a sentence.

The study's findings also revealed frequent mechanical errors, specifically misspellings and capitalization issues, which do not generally affect the meaning of the collocations in the Arab EFLLs Corpus. However, these spelling errors could have limited the identification of Verb-Noun collocations, as some collocations may have been incorrectly part-of-speech tagged. Arab EFLLs often face spelling challenges, including substituting or omitting syllables due to the differences in vowel usage between Arabic and English. Moreover, the absence of a capitalization system in Arabic contributes to capitalization errors. This outcome is consistent with previous research that identified spelling errors as a result of the differences between Arabic and English, such as (Al-Oudat, 2017; Qasem, 2020).

The results indicate that compound words is a category that causes a problem for Arab EFLLs: e.g. 'break time', and 'nowadays'. The term compound word refers to the use of two independent words, typically belonging to content word classes such as nouns or adjectives, to form a novel word (Yule, 2022). In English, Taylor (1991, p.76) explained that compounds are often formed by combining two or more words into one or more words, "without spaces or hyphens", to create a new word with a new meaning. These compound words may consist of more than one single word such as 'box room', used to describe a small room often used for storage. It is possible that there may be an intralingual error in the use of compounds by the Arab EFLLs if they are using rules from their L1 to form compounds in English. The use of compounds in English can be different from how they are formed in other languages, including Arabic. However, in some languages, including Arabic, compounds are formed by using a space or a hyphen between the words. This can lead to errors in English, where the learner may separate the words when they should be combined, or combine them when they should be separated.

Regarding the word ‘nowadays’, it is a compound word in English and should be written as one word, without a space or a separation between "now" and "adays". It is possible that Arab EFLLs are separating the word based on the rules of their L1, which could be an interlingual error. The variability in the usage of compounds in Arabic and English may account for this type of error, as Betti (2020) observed that there is no fixed pattern for compounds in either language. It is possible that writing compound words as separate words may be influenced by Arabic, given that Arabic is believed to lack compounds. While there are some compounds in Arabic, they are primarily borrowed from foreign languages, such as *capitalism*, which Mughazy (2016) translates as رأسماليه and comprises two words, رأس (literally: ‘head’) and مالية (literally: ‘money’, i.e. مال, plus the nominal suffix, typically used for *-ism* words, اية (*iyya)* ‘verse’). This can be compared to English ‘capitalism’, which is derived from ‘capital’, evolving from *capitale*, a late-Latin word based on *caput*, meaning ‘head’[7]. Errors in compound words have also been observed in other EFLLs, such as Swedish EFLLs (Kusuran, 2017).

Compound words can also be challenging for some native speakers of English. For example, Bowers et al. (2005), found that native English speakers had trouble identifying the target word in compound terms containing orthographically similar embedding words. Also, there are resources[8] that native speakers use to determine whether a compound word should be written as one word or two. The topic of errors in compound words is therefore of interest and warrants further exploration.

5.2 Conclusion

This study has found that some errors in the use of lexical collocations by Arab EFLLs can be linked to specific linguistic differences between English and Arabic.

[7] This has been taken from: (https://en.wikipedia.org/wiki/Capitalism#:~:text=%22Capitalism%22%20is%20derived%20from%20capital,to%20refer%20only%20to%20livestock)

[8] Ritter, R.M. 2003. *The Oxford style manual:[the essential handbook for all writers and editors].* Oxford University Press.

In most cases, it is not possible to produce identical linguistic patterns in English and Arabic due to their significant differences (Hadi, 2014). As a result, the closest acceptable collocation in English may involve variations in wording or pattern. To achieve accuracy and naturalness in collocational patterns, it is necessary to use common expressions that are familiar to the recipient in their target language. Therefore, in the following lines I will highlight some of the differences that I have found that led to errors or unnatural collocation use.

First of all, in the Adverb-Adjective collocation in English, the adverb modifies the adjective (e.g. 'absolutely agree'). However, in Arabic, the adjective precedes the adverb in this collocation (e.g. environmentally friendly is said as صَدِيقٌ لِلبِيئَةِ (*ṣadīq liy al-lbīʾat*)). Both quantitative and qualitative adverbs are usually written after the word they modify (e.g. جميلةٌ جداً (*jamīlat jida*) 'extremely beautiful' in which the adverb 'extremely' follows the adjective as in '*beautiful extremely'. In Arabic grammar, this structure is regarded as an absolute object (a verbal noun, modified by an adjective).

Second, the Verb-Noun and the Verb-Adverb collocations will be discussed together because they both begin with verbs. The main error that was found in these is in the use of the wrong verb tense in the collocations. The past tense in English is indicated by the placement of *-ed* at the end of the verb or by an irregular form that is known and listed in language textbooks. In Arabic, the Verb-Noun collocation is regarded as part of the verbal sentence (e.g. قَرأ كِتَابٍ *(qr'ā ktāb) 'r*ead a book'). The past tense carries perfect aspect in Arabic, and it can be used to express actions in the present or the future. Both the English simple past tense (e.g. 'wrote') and the present perfect (e.g. 'has written') correspond to the Arabic perfect tense (e.g. كَتَبَ (*katb*))[9]. This seems to create a problem because using the wrong tense and generally over-using the past tense is a common problem for Arab EFLLs. For example, Fareh (2014, p.928) found that Arab EFLLs tend to unjustifiably shift tenses (e.g. "*I have a toy at home. It is a teddy bear. I still kept it because it is the best toy I have ever had"*)*.

[9] For further information on the Arabic perfect tense, please refer to Schulz, E. 2004. *A student grammar of modern standard Arabic.* Cambridge, United Kingdom.: Cambridge University Press.

In this Chapter, I have presented the overall findings, provided answers to the research questions and discussed the results in the light of previous studies. In the following chapter, the conclusion, I will present the implications, limitations, and possible future directions of this study.

Chapter 6 Conclusion

Introduction

Having presented the key findings of this study on the use of the four lexical collocations by Arab EFLLs and NBESs in the previous chapter, in this chapter I will discuss the implications of these findings. I will also point out some limitations of this study. Finally, I will suggest some ideas for future research in this area. By summarizing and analysing the results, this chapter aims to provide a comprehensive understanding of the significance of competence in collocation in the process of language learning and its implication for language learners and teachers.

6.1 Contribution

In this section, I will suggest several contributions to the field of knowledge. These contributions will be divided into two categories: one will focus on implications for pedagogy, while the other will examine implications for EFL curriculum designers.

6.1.1 Contribution to the pedagogy

Corpus linguistics is a valuable tool for language learners at all levels of proficiency, not just at the advanced levels typically associated with postgraduate study. For example, Allan (2006) and Boulton (2012) found that a hands-on approach to corpora, where learners manually explore and analyze language data, can be particularly beneficial for advanced learners. This notion has been argued by a number of scholars in the field, including Bill Louw, a corpus linguist based in Zimbabwe who has made a persuasive case for using corpora with undergraduates (Louw and Milojkovic, 2016).

There is, therefore, a definite need to implement a corpus-driven approach when teaching Arab EFLLs. In general, a corpus-driven approach is based on the use of authentic written or spoken data as they are produced by native speakers of English. This approach allows learners to explore a corpus, and search for concordance lines to view how certain words are used in context. This in turn will

enrich their knowledge and vocabulary stock. The creation of materials from corpora using a hand-off approach and hands-on approaches can be viable, yet Boulton noted that hands-on approach to corpora, where learners engage directly with the language data, is more restricted such as training and cluster availability (2012, p.153).

For the Adverb-Adjective collocation, further work needs to be done to improve the use of adverbs by Arab EFLLs. Teachers could use corpora to encourage and facilitate vocabulary learning by generating adverb wordlists to use as materials to teach from, enabling students to engage with these lists in various ways, and getting learners to practice using adverbs in English sentences. By drawing on authentic language data, teachers can provide learners with a rich and varied set of examples that can help to reinforce their understanding of adverbs and improve their overall vocabulary acquisition. In addition, the difference between the two languages should be referred to when the learners are using these wordlists. Wordlists plays a key role in pedagogy as they serve as valuable sources for EFLLs to acquire new words and expand their vocabulary stock. While sharing the same view of Brezina and Gablasove that EFL learning is a “complex and dynamic process” (2017, p.2), I believe that creating a specialised vocabulary wordlist would assist learners to acquire domain-specific termonolgy. Therefore, creating wordlists of different adverbs and adjectives will provide more descriptive detail to the target language. This will enrich learners’ understanding of how certain words can be used in different parts of speech specifically when the Arab EFLLs rely on basic adverbs such as the use of *very*. By enhancing this focus, learners will be made aware of how to use specific words that will help them sound more acceptable. Arab EFLLs tend to have fragmented sentences, which could be due to the influence of the covert and attached pronouns in Arabic. In Arabic, pronouns can be attached to verbs, prepositions and particles, which can make the sentence structure more complex and may result in fragmented sentences in English. Furthermore, many of the fragmented sentences produced by Arab EFLLs lack a subject, which may also be an influence of this aspect of Arabic. In Arabic, the subject is often implicit and may be conveyed through the use of attached pronouns. This may cause learners to

overlook the need for a subject in English and produce sentences that are incomplete or fragmented.

The impact of L1 interference on the use of collocations needs to be considered. In terms of its pedagogical implications, teachers and language practitioners should prioritize explaining categories: both those in the L2 that differ from the learners' L1 and those that are absent. Some scholars have suggested that it is preferable to implement an interference-induced condition that could enhance the acquisition and learning of English collocations, as proposed by Pulido and Dussias (2020). The interference-induced condition suggests that EFLLs can benefit from explicit teaching that focuses on the differences and similarities between an L1 and L2, for example, highlighting how first language (L1) interference can affect the use of collocations in English and providing strategies to mitigate this interference. The main impetus of this condition is to increase learners' awareness of how their L1 might affect their use of collocations and thus this will facilitate greater accuracy in collocation use.

Based on the limited or very basic lexical choices within collocations produced by Arab EFLLs, it seems that the teaching of vocabulary in writing courses could be improved by focusing on how words are used in real contexts, not just their meanings and translations. The following are some suggestions for how to achieve this. (i) teachers should consider teaching vocabulary by providing examples of how words are used in real contexts. There are many ways to bring in real contexts by incorporating technology. One way is through the use of concordancers. This is equally important (ii) to teaching collocations because words do not exist in isolation but are used in combination with other words. For example, if the word 'argument' is being taught, learners should learn that one way to use this word is by using 'put forth an argument'. Therefore, they will get familiarised with the possible acceptable collocations in English instead of relying on their L1 in forming the collocations.

6.1.2 Contribution to EFL Curriculum Designers

The error-analysis approach above in Chapter 4 has shown that many collocations had a verb tense error. In Arabic, there is a perfect tense, which I referred to in (§Chapter 5, in the conclusion); that can correspond to the present and the future forms in English. Emphasising the differences between Arabic perfect tense and English past tense when teaching collocations will assist learners in accurately using verb tenses. Another issue arising from the error-analysis results was that some collocations are part of a bigger phrase. Thus, teaching the possible combinations that may follow these four types of collocation can assist language learners in knowing the right word that might follow a particular collocation. Teaching the possible extensions to collocations will improve learners' language proficiency and awareness of the target language's patterns. For example, the collocation *sort out*[10] can be said as:

1. *You need to sort the problem out.*

2. *You need to sort out the problem.*

The two examples show this collocation can allow an intermediary object within as in Example 1 or following it as in Example 2. Both examples are correct in English; therefore, explaining these possibilities for language learners will improve their use of collocations. However, there is a stylistic difference between the two examples which is explained by Trask (2003) as prepostion stranding. In formal contexts, the verb-adverb-preposition combination is typically kept together, whereas in colloquial contexts, the preposition is often separated and placed at the end of the sentence. It must be mentioned that, in order to investigate this case of stylistic differences, there is a need to determine the level of formality in corpora to gain insights into how language is used in various contexts, which is crucial for linguistic research. The methods for teaching English grammar to non-native learners, especially in the Arabic-speaking context, need to be developed further. Grammar books could improve their coverage of possible collocations and word combinations and what might follow or come in between the two elements combining in a collocation will be very useful in developing learners' collocational knowledge. One possible way to improve the

[10] ***Sort out*** is a phrasal verb. It is interesting to highlight this issue as it is not discussed in this thesis due to some limitations discussed in the limitations section.

comprehension and production of English collocations is to increase exposure to the language in both reception and production (Alzi'abi, 2017, p.67).

6.2 Limitations

It is essential to acknowledge and address the limitations of this study as they might affect the interpretations and the generalizability of the results. By outlining these limitations, this section aims to highlight areas for improvement when the use of collocations is being investigated.

The initial aim of this study was to include five lexical collocation sets in the comparison between Arab EFLLs and NBES. Because of the limited number of Noun-Verb collocations within the data, it was not possible to assess the use of Noun-Verb collocations. Also, as the nouns used by the Arab EFLLs and NBESs were not similar, it was hard to perform a comparative investigation of the Noun-Verb collocation.

Another important limitation in this study was in the use of the CEFR categorization. In this thesis, I used the built-in CEFR tagger in the IntelliText web interface. There were some words within the collocations that were not categorized under CEFR. The tool does not recognise all the possible forms of a lemma, which leads to some words remaining undefined according to the CEFR. For example, the word 'favour' is categorised at CEFR B1 level, but when there is a derivation, and changes the part of speech from Noun to Adverb 'favourably', it was undefined in the IntelliText CEFR tool. This is because 'favourably' is a different word that needs its own CEFR classification. Therefore, having a tool that automatically categorises all forms of a word will assist in implementing the suggested methodology for extending CEFR to measure lexical collocations.

The collocations used cannot be classified under one category by CEFR, because in terms of their vocabulary choices the Arab EFLLs reached the advanced level in some collocations, while their overall CEFR classification

remained at the basic-intermediate level. This suggests that they have developed specific knowledge and skills in the use of collocations, but their overall language proficiency level may not necessarily reflect this. Therefore, it is important to consider the context of collocation use when assessing language proficiency levels, and not rely solely on the use of collocations.

A further limitation is that, unfortunately, this study did not include an investigation of the use of phrasal verbs within the Verb-Adverb collocation. However, phrasal verbs will be the next stage of analysis that will be considered in further publication. Phrasal verbs are a vast topic that was not investigated due to time constraints. In form, phrasal verbs can be either a verb + an adverb (e.g. 'take away/turn off') or a Verb - Adverb - Preposition (e.g. 'looking forward to'). However, based on Darwin and Gray (1999), when followed by a preposition, the verb is a prepositional verb and should not be regarded as a phrasal verb, but rather it is a collocation. A further study investigating phrasal verbs, especially when a preposition is included, would be a fruitful area as many Arab EFLLs tend to face difficulty when using English prepositions (Al-Bayati, 2013; Shakir and Yaseen, 2015; Ngangbam, 2016; Al-Jarf, 2022).

A final limitation of this thesis is that I was not able to carry out my initial aim to use a more inventive approach--the collostructional analysis as proposed by Stefanowitsch and Gries (2003). Using the collostructional approach was not applicable because of some corpus data issues to do with the use of collocations by Arab EFLLs specifically. There were not enough data to perform the approach in the way suggested by Stefanowitsch and Gries. Therefore, the application of collostructional analysis will be presented in future publications.

6.3 Future Research

Although through the comparison between Arab EFLLs and NBESs in the present study has shed light on important aspects of the use of the four lexical collocations, there remains much to be explored. This section outlines several potential avenues for future research that could build upon the findings and

deepen our understanding of Arab EFLLs' use of English lexical collocations. By examining areas for further study and development, I hope to contribute to ongoing efforts to improve language learning outcomes and inform best practices in language teaching.

The thesis could be repeated using a learner corpus that includes a wider range of academic essays by Arab EFLLs to assist in understanding some of the underlying issues that might hinder learners' use of lexical collocations. Corpus linguistics is an approach that can be particularly useful for language learners, as it provides them with access to authentic language data and allows them to develop a deeper understanding of how language is used in context. Even at lower levels of language proficiency, learners can benefit from using corpora to develop their language skills. For example, learners can use corpora to identify common collocations and how certain words behave, which can help them to improve their vocabulary and fluency. They can also use corpora to study common phrases and sentence structures, which can help them to develop more natural-sounding language. While the use of corpora in language learning may have been restricted to postgraduate study in the past, there is growing recognition of the pedagogical value of corpus use at all levels of language proficiency. By introducing learners to the tools and techniques of corpus linguistics, educators can help them to develop the skills and knowledge they need to become more effective language users.

The study has contributed to our understanding of Arab EFLLs' use of English lexical collocations by shedding light on some differences that could affect the production of accurate and acceptable lexical collocations in English. It is important to note that the findings of this study only pertain to the sample that was analysed. However, there are certain aspects of the study that can be used to generalize the results, such as the influence of L1 interference, which is supported by the previous literature. To build on these findings and further develop our understanding of collocation use in writing and speaking among EFLLs in general and Arab EFLLs in particular, it may be useful to expand the sample size and incorporate a wider range of language backgrounds. This would

help to provide a more comprehensive understanding of how various linguistic backgrounds impact collocation use.

Regarding the first limitation stated in Section 6.2, further study could attempt to overcome it in order to assess the use of Noun-Verb collocations and Verb-Noun collocations. This could lead to exciting findings because of the importance and prevalence of verbal and nominal Arabic sentence structures; the Arabic nominal sentence is similar to a Noun-Verb collocation, and the verbal form is identical to the Verb-Noun collocation. This would also contribute to our understanding of the factors that lead to faults or difficulties when Arab EFLLs use these two collocations in English. Understanding these issues will assist not only in developing Arab EFLLs' collocation competence but also that of many other EFLLs, as collocations seem to be a complex area for learners from different L1 backgrounds to master (Vilkaitė, 2017; Borro, 2021).

A potential future direction could be to explore the relationship between learners' proficiency levels and the number of mistakes they make in complex collocations, using the CEFR classification as a guide. This could involve analysing data from a larger sample of learners across different proficiency levels, as well as investigating the impact of other factors such as language background or learning strategies. Additionally, future research could focus on developing instructional materials or interventions that target the specific challenges associated with mastering complex collocations, based on the insights gained from this type of analysis. Ultimately, such efforts could help to enhance language learning outcomes and support more effective and efficient language teaching practices. The following is one question that I aim to investigate in further publications:

Is there a relationship between the Common European Framework of Reference (CEFR) levels and the frequency and type of collocational errors made by Arab EFLLs? In particular:

(i) What are the most common types of errors made by Arab EFL learners across the different CEFR levels?

(ii) Is there a significant difference in the frequency and type of collocational errors made by Arab EFLLs in lower versus higher CEFR levels?

Furthermore, it was noted above in Section 6.2, that the analysis did not include an investigation into the use of phrasal verbs. Due to the significant scope of this topic, a more extensive study would be required to examine how Arab English as a Foreign Language Learners (EFLLs) use phrasal verbs. This area of research is particularly intriguing because phrasal verbs are not commonly found in the Arabic language (Mahmoud, 2005). Therefore, a further study could provide valuable insights into how Arab EFLLs incorporate the use of phrasal verbs into their writing and speaking. This is a further potentially rich future direction of research.

References

Abd Ai-Qadir, M. 2015. "Voice collocations and its role in "coherence" in the narrative text: (Ghadah Al-Samman's "Beirut 75"). *Tishreen University Journal for Research and Scientific Studies-Arts and Humanities Series.* **37**(2), pp.112-129.

Abdul Ridha, N.S. and Al-Riyahi, A.A. 2011. Lexical collocation errors in the writings of Iraqi EFL learners. *Journal of the College of Arts. University of Basrah.* **58**, pp.24-51.

Abedi, Z. and Mobaraki, M. 2014. The effect of grammatical collocation instruction on understanding ESP texts for undergraduate computer engineering students. *Journal of Language Teaching & Research.* **5**(3), pp.631-641.

Abu Khaled, D.Y. 2020. Palestinian EFL learners' use of English lexical collocations. *International Journal of Linguistics, Literature and Translation.* **3**(3), pp.60-71.

Abu-Chacra, F. 2007. *Arabic: an essential grammar.* London and New York: Routledge.

Abushihab, I., El-Omari, A.H. and Tobat, M. 2011. An analysis of written grammatical errors of Arab learners of English as a foreign language at Alzaytoonah private university of Jordan. *European Journal of Social Sciences.* **20**(4), pp.543-552.

Adolphs, S. 2006. Introducing electronic text analysis: a practical guide for language and literary studies. Routledge.

Agah, M. and Soori, A. 2015. Comparative study of collocation among the languages. *Language in India.* **15**(1), pp.527-537.

Ahmed, M.B. 2008. Collocation errors made by Iraqi EFL learners at university level. *Tikrit University Journal for Humanities.* **15**(11), pp.55-73.

Ahad, N.A., Abdullah, S., Lai, C.H. and Ali, N.M. 2012. Relative Power Performance of t-test and Bootstrap Procedure for Two-Sample. *Pertanika Journal of Science & Technology.* **20**(1), pp.43-52.

Al Aqad, M. 2013. Syntactic analysis of Arabic adverb's between Arabic and English: X Bar Theory. *International Journal of Language and Linguistics.* **1**(3), pp.70-74.

Al Jawad, A.S.H. and Mansour, A.A. 2021. An exploration of grammatical errors in written English of Libyan EFL students with special reference to Arabic as their first-language. *Journal of World Englishes and Educational Practices.* **3**(7), pp.07-25.

Al-Bayati, W.A.-W.T. 2013. Errors made by Iraqi EFL undergraduates in the use of prepositions. *Bulletin of the Transilvania University of Braşov, Series IV: Philology & Cultural Studies.* **6 (55)**(1), pp.41-56.

Al-Inbari, F.A.Y. and Al-Wasy, B.Q.M. 2022. The impact of automated writing evaluation (AWE) on EFL learners' peer and self-editing. *Education and Information Technologies.*, **28**(6645-6665), pp.1-21.

Al-Jarf, R. 2022. Undergraduate student-translators' difficulties in translating English word + preposition collocations to Arabic. *International Journal of Linguistics Studies (IJLS).* **2**(2), pp.60-72.

Al-Khasawneh, F.M. 2014. Error analysis of written English paragraphs by Jordanian undergraduate students: a case study. *International Journal of English Language, Literature and Humanities.* **2**(8), pp.85-100.

Al-Khresheh, M.H. 2010. Interlingual interference in the English language word order structure of Jordanian EFL learners. *European Journal of Social Sciences.* **16**(1), pp.105-116.

Al-Mahrooqi, R., Denman, C. and Sultana, T. 2016. Factors contributing to the survival of standard Arabic in the Arab world: an exploratory study. *Pertanika Journal of Social Sciences & Humanities.* **24**(3), pp.1177-1191.

Al-Oudat, A. 2017. Spelling errors in English writing committed by English-major students at BAU. *Journal of Literature, Languages and Linguistics.* **32**, pp.43-47.

Al-Qadi, M.M. 2017. English article system errors among Saudi Arab EFL learners: a case of the preparatory year english program learners of King Saud University. *International Journal of English Language & Translation Studies.* **5**(1), pp.69-77.

Al-Shormani, M.Q. and Al-Sohbani, Y.A. 2012. Semantic errors committed by Yemeni University learners: classifications and sources. *International Journal of English Linguistics.* **2**(6), pp.120-139.

Alanazi, M. 2017. On the production of synonyms by Arabic-speaking EFL learners. *International Journal of English Linguistics.* **7**(3), pp.17-28.

Alanazi, Z. 2022. Corpus-based analysis of near-synonymous verbs. *Asian-Pacific Journal of Second and Foreign Language Education.* **7**(1), pp.1-25.

Alangari, M.A. 2019. A corpus-based study of verb-noun collocations and verb complementation clause structures in the writing of advanced Saudi learners of English. Doctor of Philosophy thesis, University of Reading.

Alghamdi, T.A. 2019. The inter-lingual transfer of Arabic in the English writings of Arab EFL students. *International Journal of Arts and Commerce.* **8**(8), pp.15-30.

Alhassan, H., Quarcoo, M.A.B. and Addae, A. 2022. Verb-noun collocations in newspaper editorials in Ghana: a corpus-based analysis. *Contemporary Journal of African Studies.* **9**(2), pp.85-119.

Alhuthaily, Y. 2000. The Role of Verbs in Sentence Formation دور الفعل في بنية الجملة. Tunisia Dar Sahar Publication.

Allan, R. 2006. Data-driven learning and vocabulary: investigating the use of concordances with advanced learners of English. *Occasional paper-Centre for Language and Communication Studies.* (66), pp.1-52.

Almahameed, Y.S. and Al-Shaikhli, M. 2017. Understanding syntactic and semantic errors in the composition writing of jordanian EFL learners. *International Journal of Applied Linguistics and English Literature.* **6**(6), pp.158-164.

Alqaed, M.A. 2017. Perceptions on l2 lexical collocation translation with a focus on English-Arabic. *Journal of Education and Practice.* **8**(6), pp.128-133.

Alshammari, R.R.M. 2021. Adverb-adjective collocation use by Arab EFLLs and British English native speakers: a comparative corpus-based study. *The Language Scholar.* **9**, pp.24-65.

Alsop, S. and Nesi, H. 2009. Issues in the development of the British Academic Written English (BAWE) corpus. *Corpora.* **4**(1), pp.71-83.

Alsulayyi, M.N. 2015. The use of grammatical collocations by advanced Saudi EFL learners in the UK and KSA. *International Journal of English Linguistics.* **5**(1), pp.32-43.

Alzi'abi, S.E. 2017. Guessing verb–adverb collocations: Arab EFL learners' use of electronic dictionaries. *Lexikos.* **27**, pp.50-77.

Arabet, M.A. 2020. The effect of collocational success on essay grading: a case of Algerian EFL learners. *Revue Sciences Humaines.* **31**(4), pp.619-630.

Ariffin, A. and Abdi, B.M. 2020. The relationship between collocation competence and writing skills of EFL learners. *AJELP: Asian Journal of English Language and Pedagogy.* **8**(1), pp.41-52.

Astia, I. and Yunianti, S. 2020. Corpus-based analysis of the most frequent adjective on Covid-19. *Indonesian Journal of EFL and Linguistics.* **5**(2), pp.505-515.

Aziz, Y.Y. 1982. Cultural problems of English-Arabic translation. *Babel: International Journal of Translation.* **28**(1), pp.25-29.

Baheej, K. 2015. The use of prepositions in English among Arab students. *Studia Universitatis Moldaviae-Științe ale Educației.* **5**(85), pp.119-122.

Bahns, J. 1993. Lexical collocations: a contrastive view. *ELT Journal.* **47**(1), pp.56-63.

Bahumaid, S. 2006. Collocation in English-Arabic translation. *Babel.* **52**(2), pp.133-152.

Baker, M. 1992. In other words: a textbook on translation. London: Routledge.

Baker, P. 2009. The BE06 corpus of British English and recent language change. *International Journal of Corpus Linguistics.* **14**(3), pp.312-337.

Balcı, Ö. and Çakır, A. 2012. Teaching vocabulary through collocations in EFL classes: the case of Turkey. *International Journal of Research Studies in Language Learning.* **1**(1), pp.21-32.

Balıkçı, G. 2011. The use of collocations by advanced learners of English: noun–noun and adjective-noun collocations. In: *1st International Conference on Foreign Language Teaching and Applied Linguistics*, *Sarajevo*. pp.513-520.

Bardovi-Harlig, K. and Sprouse, R.A. 2018. Negative versus positive transfer. *The TESOL Encyclopedia of English Language Teaching.*, pp.1-6.

Bartsch, S. 2004. Structural and functional properties of collocations in English: a corpus study of lexical and pragmatic constraints on lexical co-occurrence. Gunter Narr Verlag.

Basal, A. 2017. Learning collocations: effects of online tools on teaching English adjective-noun collocations. *British Journal of Educational Technology.* **50**(1).

Benson, M., Benson, E. and Ilson, R. 1997. *The BBI dictionary of English: a guide to word combinations.* John Benjamins Publishing Company.

Betti, M.J. 2020. Compounding in English and Arabic: a contrastive study. *International Journal of Research in Social Sciences and Humanities.* **10**(III), pp.1-13.

Biber, D. and Barbieri, F. 2007. Lexical bundles in university spoken and written registers. *English for Specific Purposes.* **26**(3), pp.263-286.

Biber, D. and Conrad, S. 1999. Lexical bundles in conversation and academic prose. In: Hilde, H. and Oksefjell, S. eds. *Out of Corpora.* Amesterdam: Rodopi pp.181-190.

Biber, D., Conrad, S. and Cortes, V. 2004. If you look at...: lexical bundles in university teaching and textbooks. *Applied Linguistics.* **25**(3), pp.371-405.

Biber, D., Conrad, S. and Reppen, R. 1994. Corpus-based approaches to issues in applied linguistics. *Applied Linguistics.* **15**(2), pp.169-189.

Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. 1999. *Longman grammar of spoken and written English.* London: Longman.

Boonraksa, T. and Naisena, S. 2022. A study on English collocation errors of Thai EFL students. *English Language Teaching.* **15**(1), pp.164-177.

Borro, I. 2021. Enhanced incidental learning of formulaic sequences by Chinese learners of Italian. Doctor of Philosophy thesis, University of Portsmouth.

Boulton, A. 2012. Hands-on/hands-off: alternative approaches to data-driven learning. *Input, Process and Product: Developments in Teaching and Language Corpora.* Masarykova univerzita, pp.152-168.

Bowers, J.S., Davis, C.J. and Hanley, D.A. 2005. Automatic semantic activation of embedded words: Is there a “hat” in “that”? *Journal of Memory and Language.* **52**(1), pp.131-143.

Bley-Vroman, R. and Yoshinaga, N., 1992. Road and narrow constraints on the English dative alternation: Some fundamental differences between native speakers and foreign language learners

Brashi, A. 2009. Collocability as a problem in L2 production. *Reflections on English Language Teaching* **8**(1), pp.21-34.

Brashi, A.S. 2005. *Arabic collocations: implications for translations.* Doctor of Philosophy thesis, University of Western Sydney.

Brezina, V. 2018a. *Statistics in corpus linguistics: A practical guide.* Cambridge: Cambridge University Press.

Brezina, V. 2018b. Vocabulary, frequency, dispersion and diversity. *statistics in corpus linguistics: a practical guide.* Cambridge; New York.: Cambridge University Press.

Brezina, V., McEnery, T. and Wattam, S. 2015. Collocations in context: a new perspective on collocation networks. *International Journal of Corpus Linguistics.* **20**(2), pp.139-173.

Brezina, V. and Gablasova, D. 2017. How to produce vocabulary lists? Issues of definition, selection and pedagogical aims. A response to Gabriele Stein. *Applied Linguistics.* **38**(5), pp.764-767.

Brezina, V. and McEnery, T. 2020. Introduction to corpus linguistics. *The Routledge Handbook of Second Language Acquisition and Corpora.* pp.11-22.

Brown, H.D. 2000. *Principles of language learning and teaching.* Fifth Edition ed. New York: Pearson Education.

Brumfit, C. 1984. Communicative methodology in language teaching: the roles of fluency and accuracy. Cambridge: Cambridge University Press.

A frequency dictionary of Arabic: core vocabulary for learners . 2011. s.v.

Bulté, B. and Housen, A. 2012. Defining and operationalising L2 complexity. In: Housen, A., Kuiken, F., and Vedder, I. eds. *Dimensions of L2 performance and proficiency: complexity, accuracy and fluency in SLA.* John Benjamin Publishing Company, pp.23-46.

Byrne, G. 2007. A statistical primer: understanding descriptive and inferential statistics. *Evidence Based Library and Information Practice.* **2**(1), pp.32-47.

Chaaraoui, J. 2017. Grammar accuracy, language threshold level, and degree of bilingualism in the Saudi EFL learner's interlanguage. *International Journal of Language and Linguistics.* **4**(3), pp.86-101.

Chambers, F. 1997. What do we mean by fluency? *System.* **25**(4), pp.535-544.

Chang, Y. 2018. Features of lexical collocations in L2 writing: a case of Korean adult learners of English. *English Teaching.* **73**(2), pp.3-36.

Chejne, A.G. 1969. *The Arabic language: Its role in history.* reissue ed. Minneapolis: University of Minnesota Press, 1969.

Cheng, W., Greaves, C., Sinclair, J.M. and Warren, M. 2009. Uncovering the Extent of the Phraseological Tendency: Towards a Systematic Analysis of Concgrams. *Applied linguistics.* **30**(2), pp.236-252.

Church, K. and Hanks, P. 1990. Word association norms, mutual information, and lexicography. *Computational Linguistics.* **16**(1), pp.22-29.

Corder, S.P. 1975. Error analysis, interlanguage and second language acquisition. *Language teaching surveys and studies.* **8**(4), pp.201-218.

Cowie, A.P. 1998. *Phraseology: theory, analysis, and applications.* Oxford: Oxford University Press.

Cowie, A.P. and Howarth, P. 1996. Phraseological competence and written proficiency. In: Blue, G.M. and Mitchell, R. eds. *Language and Education.* Mitchell., pp.80-93.

Crompton, P. 2011. Article errors in the English writing of advanced L1 Arabic learners: the role of transfer. *Asian EFL Journal.* **50**(1), pp.4-35.

Crystal, D. and Alan, C.L.Y. 2023. *A Dictionary of Linguistics and Phonetics.* Seventh ed. John Wiley & Sons.

Dąbrowska, E. 2010. Naive v. expert intuitions: an empirical study of acceptability judgments. *The Linguistic Review.* **27**(1), pp.1-23.

Darragh, G. 2000. A to zed, a to zee: a guide to the differences between British and American English. Editorior Stanley.

Darwin, C.M. and Gray, L.S. 1999. Going after the phrasal verb: An alternative approach to classification. *Tesol Quarterly.* **33**(1), pp.65-83.

Davis, C.R. 2010. Homogeneity of variance. *Encyclopaedia of Research Design.* pp.578-580.

Dazdarevic, S., Zoranic, A. and Fijuljanin, F. 2015. Benefits of corpus-based approach to language teaching. *Balkan Distance Education Network-BADEN Newsletter Year III.* **3**(7), pp.1-4.

Deáková, J. 2013. Selected lexical differences between British and American English: a corpus-supported view. Bachelor of Arts thesis, Charles University.

Deane, P. 2005. A nonparametric method for extraction of candidate phrasal terms. In: Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05). pp.605-613.

Demir, C. 2017. Lexical collocations in English: a comparative study of native and non-native scholars of English. *Journal of Language and Linguistic Studies.* **13**(1), pp.75-87.

Derakhshan, A. and Karimi, E. 2015. the interference of first language and second language acquisition. *Theory and Practice in Language Studies.* **5**(10), pp.2112-2117.

Deveci, T. 2018. Analysis of collocations in a book chapter and learners' corpus and teaching implications. *Journal of Language Education and Research.* **4**(1), pp.24-40.

Diab, N. 1997. The transfer of Arabic in the English writings of Lebanese students. *The ESPecialist.* **18**(1), pp.71-83.

Dickins, J. 2010a. Basic sentence structure in Sudanese Arabic. *Journal of Semitic Studies.* **55**(1), pp.237-261.

Dickins, J. 2010b. Junction in English and Arabic: syntactic, discoursal and denotative features. *Journal of Pragmatics.* **42**(4), pp.1076-1136.

Dickins, J. 2011. Faʿ* l forms in Sudanese Arabic: the reassertion of morphology. *Zeitschrift Für Arabische Linguistik.* **160**(53), pp.36-67.

Dickins, J. 2020. Thematic structure and para-syntax: Arabic as a case study. Oxon, UK: Routledge.

Dickins, J., Hervey, S. and Higgins, I. 2016. Thinking Arabic translation: a course in translation method: Arabic to English. Oxon: Routledge.

Ding, C. and Reynolds, B.L. 2019. The effects of L1 congruency, L2 proficiency, and the collocate-node relationship on the processing of L2 English collocations by L1-Chinese EFL learners. *Review of Cognitive Linguistics. Published under the auspices of the Spanish Cognitive Linguistics Association.* **17**(2), pp.331-357.

Doró, K. 2007. The use of high-and low-frequency verbs in English native and non-native student writing. *Second language lexical processes: Applied Linguistic and Psycholinguistic Perspectives.* Clevedon: Multilingual Matters Ltd., pp.117-129.

Duan, M. and Qin, X. 2012. Collocation in English teaching and learning. *Theory and Practice in Language Studies.* **2**(9), pp.1890-1894.

Dukali, A.A. 2018. An Error analysis of the use of lexical collocations in the academic writing of Libyan EFL university students. *Journal of Second Language Teaching & Research.* **6**(2), pp.55-91.

Dulay, H.C. and Burt, M.K. 1972. GOOFING: An indicator of children's second language learning strategies 1. *Language Learning.* **22**(2), pp.235-252.

Dunning, T. 1993. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics.* **19**(1), pp.61-74.

Durrant, P. 2013. Discipline and level specificity in university students' written vocabulary. *Applied Linguistics.* **35**(3), pp.328-356.

El-Dakhs, D.A.S. 2015. The lexical collocational competence of Arab undergraduate EFL learners. *International Journal of English Linguistics.* **5**(5), pp.60-74.

El-Gazzar, N.K. 2006. Lexical cohesive devices in Arab students' academic writing: Implications for teaching vocabulary. Master of Arts thesis.

Essberger, J. 2000. *English prepositions list*. [online]. 7th Edition ed.

Evert, S. 2005. The statistics of word cooccurrences: word pairs and collocations. thesis, Universität Stuttgart.

Evert, S. 2008. Corpora and collocations. *corpus linguistics. an international handbook.* Berlin: Mouton de Gruyter., pp.1-53.

Evert, S. 2010. *Computational approaches to collocations.* [Online]. Available from: http://www.col-locations.de/

Evert, S., Uhrig, P., Bartsch, S. and Proisl, T. 2017. E-view-alation—a large-scale evaluation study of association measures for collocation identification. *Proceedings of eLex.* pp.531-549.

Faaß, G. and Eckart, K. 2013. SdeWaC–a corpus of parsable sentences from the web. In: Language Processing and Knowledge in the Web: 25th International Conference, GSCL 2013, Darmstadt, Germany, September 25-27, 2013. Proceedings, Darmstadt, Germany: Springer, pp.61-68.

Fan, M. 2009. An exploratory study of collocational use by ESL students–a task-based approach. *System.* **37**(1), pp.110-123.

Fareh, S. 2014. Macrolinguistic errors in Arab EFL learners' essays. *Procedia-Social and Behavioral Sciences.* **141**, pp.923-933.

Farghal, M. and Al-Hamly, M. 2007. Lexical collocations in EFL writing. *Journal of Asia TEFL.* **4**(2), pp.69-94.

Faris, A.A.H. and Sahu, R.A. 2013. The translation of english collocations into Arabic: problems and solutions. *Journal of the College of Arts. University of Basra.* **64**, pp.51-66.

Farooqui, A.S. 2016. A corpus-based study of academic-collocation use and patterns in postgraduate computer science students' writing. Doctor of Philosophy thesis, University of Essex.

Fawi, F. and Delmonte, R. 2015. Italian-Arabic domain terminology extraction from parallel corpora. *Proceedings of the Second Italian Conference on Computational Linguistics - CLiC-it 2015.* **130**, pp.130-134.

Fillmore, C.J. 1992. Corpus linguistics or computer-aided armchair linguistics. In: *Directions in Corpus Linguistics. Proceedings of Nobel Symposium*: De Gruyter Mouton pp.35-60.

Firth, J. 1968. A synopsis of linguistic theory, 1930-1955. In: Palmer, F.R. ed. *In Selected papers of J. R. Firth 1952-1959.* London: Longman., pp.10-32.

Firth, J.R. 1957. *Papers in linguistics 1934-1951.* Oxford, England.: Oxford University Press.

Flowerdew, L. 2014. Corpus-based analyses in eap. *Academic Discourse.* Routledge, pp.105-124.

Futagi, Y., Deane, P., Chodorow, M. and Tetreault, J. 2008. A computational approach to detecting collocation errors in the writing of non-native speakers of English. *Computer Assisted Language Learning.* **21**(4), pp.353-367.

Gablasova, D., Brezina , V. and McEnery, T. 2017. Collocations in corpus-based language learning research: identifying, comparing, and interpreting the evidence. *Language Learning.* **67**(S1), pp.155-179.

Ganji, M. 2012. On the effect of gender and years of instruction on Iranian EFL learners' collocational competence. *English Language Teaching.* **5**(2), pp.123-133.

Gardner, D. and Davies, M. 2007. Pointing out frequent phrasal verbs: a corpus-based analysis. *TESOL Quarterly.* **41**(2), pp.339-359.

Garner, J., Crossley, S. and Kyle, K. 2018. Beginning and intermediate L2 writer's use of N-grams: an association measures study. *International Review of Applied Linguistics in Language Teaching.* **58**(1), pp.51-74.

Gass, S.M. and Selinker, L. 2008. *Second language acquisition: an introductory course.* Third edition. New York and London: Routledge.

Gerald, B. 2018. A brief review of independent, dependent and one sample t-test. *International Journal of Applied Mathematics and Theoretical Physics.* **4**(2), pp.50-54.

Ghazala, H. 1993. Translating collocations: Arabic-English (in Arabic). Turjuman.

Ghazala, H. 2007. *Dictionary of collocations English-Arabic.* Beirut.: Dâr al-'ilm li-l-malâyîn.

Gholitabar, M. and Kamali, A.D. 2012. The Quran and the development of Arabic linguistics. *2012 International Conference on Language, Medias and Culture.* **33**, pp.26-30.

Gibbons, J.D. and Chakraborti, S. 1991. Comparisons of the Mann-Whitney, student's t, and alternate t tests for means of normal distributions. *The Journal of Experimental Education.* **59**(3), pp.258-267.

Gizatova, G., Shangaraeva, L., Zorina, A. and Zakirova, L. 2020. Practical applications of using language corpora in an EFL classroom. In: *INTED2020 Conference, 2020, Valencia, Spain*.

Granger, S. 1998a. The computer learner corpus: a versatile new source of data for SLA research. London. New York: Addison Wesley Longman.

Granger, S. 1998b. Prefabricated patterns in advanced EFL writing: collocations and lexical phrases. *Phraseology: Theory, Analysis, and Applications.* Oxford: Oxford University Press, pp.145-160.

Granger, S., Hung, J. and Petch-Tyson, S. 2002. Computer learner corpora, second language acquisitio, and foreign language teaching. Amesterdam: John Benjamins.

Greaves, C. and Warren, M. 2010. What can a corpus tell us about multi-word units? *The Routledge Handbook of Corpus Linguistics.* Routledge, pp.212-226.

Greenbaum, S. 1970. Verb-intensifier collocations in English: Janua Linguarum, Series Minor 86. *The Hague: Mouton.*

Greenbaum, S. 1974. some verb-intensifier collocations in American and British English. *American Speech.* **49**(1/2), pp.79-89.

Gries, S. 2010. Useful statistics for corpus linguistics. *A Mosaic of Corpus Linguistics: Selected Approaches.* **66**, pp.269-291.

Gries, S. 2017. Basic significance testing. In: Podesva, R.J. and Sharma, D. eds. *Research Methods in Linguistics.* Cambridge: Cambridge University Press, pp.316-336.

Gries, S. and Newman, J. 2017. Creating and using corpora. In: Podesva, R. and Sharma, D. eds. *Research Methods in Linguistics.* Cambridge: Cambridge University Press, pp.257-287.

Gries, S.T. and Ellis, N.C. 2015. Statistical measures for usage-based linguistics. *Language Learning.* **65**(S1), pp.228-255.

Grimm, P.S. 2009. Collocation in modern standard Arabic revisited. *Zeitschrift für arabische Linguistik.* (51), pp.22-41.

Guz, E. 2016. Refining the methodology for investigating the relationship between fluency and the use of formulaic language in learner speech. *Research in Language.* **14**(2), pp.95-122.

Gyllstad, H. 2012. Grammatical collocation. *The Encyclopedia of Applied Linguistics.*, pp.1-6.

Habtoor, H.A. and Al-Swaidan, R.H. 2019. Familiarity with collocations in EFL context and strategies utilized in translating them into Arabic. *Journal of Language Teaching and Research.* **10**(4), pp.724-737.

Haden, P. 2019. Inferential statistics. The Cambridge handbook of computing education research. pp.133-172.

Hadi, S.A.K. 2014. Iraqi EFL learners' use of English lexical collocations. *The Journal of Babylon University.* **22**(3), pp.566-579.

Hafiz, M.S., Abdul-Majeed, A. and Gul Sher, K. 2018. Analysis of syntactic errors in English writing: a case study of Jazan University preparatory year students. *Journal of Education and Practice.* **9**(11), pp.113-120.

Hago, O.E. and Ali, M.H. 2015. Assessing English syntactic structures experienced by Sudanese female students at secondary schools, (2013-2014). *Arab World English Journal (AWEJ).* **6**(2), pp.213-241.

Hair, J., Hollingsworth, C.L., Randolph, A.B. and Chong, A.Y.L. 2017. An updated and expanded assessment of PLS-SEM in information systems research. *Industrial Management & Data Systems.* **117**(3), pp.442-458.

Halliday, M.A.K. 1976. *System and function in language: selected papers.* Oxford University Press.

Hammarberg, B. 2010. The languages of the multilingual: some conceptual and terminological issues. *IRAL-International Review of Applied Linguistics in Language Teaching.* **48**(2-3), pp. 91-104.

Harlow, L.L. 1986. Behavior of some elliptical theory estimators with nonnormal data in a covariance structures framework: A Monte Carlo study. thesis, ProQuest Information & Learning.

Harwood, N., Austin, L. and Macaulay, R. 2009. Proofreading in a UK university: proofreaders' beliefs, practices, and experiences. *Journal of Second Language Writing.* **18**(3), pp.166-190.

Hashim, A. 2017. Crosslinguistic influence in the written English of Malay undergraduates. *Journal of Modern Languages.* **12**(1), pp.60-76.

Heydari, P. and Bagheri, M.S. 2012. Error analysis: sources of L2 Learners' errors. *Theory and Practice in Language Studies.* **2**(8), pp.1583-1589.

Hong, A.L., Rahim, H.A., Hua, T.K. and Salehuddin, K. 2011. Collocations in Malaysian English learners' writing: a corpus-based error analysis. *3L:The Southeast Asian Journal of English Language Studies.* **17**(Special Issue), pp.31-44.

Howarth, P. 1998. Phraseology and second language proficiency. *Applied Linguistics.* **19**(1), pp.24-44.

Hu, K. 2020. Corpus-based translation studies and translation cognition research: similarity and convergence. In Hu, K., Kim, K. eds. *Corpus-based Translation and Interpreting Studies in Chinese Contexts.* Palgrave Macmillan, Cham, pp.221-236.

Hughes, A. 2003. *Testing for language teachers.* Second Edition. Cambridge: Cambridge University Press.

Hughes, R. 2002. *Teaching and researching speaking Rebecca Hughes.* Great Britain Pearson Education Limited

Hunston, S. 2002. *Corpora in applied linguistics.* Cambridge: Cambridge University Press.

Hunston, S. 2008. Starting with the small words: patterns, lexis and semantic sequences. *International Journal of Corpus Linguistics.* **13**(3), pp.271-295.

Hunston, S. 2010. Corpus approaches to evaluation: phraseology and evaluative language. New York: Routledge.

Husamaddin, K. 1985. Idiomatic expressions. Cairo: Maktabat Al-Anjilu Al-Masriyah.

Igaab, Z.K. and Abdulhasan, H. 2018. Collocation in English and Arabic: a contrastive study. *English Language and Literature Studies.* **8**(4), pp.89-103.

Izwaini, S. 2015. Patterns of lexical collocation in Arabic. *Zeitschrift für Arabische Linguistik.* (61), pp.72-100.

Izwaini, S. 2016. The translation of Arabic lexical collocations. *Translation and Interpreting Studies.* **11**(2), pp.306-328.

Jaspal, R. and Coyle, A. 2010. "Arabic is the language of the Muslims–that's how it was supposed to be": exploring language and religious identity through

reflective accounts from young British-born South Asians. *Mental Health, Religion and Culture.* **13**(1), pp.17-36.

Johansson, S. 2008. Contrastive analysis and learner language: A corpus-based approach. Compendium for the course ENG2152–Contrastive and Learner Language Analysis. University of Oslo.

Jones, T.A. 1969. Skewness and kurtosis as criteria of normality in observed frequency distributions. *Journal of Sedimentary Research.* **39**(4), pp.1622-1627.

Kampookaew, P. 2020. An analysis of grammatical errors made by Thai EFL university students in an EAP writing class: issues and recommendations. *rEFLections.* **27**(2), pp.246-273.

Ke, Y. 2013. The phraseology of phrasal verbs in English: a corpus study of the language of Chinese learners and native English writers. Doctor of Philosophy thesis, University of Birmingham.

Kennedy, G. 2014. *An introduction to corpus linguistics.* New York: Routledge.

Keshavarz, M.H. 2012. Contrastive analysis, error analysis, and interlanguage. Rahnama Press.

Khaled, D.Y.A. 2020. Palestinian EFL learners' use of English lexical collocations. *International Journal of Linguistics, Literature and Translation.* **3**(3), pp.60-71.

Khansir, A.A. 2012. Error analysis and second language acquisition. *Theory and Practice in Language Studies.* **2**(5), pp.1027-1032.

Khatter, S. 2019. An analysis of the most common essay writing errors among EFL Saudi female learners (Majmaah University). *Arab World English Journal.* **10**(3), pp.364-381.

Khoja, H.Y.Y. 2019. Exploring the use of collocation in the writing of foundation-year students at King Abdulaziz University. Doctor of Philosophy thesis, The University of Leeds.

Khorshid, S.R. and Alishah, A.R. 2021. An analysis of the frequency of acceptable and unacceptable collocations among Iranian EFL learners. *International Journal of Education and Social Science Research.* **4**(3), pp.86-94.

Khuwaileh, A.A. 2000. Vocabulary in LSP: a case study of phrases and collocations. *Babel.* **46**(2), pp.97-111.

Kilgarriff, A. 2006. Collocationality (and how to measure it). In: In Proceedings of the 12th Euralex International Congress on Lexicography (EURALEX). Turin, Italy.: Springer-Verlag., pp.997–1004.

Kilgarriff, A., Charalabopoulou, F., Gavrilidou, M., Johannessen, J.B., Khalil, S., Johansson Kokkinakis, S., Lew, R., Sharoff, S., Vadlapudi, R. and Volodina, E. 2014. Corpus-based vocabulary lists for language learners for nine languages. *Language Resources & Evaluation.* **48**(1), pp.121-163.

Kniaź, M. and Zawrotna, M. 2021. Embedded English verbs in Arabic-English code-switching in Egypt. *International Journal of Bilingualism.* **25**(3), pp.622-639.

Knowles, G. and Don, Z.M. 2004. The notion of a “lemma”: Headwords, roots and lexical sets. *International Journal of Corpus Linguistics.* **9**(1), pp.69-81.

Kruskal, W.H. and Wallis, W.A. 1952. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association.* **47**(260), pp.583-621.

Kunchukuttan, A. and Damani, O.P. 2008. A system for compound noun multiword expression extraction for Hindi. In: *Proceedings of ICON-2008: 6th International Conference on Natural Language Processing., India*. Macmillan Publishers., pp.20-29.

Kusuran, A. 2017. *L2 English spelling error analysis: an investigation of English spelling errors made by Swedish senior high school students*. Faculty of Humanities and Social Sciences, diva portal.org. Available from: http://www.diva-portal.org/smash/get/diva2:1078118/FULLTEXT01.pdf

Kyle, K. and Crossley, S.A. 2018. Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *The Modern Language Journal.* **102**(2), pp.333-349.

Labov, W. 1973. The boundaries of words and their meanings. In: Bailey, C.J. and Shuy, R.W. eds. *New Ways of Analyzing Variation in English.* Washington DC: Georgetown University Press, pp.340-373.

Lado, R. 1957. *Linguistics across cultures.* MI: University of Michigan Press.

Lahuerta, A.C. 2018. Study of accuracy and grammatical complexity in EFL writing. *International Journal of English Studies.* **18**(1), pp.71-89.

Lambert, V.A. and Lambert, C.E. 2012. Qualitative descriptive research: an acceptable design. *Pacific Rim International Journal of Nursing Research.* **16**(4), pp.255-256.

Langsford, S., Perfors, A., Hendrickson, A.T., Kennedy, L.A. and Navarro, D.J. 2018. Quantifying sentence acceptability measures: reliability, bias, and variability. *glossa: A Journal of General Linguistics.* **3**(1), pp.1-34.

Lau, J.H., Clark, A. and Lappin, S. 2017. Grammaticality, acceptability, and probability: a probabilistic view of linguistic knowledge. *Cognitive Science.* **41**(5), pp.1202-1241.

Laufer, B. 2011. Verb-noun collocations in second language writing: a corpus analysis of learners' English. *Langauge Learning.* **61**(2), pp.647-672.

Laufer, B. and Girsai, N. 2008. Form-focused instruction in second language vocabulary learning: a case for contrastive analysis and translation. *Applied Linguistics.* **29**(4), pp.694-716.

Laufer, B. and Waldman, T. 2011. Verb-noun collocations in second language writing: a corpus analysis of learners' English. *Language Learning.* **61**(2), pp.647-672.

Lee, D.Y. 2001. Genres, registers, text types, domains and styles: clarifying the concepts and nevigating a path through the BNC jungle. *Language Learning & Technology.* **5**(3), pp.37-72.

Lee, S. 2006. A Corpus-based analysis of Korean EFL learner's use of amplifier collocations. *English Teaching.* **61**(1), pp.3-17.

Leech, G. 1981. *Semantics: the study of meaning.* Second Edition. Harmondsworth: Penguin Books.

Leech, G. 1992a. 100 million words of English: the British national corpus (BNC). *Language Research.* **28**(1), pp.1-13.

Leech, G. 1992b. Corpora and theories of linguistic performance. In: Svartvik, J. ed. *Directions in corpus linguistics.* Berlin: Mouton de Gruyter., pp.105-122.

Leech, G. 1992c. Corpora and theories of linguistic performance. Berlin: Mouton de Gruyter., pp.105-122.

Leech, G. 2002. The importance of reference corpora. *UZEI.* Hizkuntza-corpusak. Oraina eta geroa. , pp.1-11.

Leech, G., Rayson, P. and Wilson, A. 2001. Word frequencies in written and spoken English: based on the British national corpus. London: Longman.

Lennon, P. 1996. Getting 'easy' verbs wrong at the advanced level. *IRAL-International Review of Applied Linguistics in Language Teaching.* **34**(1), pp.24-36.

Lewis, M. and Conzett, J. 2000. *Teaching collocation: further developments in the lexical approach.* Hove: Language Teaching Publications.

Liu, C.P. 1999. An analysis of collocation errors in EFL Writings. In: *The proceeding of the English international symposium on English teaching.*: Crane Taipei, pp.483-494.

Liu, D. 2011. The most frequently used English phrasal verbs in American and British English: a multicorpus examination. *Tesol Quarterly.* **45**(4), pp.661-688.

Louw, B. and Milojkovic, M. 2016. *Corpus stylistics as contextual prosodic theory and subtext.* Amsterdam / Philadelphia: John Benjamins Publishing Company.

Ludbrook, J. 2013. Should we use one-sided or two-sided P values in tests of significance? *Clin Exp Pharmacol Physiol.* **40**(6), pp.357-361.

Mahmoud, A. 2005. Collocation errors made by Arab learners of English. *Asian EFL Journal.* **5**(2), pp.117–126.

Mammeri, S. 2015. A Morphosyntactic study of EFL students' written compositions: a corpus based analysis. *Arab World English Journal.* pp.112-126.

Manning, C.D., Manning, C.D. and Schütze, H. 1999. *Foundations of statistical natural language processing.* MIT press.

Martin, W. 2008. A unified approach to semantic frames and collocational patterns. In: Granger, S. and Meunier., F. eds. *Phraseology: an interdisciplinary perspective.* Amesterdam/Philadelphia: John Benjamins Publishing Company.

Martyńska, M. 2004. Do English language learners know collocations? *Investigationes Linguisticae.* **11**, pp.1-12.

The concise Oxford dictionary of linguistics. Third Edition ed. 2014. s.v.

McConnell-Ginet, S. 1982. Adverbs and logical form: a linguistically realistic theory. *Language.* **58**(1), pp.144-184.

McEnery, T. and Hardie, A. 2011. *Corpus linguistics: method, theory and practice.* Cambridge: Cambridge University Press.

McEnery, T. and Hardie, A. 2012. *Corpus linguistics: method, theory and practice.* [Online]. [Accessed 16 January 2020]. Available from: http://corpora.lancs.ac.uk/clmtp/2-stat.php

McEnery, T. and Wilson, A. 1996. *Corpus linguistics.* Edinburgh.: Edinburgh University Press.

McEnery, T. and Wilson, A. 2001. *Corpus linguistics. an introduction.* Edinburgh: Edinburgh University Press.

McEnery, T., Xiao, R. and Tono, Y. 2006. Corpus-based language studies: an advanced resource book. Routledge.

Mel'čuk, I. 1998. Collocations and lexical functions. In: Cowie, A. P. ed. *Phraseology. Theory, analysis, and applications.* Oxford: Clarendon Press., pp.23-53.

Men, H. 2015. Vocabulary increase and collocation learning: a corpus-based cross-sectional study of Chinese EFL learners. Doctor of Philosophy. thesis, Birmingham City University.

Meyer, C.F. 2002. *English corpus linguistics: An introduction.* Cambridge: Cambridge University Press.

Mheel, A.-S., Abdo, I.B. and Al-Amer, K. 2018. Difficulties in translating compound nouns from English into Arabic case study: Jordan University/Aqaba Branch. *Studies in Literature and Language.* **17**(2), pp.20-24.

Millon, C. 2011. Quenched light, or seeing through a glass darkly – a collocation-based View of Larkin's atheism and depression. *Belgrade English Language and Literature Studies.* **3**(1), pp.127-144.

Mishra, P., Pandey, C.M., Singh, U., Gupta, A., Sahu, C. and Keshri, A. 2019. Descriptive statistics and normality tests for statistical data. *Annals Of Cardiac Anaesthesia.* **22**(1), p67.

Moehkardi, R.R.D. 2002. Grammatical and lexical English collocations: Some possible problems to Indonesian learners of English. *HUMANIORA.* **14**(1), pp.53-62.

Mohamed, G.E.-d.H. and Eltayeb, M. 2016. Difficulties and problems that face Sudanese EFL learners students in learning English collocations. *Journal of Humanities Sciences.* **17**(2), pp.173-190.

Moisl, H. 2015. *Cluster analysis for corpus linguistics.* Berlin-New York: Walter de Gruyter GmbH & Co KG.

Mughazy, M. 2016. *The Georgetown guide to Arabic-English translation.* Washington, DC.: Georgetown University Press.

Mumin, M.A. 2022. International postgraduates' perception on academic writing: a lesson on criticality. *World Journal of English Language.* **12**(4), pp.1-46.

Mustafa, B.A. 2010. Collocation in English and Arabic: a linguistic and cultural analysis. *Journal of the College of Basic Education.* **15**(65), pp.29-43.

Nachar, N. 2008. The Mann-Whitney U: a test for assessing whether two independent samples come from the same distribution. *Tutorials in Quantitative Methods for Psychology.* **4**(1), pp.13-20.

Namvar, F. 2012. The relationship between language proficiency and use of collocation by Iranian EFL students. *The Southeast Asian Journal of English Language Studies.* **18**(3), pp.41-52.

Nasr, S.H. 2015. The Study Quran. A new Translation and Commentary. New York: HarperOne.

Nesselhauf, N. 2003. The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics.* **24**(2), pp.223-242.

Nesselhauf, N. 2005. *Collocations in a learner corpus.* John Benjamins Publishing.

Ngangbam, H. 2016. An analysis of syntactic errors committed by students of English language class in the written composition of Mutah University: A case study. *European Journal of English Language, Linguistics and Literature.* **3**(1), pp.1-13.

Nofal, K.H. 2012. Collocations in English and Arabic: a comparative study. *English Language and Literature Studies.* **2**(3), pp.75-93.

Nuruzzaman, M., Islam, A.S. and Shuchi, I.J. 2018. An analysis of errors committed by Saudi Non-English major students in the English Paragraph Writing: a study of comparisons. *Advances in Language and Literary Studies.* **9**(1), pp.31-39.

O'Donnell, M.B. 2008. KWICgrouper–designing a tool for corpus-driven concordance analysis. *International Journal of English Studies.* **8**(1), pp.107-122.

Oakes, M. 2019. *Statistics for corpus linguistics.* Edinburgh University Press.

Oakes, M.P. 1998. *statistics for corpus linguistics.* Edinburgh: Edinburgh University Press.

Obeidat, A. and Sepora, T. 2019. Collocation translation errors from Arabic into English: a case study of naguib mahfouz's novel "Awlad Haratina".

International Journal of Humanities, Philosophy and Language. **2**(7), pp.129-138.

Obukadeta, P. 2019. Collocations in a learner English corpus: analysis of Yoruba-speaking Nigerian English learners' use of collocations. Doctor of Philosophy. thesis, Kingston University.

Odlin, T. 1989. *Language transfer.* Cambridge: Cambridge University Press.

Orenha-Ottaiano, A. 2012. English collocations extracted from a corpus of university learners and its contribution to a language teaching pedagogy. *Acta Scientiarum. Language and Culture.* **34**(2), pp.241-251.

Ostertagova, E. and Ostertag, O. 2013. Methodology and application of one-way ANOVA. *American Journal of Mechanical Engineering.* **1**(7), pp.256-261.

Ostertagova, E., Ostertag, O. and Kováč, J. 2014. Methodology and application of the Kruskal-Wallis test. In: *Applied Mechanics and Materials*: Trans Tech Publ, pp.115-120.

Palmer, H.E. 1933. Second interim report on English collocations. In: The Tenth Annual Conference of English Teachers under the Auspices of the Institute for Research in English Teaching., Tokyo, Japan.

Paquot, M. 2019. The phraseological dimension in interlanguage complexity research. *Second Language Research.* **35**(1), pp.121-145.

Pawley, A. and Syder, F.H. 1983. Natural selection in syntax: notes on adaptive variation and change in vernacular and literary grammar. *Journal of Pragmatics.* **7**(5), pp.551-579.

Pecina, P. 2005. An extensive empirical study of collocation extraction methods. In: *Proceedings of the ACL Student Research Workshop*, pp.13-18.

Peters, E. 2016. The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research.* **20**(1), pp.113-138.

Pinto, P.T., Rees, G.P. and Frankenberg-Garcia, A. 2021. Identifying collocation issues in English L2 research article writing. In: Charles, M. and Frankenberg-Garcia, A. eds. *Corpora in ESP/EAP Writing Instruction: Preparation, Exploitation, Analysis.* London: Routledge, pp.147-170.

Pirhayati, M. 2021. A parallel corpus-based study of collocations from English to Persian: criticism, and resolution. *Journal of Applied Linguistics and Language Research.* **8**(3), pp.109-126.

Potts, A., Bednarek, M. and Caple, H. 2015. How can computer-based methods help researchers to investigate news values in large datasets? A corpus

linguistic study of the construction of newsworthiness in the reporting on Hurricane Katrina. *Discourse & Communication.* **9**(2), pp.149-172.

Pulido, M.F. and Dussias, P.E. 2020. Desirable difficulties while learning collocations in a second language: Conditions that induce L1 interference improve learning. *Bilingualism: Language and Cognition.* **23**(3), pp.652-667.

Pym, A. 1992. Translation error analysis and the interface with language teaching. In: Dollerup, C. and Loddegaard, A. eds. *The Teaching of Translation.* Amsterdam: John Benjamins, pp.279-288.

Qasem, F.A.A. 2020. Exploring spelling errors with relation to the phonological syllable structure in the writings of Saudi ESL learners. *International Journal of English Linguistics.* **10**(6), pp.152-161.

Quirk, R. and Greenbaum, S. 1973. *A concise grammar of contemporary english.* Harcourt School.

Rabeh, F. 2010. *Problems in translating collocations - the case of master I students of applied languages.* Master degree in Applied Language Studies thesis, Mentouri University-Constantine.

Radford, A. 2009. *An introduction to English sentence structure.* Cambridge: Cambridge University Press.

Rainer, F. 2003. Studying restrictions on patterns of word-formation by means of the Internet. *Italian Journal of Linguistics.* **15**(1), pp.131-139.

Rajab, A.S., Darus, S. and Aladdin, A. 2016. An investigation of semantic interlingual errors in the writing of Libyan English as foreign language learners. *Arab World English Journal.* **7**(4), pp.277- 296.

Randall, M. and Groom, N. 2009. The BUiD Arab learner corpus: a resource for studying the acquisition of L2 English spelling. In: *Proceedings of the Corpus Linguistics Conference (CL), Liverpool, UK*.

Rasinger, S.M. 2010. Quantitative methods: concepts, frameworks and issues. *Research Methods in Linguistics.* pp.49-67.

Richards, J.C. 2015. A non-contrastive approach to error analysis. *Error Analysis.* Routledge, pp.172-188.

Ritter, R.M. 2003. The Oxford style manual:[the essential handbook for all writers and editors]. Oxford University Press.

Sabah, S.S. 2015. Negative transfer: Arabic language interference to learning English. *Arab World English Journal (AWEJ).* **4**(Special Issue on Translation), pp.269-288.

Sabah, W. 2022. Investigating difficulties encounter EFL Iraqi secondary school learners in verb-noun collocations. *Midad Al-Adab Refereed Journal.* **1**(27), pp.665-696.

Sakaria, S. and Priyana, J. 2018. Code-switching: a pedagogical strategy in bilingual classrooms. *American Journal of Educational Research.* **6**(3), pp.175-180.

Saldanha, G. 2009. Principles of corpus linguistics and their application to translation studies research. *Tradumàtica: traducció i tecnologies de la informació i la comunicació.* (7), pp.1-7.

Saleh, M.K. and Fatah, A.H. 2022. A construal-based approach to Arabic Tenses. *The Scientific Journal of Cihan University–Sulaimaniya.* **6**(2), pp.60-77.

Sarfraz, S. 2011. Error analysis of the written English essays of Pakistani undergraduate students: a case study. *Asian Transactions on Basic & Applied Sciences.* **1**(3), pp.29-51.

Sari, B.N. and Gulö, I. 2019. Observing grammatical collocation in students' writings. *Teknosastik.* **17**(2), pp.25-31.

Sawalmeh, M.H.M. 2013. Error analysis of written English essays: the case of students of the preparatory year program in Saudi Arabia. *English for Specific Purposes World.* **14**(40), pp.1-17.

Schmitt, N. 2012. Formulaic language and collocation. *The Encyclopedia of Applied Linguistics.* pp.1-10.

Schulz, E. 2004. *A Student grammar of modern standard Arabic.* Cambridge, United Kingdom.: Cambridge University Press.

A dictionary of sociology. 3rd ed. 2014. s.v.

Selinker, L. 1983. Language transfer. In: Gass, S.M. and Selinker, L. eds. *Language learning. issues in second language research.* Newbury House Publishers.

Seretan, V. 2018. Bridging collocational and syntactic analysis. *Lexical Collocation AnalysisAdvances and Applications.* Switzerland: Springer, pp.23-38.

Shakir, H. and Yaseen, B.H. 2015. Acquisition of English prepositions among Iraqi secondary school students in Kuala Lumpur-Malaysia. *Advances in Language and Literary Studies.* **6**(5), pp.157-162.

Shammas, N.A. 2013. Collocation in English: comprehension and use by ma students at Arab Universities. *International Journal of Humanities and Social Science.* **3**(9), pp.107-122.

Sharoff, S. 2006. Open-source corpora: using the net to fish for linguistic data. *International Journal of Corpus Linguistics.* **11**(4), pp.435-462.

A frequency dictionary of Russian: Core vocabulary for learners. 2013. s.v.

Sharoff, S. 2020. Know thy corpus! Robust methods for digital curation of Web corpora. *Proc LREC.*

Sharoff, S. 2021. Genre annotation for the web: text-external and text-internal perspectives. *Register studies.* **3**(1), pp.1-32.

Shehata, A. 2008. L1 Influence on the reception and production of collocations by advanced ESL/EFL Arabic learners of English. Master of Arts thesis, Ohio University.

Shin, D. and Nation, P. 2008. Beyond single words: the most frequent collocations in spoken English. *ELT Journal.* **62**(4), pp.339-348.

Shooshtari, Z.G. and Karami, N. 2013. Lexical collocation instruction and its impact on Iranian non-academic EFL learners' speaking ability. *Journal of Language Teaching and Research.* **4**(4), pp.767-776.

Siddig, O.A., Siddig, B.E. and Elmahi, A. 2022. contrastive analysis of most relevant features of English and Arabic languages. *Journal of Positive School Psychology.* **6**(7), pp.4198-4210.

Sinclair, J. 1991. *Corpus, concordance, collocation.* Oxford University Press.

Sinclair, J. 2004. Trust the text: language, corpus and discourse. Routledge.

Siyanova, A. and Schmitt, N. 2008. L2 learner production and processing of collocation: a multi-study perspective. *The Canadian Modern Language Review.* **64**(3), pp.429-458.

Sorace, A. and Keller, F. 2005. Gradience in linguistic data. *Lingua.* **115**(11), pp.1497-1524.

Skitina, N., Poliakova, N. and Shabanova, V. 2017. Corpus-based approach to teaching phraseological units for foreign learners. In: *EDULEARN17 Proceedings*: IATED, pp.9014-9019.

Stefanowitsch, A. and Gries, S.T. 2003. Collostructions: investigating the interaction of words and constructions. *International Journal of Corpus Linguistics.* **8**(2), pp.209-243.

Storjohann, P. 2003. The lexicographic use of corpora and computational tools for disambiguation. In: *Proceedings of the Corpus Linguistics 2003 conference. UCREL technical paper number 16, Lancaster*. University of Lancaster, pp.754-762.

Suleiman, A.A. 2022. F1 impact on reception and production of delexical verbs collocations by Arabic EFL learners. *International Journal of Science and Research (IJSR).* **11**(5).

Tajareh, M.J., Khodareza, M. and Tonekabon, I. 2015. The effect of error analysis on Iranian intermediate EFL learners of L2 adjective knowledge. *Journal of Teaching English Language Studies.* **3**(4), pp.7-20.

Takač, V.P. and Lukač, M. 2013. How word choice matters: an analysis of adjective-noun collocations in a corpus of learner essays. *Jezikoslovlje.* **14**(2-3), pp.385-402.

Talakoob, F. and Koosha, M. 2017. Productive and receptive collocational knowledge of Iranian EFL learners at different proficiency levels. *International Journal of Applied Linguistics & English Literature.* **6**(7), pp.11-16.

Tang, B. 2004. Collocation errors in the intermediate EFL learners' interlanguage. *Sino-US English Teaching.* **1**(12), pp.39-47.

Taylor, C. 2008. What is corpus linguistics? What the data says. *ICAME journal.* **32**, pp.179-200.

Taylor, D. 1991. Compound word stress. *ELT Journal.* **45**(1), pp.67-73.

Thyab, R.A. 2016. Mother-tongue interference in the acquisition of English articles by L1 Arabic students. *Journal of Education and Practice.* **7**(3), pp.1-4.

Thanopoulos, A., Fakotakis, N. and Kokkinakis, G. 2002. Comparative Evaluation of Collocation Extraction Metrics. In: *LREC*, pp.620-625.

Trafimow, D. and MacDonald, J.A. 2017. Performing inferential statistics prior to data collection. *Educational and Psychological Measurement.* **77**(2), pp.204-219.

Trask, R.L. 2003. Mind the gaffe: the Penguin guide to common errors in English. Penguin UK.

Trinant, K. and Kijpoonphol, W. 2021. Lexical collocations in a sample corpus of tourism research articles (SCTRA). *NKRAFA Journal of Humanities and Social Sciences.* **9**, pp.94-108.

Trudgill, P. and Hannah, J. 2013. International English: a guide to the varieties of standard English. 6th ed. Routledge.

Uchihara, T., Eguchi, M., Clenton, J., Kyle, K. and Saito, K. 2022. To what extent is collocation knowledge associated with oral proficiency? A corpus-based approach to word association. *Language and Speech.* **65**(2), pp.311-336.

Van de Cruys, T. 2011. Two Multivariate Generalizations of Pointwise Mutual Information. In: *Proceedings of the Workshop on Distributional Semantics and Compositionality.*, pp.16-20.

Vergura, S., Acciani, G., Amoruso, V., Patrono, G.E. and Vacca, F. 2008. Descriptive and inferential statistics for supervising and monitoring the operation of PV plants. *IEEE Transactions on Industrial Electronics.* **56**(11), pp.4456-4464.

Verma, J. and Abdel-Salam, A.-S.G. 2019. *Testing statistical assumptions in research.* John Wiley & Sons.

Vilkaitė, L. 2017. Incidental acquisition of collocations in L2: effects of adjacency and prior vocabulary knowledge. *ITL-International Journal of Applied Linguistics.* **168**(2), pp.248-277.

Wang, T. and Good, R.L. 2007. The repetition of collocations in EFL Textbooks: a corpus study. *Online Submission.*

Wang, Y. 2016. The idiom principle and L1 influence: a contrastive learner-corpus study of delexical verb + noun collocations. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Weinreich, U. 1953. Languages in contact: findings and problems. Mouton: The Hague.

Willems, K. 2012. Intuition, introspection and observation in linguistic inquiry. *Language Sciences.* **34**(6), pp.665-681.

Williams, B. 2002. Collocation with advanced levels. [Online]. Available from: www.Teachingenglish.org.uk/think/vocabulary/collocation1.html

Williams, G. and Millon, C. 2011. Palmer, Firth and Internet: Drawing together collocational threads. In: *Corpus Linguistics 2011, Birmingham, United Kingdom.,* pp.1-34.

Wolff, D. 2009. Content and language integrated learning. In: Knapp, K. and Seidlhofer, B. eds. *Handbook of foreign language communication and learning.* De Gruyter Mouton, pp.545-572.

Wood, D. 2019. Classifying and identifying formulaic language. *The Routledge Handbook of Vocabulary Studies.* Routledge, pp.30-45.

Wray, A. 2004. Here's one I prepared earlier. *Formulaic Sequences: Acquisition, Processing, and Use.* **9**, pp.249-268.

Youcef, B. 2022. A Corpus-based study of adjective-noun collocation in descriptive essays by Algerian students of EFL. *Rufuf Journal.* **10**(1), pp.926-935.

Yule, G. 2022. *The study of language.* Cambridge, United Kingdom.: Cambridge University Press.

Yumanee, C. and Phoocharoensil, S. 2013. Analysis of collocational errors of Thai EFL students. *LEARN Journal: Language Education and Acquisition Research Network.* **6**(1), pp.88-98.

Zernik, U. 1990. Lexical acquisition: where is the semantics?. *Machine Translation.* **5**, pp.155-174.

Zhang, X. 1993. English collocations and their effect on the writing of native and non-native college freshmen. Doctor of Philosophy thesis, Indiana University of Pennsylvania.

Zhou, X. 2016. A Corpus-based study on high frequency verb collocations in the case of "have". In: *International Forum of Teaching and Sudies*: American Scholars Press, Inc., pp.42-50.

Zughoul, M.R. 1991. Lexical choice: towards writing problematic word lists. *International Review of Applied Linguistics.* **29**(1), pp.45-60.

Appendix A

A.1 Frequency analysis tables for the comparison of collocations in British and American English

Table A-1: The noun collocates for the verb lemma *MAKE* in BE06 and NBES (BAWE corpus)

Colocations_BE06	NF	Collocations_NBES	NF	Collocations_Arab EFLLs	NF
MAKE-- DECISION	24	MAKE-- DECISION	39.85	MAKE -- LIFE	49.82
MAKE-- SENSE	23	MAKE-- SENSE	30.46	MAKE – DECISION	18.12
MAKE-- PROGRESS	14	MAKE-- PROFIT	17.08	MAKE – SENSE	12.46
MAKE—DIFFERENCE	18	MAKE-- MISTAKE	10.31	MAKE – DIFFERENCE	7.93
MAKE-STATEMENT	10	MAKE-- USE	34.00	MAKE -- MISTAKE	6.79
MAKE—WAY	22	MAKE—DISTINCTION	8.00	MAKE –LIVING	5.66
MAKE-- MISTAKE	9	MAKE—JUDGMENT	7.08	MAKE – SMOKER	6.79
MAKE—EFFORT	8	MAKE—EFFORT	7.69	MAKE – CHOICE	5.66
MAKE-- MONEY	11	MAKE--REFERENCE	7.69	MAKE – PERSON	9.06
MAKE-- EXCUSE	6	MAKE --COMPARISON	7.54	MAKE – PLAYER	4.53
MAKE—RECOMMENDATION	5	MAKE—CHOICE	6.62	MAKE – ACCIDENT	4.53
MAKE--CONTRIBUTION	6	MAKE-- PROCESS	10.77	MAKE – WORK	7.93
MAKE-- USE	17	MAKE--CONTRIBUTION	4.62	MAKE – FRIEND	10.19
MAKE--APPEARANCE	5	MAKE—INTERFERENCE	3.54	MAKE – EVERYTHING	4.53
MAKE—POINT	10	MAKE—ASSUMPTION	5.38	MAKE – STUDENT	9.06
MAKE—BREAKFAST	4	MAKE—CLAIM	6.92	MAKE – BODY	6.79
MAKE—CLOTHING	4	MAKE-STATEMENT	5.23	MAKE – MONEY	5.66
MAKE—LIFE	11	MAKE-- MONEY	4.92	MAKE – COMMUNICATION	4.53
MAKE—CHOICE	6	MAKE—DIFFERENCE	6.77	MAKE – CHANGE	5.66
MAKE—ARRANGEMENT	4	MAKE—PREDICTION	3.08	MAKE—SHOPPING	4.53

Table A-2 The noun collocates for the verb lemma *HAVE* in BE06 and NBES (BAWE corpus)

Colocations_BE06	NF	Collocations_NBES	NF	Collocations_Arab EFLLs	NF
HAVE--IDEA	44	HAVE-- EFFECT	68.15	HAVE—FUN	70.20
HAVE -- ACCESS	27	HAVE—ACCESS	26.46	HAVE—LUNCH	27.18
HAVE—CONTACT	23	HAVE--IMPACT	27.69	HAVE—RIGHT	36.23
HAVE—EFFECT	27	HAVE—IMPLICATION	19.38	HAVE-- ABILITY	15.85
HAVE—CHANCE	19	HAVE—DIFFICULTY	17.54	HAVE—EFFECT	27.18
HAVE—IMPACT	16	HAVE—POWER	32.46	HAVE—ADVANTAGE	19.25
HAVE—CAPACITY	11	HAVE—ADVANTAGE	18.77	HAVE—POWER	15.85
HAVE—MISGIVING	4	HAVE—ABILITY	20.15	HAVE—DINNER	13.59
HAVE—PLENTY	8	HAVE—POTENTIAL	18.92	HAVE--CHANCE	13.59
HAVE—RIGHT	27	HAVE—KNOWLEDGE	19.08	HAVE—REST	10.19
HAVE—RESPONSIBILITY	9	HAVE—OPPORTUNITY	14.62	HAVE—ACCIDENT	13.59
HAVE—FUN	8	HAVE—RIGHT	22.92	HAVE—OPPORTUNITY	9.06
HAVE—BREAKFAST	6	HAVE—CHANCE	8.31	HAVE—DISADVANTAGE	12.46
HAVE—IMPLICATION	7	HAVE—CONSEQUENCE	9.69	HAVE—CONVERSATION	6.79
HAVE—TROUBLE	8	HAVE—INFLUENCE	14.62	HAVE—SIDE	12.46
HAVE—DIFFICULTY	8	HAVE—CAPACITY	8.31	HAVE—BENEFIT	13.59
HAVE—POWER	15	HAVE-- INCENTIVE	5.23	HAVE—PROBLEM	20.38
HAVE—DOUBT	9	HAVE—RESPONSIBILITY	8.15	HAVE—FREEDOM	7.93
HAVE--BATH	6	HAVE—TENDENCY	5.23	HAVE—CANCER	4.53
HAVE—CONFIDENCE	7	HAVE—CHOICE	9.08	HAVE—TIME	33.97

Table A-3 The noun collocates for the verb lemma *DO* in BE06 and NBES (BAWE corpus)

Colocations_BE06	NF	Collocations_NBES	NF	Collocations_Arab EFLLs	NF
DO—EXPLODE	8	DO—JOB	22	DO—SURGERY	27.18
DO—HAVE	7	DO—TRICK	6	DO—EXERCISE	9.06
DO—JOB	25	DO—HOMEWORK	4	DO—MISTAKE	5.66
DO—BUSINESS	34	DO—ACT	9	DO—ACTIVITY	10.19
DO—MATHEMATICS	8	DO—WORK	15	DO—ACCIDENT	7.93
DO—WORK	49	DO—HOME	7	DO—WORK	12.46
DO—DEED	5	DO—BIT	4	DO—TASK	5.66
DO—RESEARCH	26	DO—RESEARCH	4	DO—HOMEWORK	4.53
DO—EXERCISE	12	DO—PARENT	22	DO—PROJECT	4.53
DO—HARM	8			DO—RESEARCH	6.79
DO—HOUSEWORK	4			DO—BUSINESS	4.53
DO—BIND	9			DO—SPORT	5.66
DO—FACT	23			DO—KIND	4.53
DO—EXPERIMENT	15			DO—HARM	4.53
DO—JUSTICE	9			DO—JOB	4.53
DO—TASK	12			DO—SHOPPING	4.53
DO—CALCULATION	8				
DO—SHOPPING	4				
DO—VIOLENCE	7				
DO—SENSATION	4				

Table A-4 The noun collocates for the verb lemma *GET* in BE06 and NBES (BAWE corpus)

Colocations_BE06	NF	Collocations_NBES	NF	Collocations_Arab EFLLs	NF
GET—MEAL	16	GET—JOB	4.46	GET--INFORMATION	10.19
GET—MUSCLE	11	GET—FORMULA	4.46	GET—MARK	5.66
GET—WORKOUT	8	GET—INFORMATION	3.23	GET—JOB	7.93
GET—GRIPPE	9	GET—MONEY	1.85	GET—CITIZENSHIP	4.53
GET—BIT	9	GET—CHANCE	1.38	GET—GRADE	4.53
GET—TROUBLE	6	GET—GRIPPE	0.77	GET—CHANCE	4.53
GET—CHANCE	7	GET—ACCESS	1.54	GET—BENEFIT	5.66
GET—JOB	8	GET—PERMISSION	0.77	GET—ENERGY	5.66
GET—GUN	5	GET—NAME	1.38	GET--HOME	5.66
GET—TOUCH	6	GET—INSIGHT	0.92		
GET—IMPRESSION	4	GET—MARK	1.08		
GET—HELP	7	GET—ANSWER	1.23		
GET—ADVICE	4	GET—ITEM	1.08		
GET—HOME	7	GET—IDEA	1.69		
GET—SENSE	5	GET--FEEDBACK	0.77		
GET—CAR	5	GET—PATENT	0.62		
GET—MONEY	5	GET—DETAIL	1.08		
GET—IDEA	5	GET—RETURN	1.23		
GET—WORD	5	GET—SENSE	1.69		
GET—HAND	8	GET—PRODUCT	0.92		

Table A-5 The noun collocates for the verb lemma ***TAKE in*** **BE06 and NBES *(BAWE corpus)***

Colocations_BE06	NF	Collocations_NBES	NF	Collocations_Arab EFLLs	NF
TAKE --PLACE	103	TAKE -- PLACE	124.15	TAKE –CARE	48.69
TAKE --ACCOUNT	40	TAKE – ACCOUNT	101.08	TAKE – ACTION	24.91
TAKE -- PART	45	TAKE – CONSIDERATION	25.08	TAKE – PICTURE	21.51
TAKE --CARE	23	TAKE – ADVANTAGE	22.15	TAKE –GAP	12.46
TAKE -- STEP	22	TAKE --PART	28.62	TAKE – REST	12.46
TAKE -- ADVANTAGE	13	TAKE -- ACTION	14.46	TAKE –PHOTO	12.46
TAKE -- SIP	6	TAKE –CARE	12.00	TAKE –PLACE	22.65
TAKE -- ACTION	10	TAKE – RESPONSIBILITY	24.41	TAKE – ADVANTAGE	12.46
TAKE – OPPORTUNITY	9	TAKE – FORM	15.5	TAKE – TAXI	7.93
TAKE – TOLL	5	TAKE –TIME	14.92	TAKE – PHOTOGRAPH	5.66
TAKE – RESPONSIBILITY	7	TAKE –STEP	13.24	TAKE – YEAR	12.46
TAKE – NOTICE	8	TAKE – PRECEDENCE	12.36	TAKE – PHOTOES	4.53
TAKE – FORM	11	TAKE –MEDICATION	11.09	TAKE – RESPONSIBILITY	5.66
TAKE – BREATH	7	TAKE – RISK	9.81	TAKE – CONSIDERATION	4.53
TAKE – TIME	19	TAKE -- PRECAUTION	8.96	TAKE – JOB	6.79
TAKE -- SHOT	6	TAKE – DRUG	8.77	TAKE – NOTE	4.53
TAKE – LOOK	15	TAKE – MEASUREMENT	8.36	TAKE – PART	6.79
TAKE -- CHANCE	7	TAKE – READING	7.37	TAKE – COURSE	6.79
TAKE -- HINT	4	TAKE – NOTE	4.46	TAKE – HOUR	7.93
TAKE --RISK	8	TAKE – INITIATIVE	2.31	TAKE – BREAK	4.53

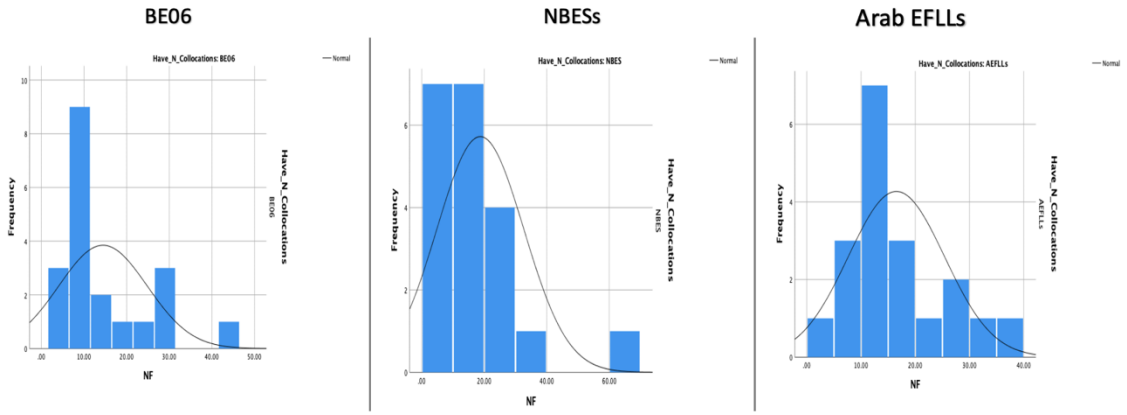


Figure A.1-1: Histograms showing the distribution of the noun collocates for the verb *MAKE* in BE06, NBES and Arab EFLLs.

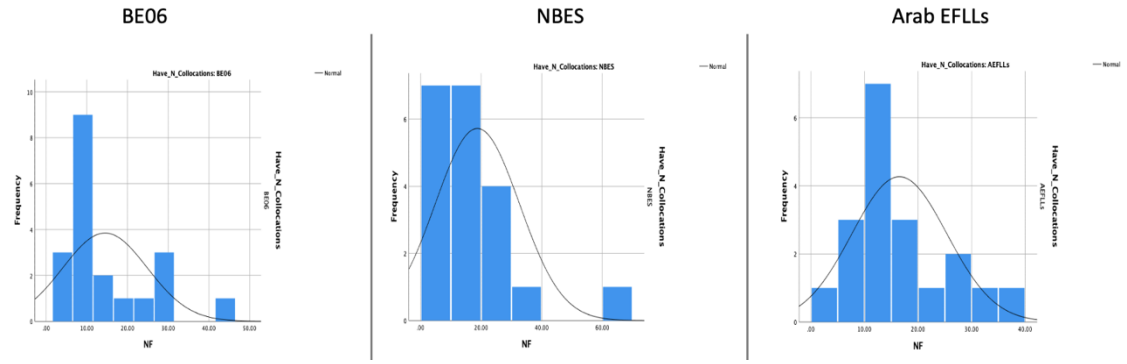


Figure A.1-2: Histograms showing the distribution of the noun collocates for the verb *HAVE* in BE06, NBES and Arab EFLLs.

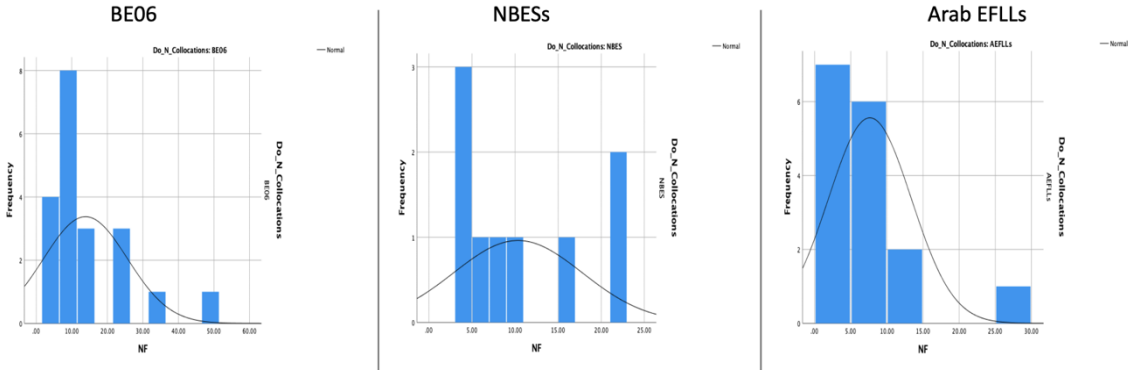


Figure A.1-3: Histograms showing the distribution of the noun collocates for the verb *DO* in BE06, NBES and Arab EFLLs.

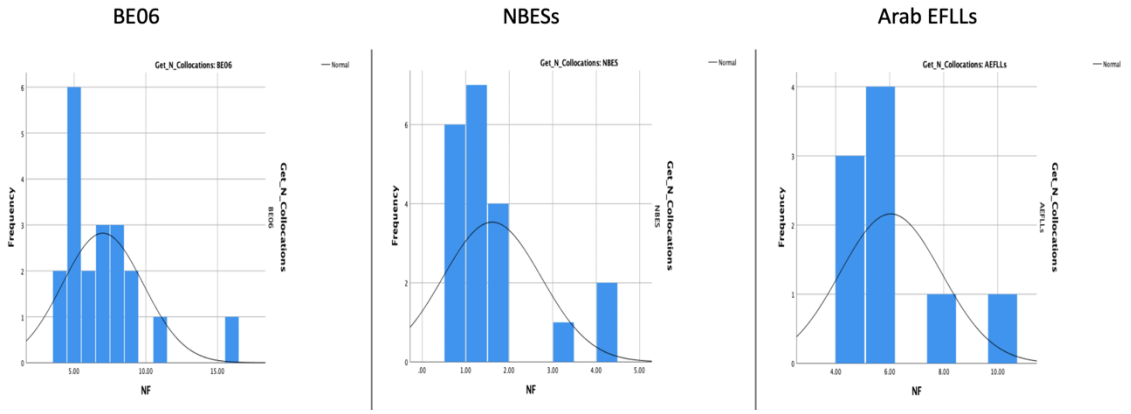


Figure A.1-4: Histograms showing the distribution of the noun collocates for the verb *GET* in BE06, NBES and Arab EFLLs.

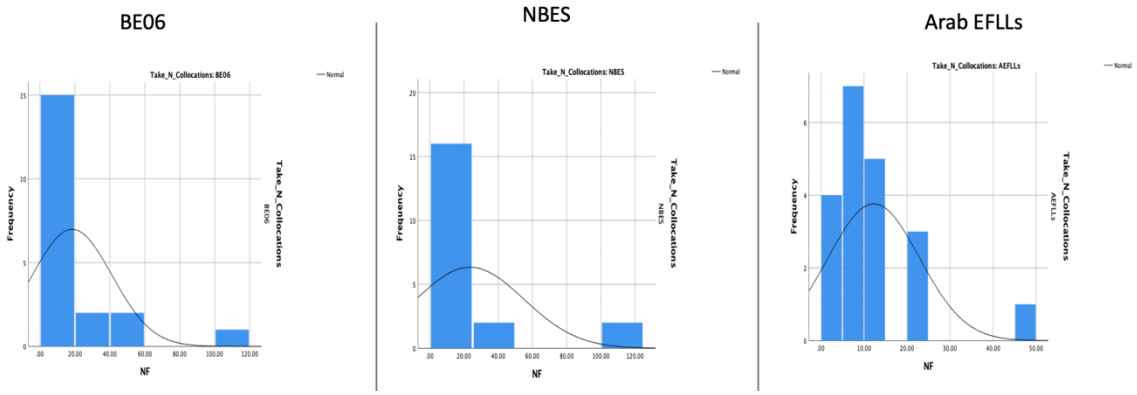


Figure A.1-5: Histograms showing the distribution of the noun collocates for the verb *TAKE* in BE06, NBES and Arab EFLLs.

Table A-6: Verb-Noun collocations in the BNC and I-EN

Corpora/ Verbs	BNC	Log-Likelihood (LL)	I-EN	Log-Likelihood (LL)	Commentary
DO + **NOUN**	1. *SOMETHING* 2. *ANYTHING* 3. *JOB* 4. *THING* 5. *NOTHING* 6. *JUSTICE* 7. *EVERYTHING* 8. *WORK* 9. *HARM* 10. *BUSINESS* 11. WORRY 12. SHOPPING 13. *TRICK* 14. *BIDDING* 15. *BATTLE* 16. WASHING 17. *HOMEWORK* 18. *DAMAGE* 19. *SORT* 20. *DUTY*	1. 2785.63 2. 2288.75 3. 992.07 4. 773.74 5. 606.34 6. 372.18 7. 323.91 8. 242.45 9. 239.78 10. 158.02 11. 138.74 12. 131.48 13. 128.78 14. 69.64 15. 62.82 16. 72.77 17. 113.13 18. 46.88 19. 36.98 20. 35.19	1. *SOMETHING* 2. *ANYTHING* 3. *THING* 4. *JOB* 5. *EVERYTHING* 6. CUSS 7. *NOTHING* 8. *WORK* 9. RESEARCH 10. *BUSINESS* 11. *TRICK* 12. *HARM* 13. *HOMEWORK* 14. *JUSTICE* 15. *DAMAGE* 16. *BIDDING* 17. STUFF 18. *SORT* 19. *DUTY* 20. *BATTLE*	1. 4922.93 2. 4641.93 3. 2402.28 4. 1079.75 5. 738.44 6. 591.78 7. 858.58 8. 333.47 9. 289.42 10. 333.86 11. 257.22 12. 332.38 13. 341.6 14. 178.16 15. 88.4 16. 115.78 17. 86.19 18. 47.09 19. 22.55 20. 23.03	- Seventeen out of the twenty nouns were found as the top twenty collocates in both corpora accounting for 80%. -The remaining collocates were found in both corpora but were in the top twenty for only one of the two corpora.
MAKE + **NOUN**	1. *SENSE* 2. *DECISION* 3. *DIFFERENCE* 4. *STATEMENT* 5. *USE* 6. *MISTAKE* 7. *PROFIT* 8. MONEY 9. *PROGRESS* 10. CONTACT 11. CONTRIBUTION 12. ARRANGEMENT 13. *EFFORT* 14. *AMENDS* 15. *CHOICE* 16. WAY 17. *RECOMMENDATION* 18. *JUDGMENT* 19. *PAYMENT* 20. *POINT*	1. 1835.46 2. 1352.49 3. 858.46 4. 783.76 5. 623.84 6. 487.22 7. 449.65 8. 424.25 9. 387.33 10. 373.8 11. 360.25 12. 358.73 13. 345.53 14. 343.47 15. 319.27 16. 296.8 17. 290.86 18. 281.17 19. 244.8 20. 231.64	*1. SENSE* *2. DECISION* *3.* MONEY *4. DIFFERENCE* *5. CHOICE* *6.* CHANGE *7. MISTAKE* *8.* LIVING *9.* THING *10. PROFIT* *11. PROGRESS* *12.* LIFE *13. EFFORT* *14. JUDGMENT* *15. STATEMENT* *16. POINT* *17. USE* *18. RECOMMENDATION* *19. PAYMENT* *20. AMENDS*	1. 6728.57 2. 4827.83 3. 3995.42 4. 2250.07 5. 1327.26 6. 1052.16 7. 1611.35 8. 892.35 9. 750.14 10. 691.56 11. 525.79 12. 494.15 13. 511.97 14. 465.17 15. 437.11 16. 429.9 17. 437.54 18. 297.56 19. 284.81 20. 246.53	- Fifteen out of the twenty nouns above were found as the top twenty collocates in both corpora accounting for 75% -The remaining collocates were found in both corpora but were in the top twenty for only one of the two corpora.
GET + **NOUN**	1. *JOB* 2. *TOUCH* 3. *MONEY* 4. *GRIPPE* 5. AWAY FROM 6. *CHANCE*	1. 579.42 2. 501.81 3. 414.78 4. 343.1 5. 323.78 6. 294.95 7. 283.49	*1. JOB* *2. INFORMATION* *3. MONEY* 4. ATTENTION 5. *CHANCE* 6. *PEOPLE*	1. 941.89 2. 573.3 3. 561.86 4. 499.42 5. 1041.92 6. 417.35 7. 441.88	- Seventeen out of the twenty nouns above were found as the top twenty collocates in both corpora accounting for 85%.

	7. HOME 8. *SOMETHING* 9. *ANYTHING* 10. *THING* 11. *MESSAGE* 12. *HAND* 13. *HOLD* 14. *TICKET* 15. ROUND 16. *INFORMATION* 17. *IDEA* 18. *PEOPLE* 19. *PERMISSION* 20. *TROUBLE*	8. 278.97 9. 199.38 10. 195.78 11. 191.11 12. 159.59 13. 123.15 14. 111.56 15. 111.56 16. 105.26 17. 94.7 18. 933.33 19. 90.99 20. 62.44	7. *TOUCH* 8. *IDEA* 9. *SOMETHING* 10. *HAND* 11. *THING* 12. *MESSAGE* 13. *PERMISSION* 14. *ANYTHING* 15. *GRIPPE* 16. *HOLD* 17. *TICKET* 18. *TROUBLE* 19. CREDIT 20. FEELING	8. 702.3 9. 349.6 10. 372.67 11. 312.16 12. 335.58 13. 372.76 14. 205.22 15. 199.04 16. 157.75 17. 169.44 18. 209.24 19. 172.53 20. 310.33	-The remaining collocates were found in both corpora but were in the top twenty for only one of the two corpora.
TAKE **+** **NOUN**	1. *PLACE* 2. *ACCOUNT* 3. *ADVANTAGE* 4. *PART* 5. *ACTION* 6. *CARE* 7. *STEP* 8. *RESPONSIBILITY* 9. *RISK* 10. *TIME* 11. *OPPORTUNITY* 12. *NOTICE* 13. FORM 14. CHARGE 15. *PHOTOGRAPH* 16. *CHANCE* 17. ADVICE 18. NOTE 19. DECISION 20. PR*ECAUTION*	1. 5618.74 2. 3306.26 3. 2491.16 4. 1966.52 5. 1545.21 6. 1533.61 7. 631.09 8. 594.75 9. 454.98 10. 417.09 11. 323.6 12. 320.24 13. 259.73 14. 217.35 15. 151.92 16. 259.2 17. 194.85 18. 240.1 19. 148.72 20. 162.94	*1. ADVANTAGE* *2. PLACE* *3. CARE* *4. ACTION* *5. ACCOUNT* *6. TIME* *7.* LOOK *8. PART* *9. RESPONSIBILITY* *10. STEP* *11.* PICTURE *12.* NOTE *13. RISK* *14. CHANCE* *15.* CONTROL *16.* CONSIDERATION *17. PRECAUTION* *18. PHOTO* *19. OPPORTUNITY* *20. NOTICE*	1. 5269.51 2. 5193.06 3. 4087.79 4. 2775.29 5. 2713.29 6. 2444.23 7. 2273.16 8. 1763.14 9. 1806.82 10. 1157.97 11. 1046.72 12. 733.11 13. 562.32 14. 503.83 15. 444.63 16. 488.09 17. 369.12 18. 307.78 19. 245.38 20. 247.39	- Fifteen out of the twenty verbs above were found as the top twenty collocates in both corpora accounting for 75%. -The remaining collocates were found in both corpora but were in the top twenty for only one of the two corpora.

Table A-7:Noun-Noun Collocations in the BNC and I-EN

Corpora/ Nouns	BNC	Log-Likelihood (LL)	I-EN	Log-Likelihood (LL)	Commentary
***TIME* + NOUN**	1. *CONSUMING* 2. *LIMIT* 3. WRITING 4. *EFFORT* 5. *MONEY* 6. *YEAR* 7. *ENERGY* 8. *PERIOD* 9. *INTERVAL* 10. *SPACE* 11. SCALE 12. *ZONE* 13. *SPAN* 14. *DAY* 15. *LAG* 16. DEPOSIT 17. RESOURCE 18. PLACE 19. DIAGNOSIS 20. *DELAY*	1. 517.76 2. 486.53 3. 250.35 4. 223.18 5. 222.99 6. 216.94 7. 203.03 8. 169.99 9. 164.49 10. 145.7 11. 144.15 12. 136.46 13. 99.81 14. 91.03 15. 75.84 16. 66.66 17. 64.28 18. 51.09 19. 31.88 20. 31.48	*1. PERIOD* *2.* POINT *3. CONSUMING* 4. FRAME 5. THREAD 6. *INTERVAL* *7. MONEY* *8. EFFORT* 9. *ENERGY* *10. ZONE* 11. CONSTRAINT *12. LIMIT* 13. COURSE 14. *YEAR* 15. *SPACE* *16. SPAN* 17. WINDOW 18. *DELAY* 19. WRITER *20. LAG*	1. 2867.4 2. 2161.4 3. 1576.94 4. 1437.82 5. 1127.94 6. 958.47 7. 778.11 8. 690.73 9. 573.11 10. 532.69 11. 424.53 12. 363.46 13. 349.28 14. 312.61 15. 275.07 16. 238.93 17. 172.31 18. 198.64 19. 165.44 20. 165.05	Fourteen out of the twenty nouns were in the top twenty of collocates in both corpora, accounting for 85%. The remaining collocates were found in both corpora but were in the top twenty for only one of the two corpora.
***WORK* + NOUN**	1. ART 2. EXPERIENCE 3. FORCE 4. ORGANIZATION 5. ETHIC 6. PRACTICE 7. PROGRESS 8. LEISURE 9. LOAD 10. PERMIT 11. PLACEMENT 12. ENVIRONMENT 13. SITUATION 14. GROUP 15. DEPARTMENT 16. COLLEAGUE 17. SCHEDULE 18. FICTION 19. PLACE 20. COMMITMENT	1. 413.96 2. 309.1 3. 245.09 4. 137.92 5. 106.23 6. 98.14 7. 82.86 8. 74.85 9. 66.89 10. 64.64 11. 58.59 12. 41.71 13. 40.26 14. 39.06 15. 35.17 16. 34.43 17. 34.06 18. 33.4 19. 21.68 20. 28.65	*1. ENVIRONMENT* *2. ETHIC* *3. PROGRESS* *4. FORCE* *5. ART* *6. EXPERIENCE* 7. VISA 8. HOME *9. SCHEDULE* *10. LOAD* 11. CAPACITY *12. COLLEAGUE* 13. PERFORMANCE 14. *FICTION* 15. DEDICATION *16. GROUP* *17. PLACE* *18. PERMIT* *19. SITUATION* 20. *ORGANIZATION*	1. 478.3 2. 415.36 3. 380.08 4. 313.58 5. 299.14 6. 191.64 7. 182.72 8. 65931 9. 132.76 10. 127.57 11. 84.24 12. 74.63 13. 68.51 14. 69.09 15. 68.47 16. 55.86 17. 53.17 18. 42.81 19. 30.03 20. 23.9	Fifteen out of the twenty nouns above were found as the top twenty collocates in both corpora accounting for 75%. The remaining collocates were found in both corpora but were in the top twenty for only one of the two corpora.
***LIFE* + NOUN**	1. EXPECTANCY 2. CYCLE 3. IMPRISONMENT 4. ASSURANCE 5. DEATH 6. SENTENCE 7. INSURANCE 8. STYLE 9. HISTORY 10. SPAN 11. EARTH 12. LIMB 13. EVENT 14. MISERY 15. STORY	1. 1520.57 2. 556.56 3. 543.18 4. 532.12 5. 424.2 6. 374.58 7. 295.61 8. 155.33 9. 107.37 10. 106 11. 80.16 12. 78.84 13. 77.54 14. 71.6 15. 61.77	*1. EXPECTANCY* *2. CYCLE* *3. INSURANCE* *4. SPAN* 5. *DEATH* 6. *EVENT* 7. POLITICS 8. LIBERTY *9. STYLE* *10. EXPERIENCE* *11. IMPRISONMENT* *12. SENTENCE* *13. EARTH* *14. HISTORY* *15. SUPPORT*	1. 4210.02 2. 1039.16 3. 818.32 4. 671.55 5. 580.12 6. 536.52 7. 423.74 8. 369.06 9. 323.18 10. 320.96 11. 285.86 12. 278.58 13. 227.61 14. 203.28 15. 184.37	Sixteen out of the twenty nouns above were found as the top twenty collocates in both corpora accounting for 80%. The remaining collocates were found in both corpora but were in the top twenty for only one of the two corpora.

	16. EXPERIENCE 17. SUPPORT 18. SAVING 19. WORK 20. CHANCE	16. 58.5 17. 47.66 18. 45.68 19. 40.67 20. 31.88	*16. SAVING* *17. LIMB* *18. STORY* 19. STRENGTH 20. GOAL	16. 112.61 17. 84.17 18. 66.65 19. 65.53 20. 45.13	
***WAY* + NOUN**	1. ROUND 2. LIFE 3. HOME 4. THINKING 5. PEOPLE 6. UPSTAIRS 7. DOWNSTAIRS 8. LIVING 9. SALVATION 10. SPEAKING 11. LONDON 12. UNDERSTANDING 13. ROME 14. ACCOUNTING 15. RECOVERY 16. SOCIETY 17. GOD	1. 1285.95 2. 1127.84 3. 550.31 4. 132.93 5. 68.96 6. 37.09 7. 25.67 8. 20.82 9. 14.59 10. 14.28 11. 10.35 12. 7.31 13. 6.64 14. 6.5 15. 5.93 16. 4.95 17. 4.92	1. *LIFE* 2. *HOME* 3. *THINKING* 4. *PEOPLE* 5. *LIVING* 6. ROUND 7. STREET 8. EVERYONE 9. *GOD* 10. *SALVATION* 11. *DOWNSTAIRS* 12. CONGRESS 13. *SOCIETY* 14. ANYONE 15. CIRCULATION 16. PROSPERITY 17. HEAVEN	1. 661.54 2. 347.12 3. 309.54 4. 221.47 5. 214.56 6. 183.86 7. 33.93 8. 31.59 9. 28.04 10. 14.5 11. 14.49 12. 13.66 13. 13.49 14. 27.85 15. 21.21 16. 17.71 17. 16.21	It was hard to extract examples for this combination as the noun *way* was used to describe how to get to certain places or destinations. Therefore, 9 out of the 17 listed were found in both corpora yet they were not from the top twenty.

Table A-8: Adjective-Noun collocations in the BNC and I-EN

Corpora/ Adjectives	BNC	Log-Likelihood (LL)	I-EN	Log-Likelihood (LL)	Commentary
NEW **+** **NOUN**	1. *TECHNOLOGY* 2. SETTLEMENT 3. *PRODUCT* 4. *SYSTEM* 5. CONSTITUTION 6. *GENERATION* 7. GOVERNMENT 8. *LAW* 9. *IDEA* 10. *JOB* 11. DEVELOPMENT 12. BUILDING 13. *CAR* 14. *ERA* 15. *HOME* 16. *VERSION* 17. SERIES 18. LEGISLATION 19. MEMBER 20. *MODEL*	1. 2099.22 2. 1486.64 3. 1204.77 4. 1167.38 5. 774.61 6. 771.4 7. 717.21 8. 691.82 9. 679.41 10. 649.17 11. 633.88 12. 597.51 13. 583.38 14. 481.06 15. 478.15 16. 446.9 17. 435.92 18. 435.19 19. 414.56 20. 394.69	*1. TECHNOLOGY* *2.* WAY *3.* BOOK *4.* WINDOW *5.* MEDIUM *6. GENERATION* *7.* FEATURE *8. PRODUCT* *9. IDEA* *10. LAW* *11. JOB* *12. VERSION* *13. SYSTEM* *14.* TOOL *15.* APPROACH *16.* PROJECT *17. MODEL* *18. CAR* *19. HOME* *20. ERA*	1. 3778.04 2. 1681.41 3. 1510.41 4. 1505.81 5. 1425.71 6. 1388.28 7. 1338.84 8. 1218.37 9. 1312.01 10. 1102.76 11. 1020.49 12. 1004.66 13. 899.7 14. 891.59 15. 836.52 16. 790.9 17. 744.76 18. 736.45 19. 586.31 20. 584.75	-Twelve out of the twenty nouns were found as the top twenty collocates in both corpora accounting for 60%. -The remaining collocates were found in both corpora but were in the top twenty for only one of the two corpora.
GOOD **+** **NOUN**	1. *IDEA* 2. *NEWS* 3. *REASON* 4. DEAL 5. *THING* 6. *JOB* 7. *EXAMPLE* 8. *PRACTICE* 9. *TIME* 10. *QUALITY* 11. *FORTUNE* 12. *FRIEND* 13. *FAITH* 14. *LUCK* 15. CONDITION 16. *CHANCE* 17. HUMOUR 18. *HEALTH* 19. SENSE 20. VALUE	1. 3438.71 2. 2179.4 3. 1700.71 4. 1192.43 5. 1184.87 6. 1046.51 7. 1041.21 8. 784.93 9. 759.8 10. 747.03 11. 701.15 12. 657.39 13. 651 14. 623.06 15. 573.28 16. 530.66 17. 425.78 18. 403.97 19. 339.49 20. 334.26	21. *IDEA* 22. *THING* 23. *NEWS* 24. *REASON* 25. *JOB* 26. *LUCK* 27. *FRIEND* 28. *EXAMPLE* 29. *FAITH* 30. DEAL 31. *TIME* 32. POINT 33. *QUALITY* 34. WAY 35. *CHANCE* 36. PLACE 37. INTENTION 38. *HEALTH* 39. *PRACTICE* 40. *FORTUNE*	1. 7522.74 2. 5740.59 3. 4045.81 4. 2658.8 5. 2358.49 6. 1826.96 7. 1372.29 8. 1229.2 9. 1191.75 10. 1064.95 11. 1064.64 12. 1044.62 13. 910.15 14. 896.81 15. 800.92 16. 723.91 17. 563.48 18. 530.06 19. 491.56 20. 485.15	-Fifteen out of the twenty nouns above were found as the top twenty collocates in both corpora accounting for 75%. -The remaining collocates were found in both corpora but were in the top twenty for only one of the two corpora.
FIRST **+** **NOUN**	1. *TIME* 2. *PLACE* 3. *HALF* 4. *YEAR* 5. *STEP* 6. *DAY* 7. *QUARTER* 8. *THING* 9. INSTANCE 10. *ROUND* 11. *MONTH* 12. SIGHT 13. MEETING	1. 12948.39 2. 2906.2 3. 2466.4 4. 2302.19 5. 1839.92 6. 1502.45 7. 1497.47 8. 1443.18 9. 1145.68 10. 1111.77 11. 1088.95 12. 1056.44 13. 895.52	*1. TIME* *2. PLACE* *3. STEP* *4. YEAR* *5. DAY* *6. THING* *7. WEEK* *8.* DATE *9. MONTH* *10.* DRAFT *11. HALF* *12. ROUND* *13.* GLANCE	1. 13646.43 2. 6386.13 3. 4759.46 4. 3613.51 5. 2362.38 6. 2002.58 7. 1665.03 8. 1581.5 9. 1515.38 10. 1177.69 11. 1090.51 12. 1072.61 13. 1021.64	-Fourteen out of the twenty nouns above were found as the top twenty collocates in both corpora accounting for 70%. -The remaining collocates were found in both corpora but were in the top twenty for only one of the two corpora.

	14. *WEEK* 15. *STAGE* 16. FLOOR 17. EDITION 18. *CLASS* 19. *PART* 20. APPEARANCE	14. 735.53 15. 702.85 16. 635.03 17. 575.96 18. 565.84 19. 519.65 20. 487.77	*14.* COMMENT *15.* LINE *16. QUARTER* *17. CLASS* *18.* PERSON *19. STAGE* *20. PART*	14. 842.58 15. 812.38 16. 787.57 17. 658.53 18. 647.59 19. 624.47 20. 619.83	
OTHER **+** **NOUN**	1. *HAND* 2. *WORD* 3. *PEOPLE* 4. *SIDE* 5. *THING* 6. *COUNTRY* 7. *MEMBER* 8. *WAY* 9. *AREA* 10. *PART* 11. *FACTOR* 12. *PERSON* 13. PARTY 14. END 15. *DAY* 16. *GROUP* 17. ASPECT 18. FORM 19. *SOURCE* 20. *TYPE*	1. 9904.01 2. 5562.1 3. 5272.99 4. 5032.15 5. 3545.09 6. 3185.21 7. 2194.53 8. 2042.36 9. 1330.34 10. 1273.91 11. 1205.14 12. 1196.25 13. 1162.07 14. 1155.45 15. 1021.1 16. 1018.69 17. 833.9 18. 829.76 19. 693.58 20. 619.1	*1. HAND* *2. WORD* *3. COUNTRY* *4. PEOPLE* *5. THING* *6. SIDE* *7. FACTOR* *8. WAY* *9. MEMBER* 10. *GROUP* 11. STUDY 12. NATION *13. AREA* *14. DAY* *15. TYPE* 16. FORM *17. PART* *18. SOURCE* 19. READER 20. *PERSON*	1. 16268.75 2. 11095.18 3. 7793.9 4. 7505.41 5. 5239.37 6. 4498.76 7. 2647.21 8. 2378.45 9. 1877.36 10. 1711.33 11. 1568.62 12. 1510.78 13. 1441.99 14. 1453.53 15. 1341.72 16. 1300.26 17. 1281.13 18. 1206.5 19. 1177.32 20. 1138.07	Sixteen out of the twenty adjectives above were found as the top twenty collocates in both corpora accounting for 80%. The remaining collocates were found in both corpora but were in the top twenty for only one of the two corpora.

Table A-9: Mann-Whitney U Test for Verb-Noun collocations in BE and AE

Mann-Whitney U Test for Verbs	Corpora	U	Z	P
DO + NOUN	BNC	94	-2.867	0.004
	I-EN	133	-1.812	0.072
MAKE + NOUN	BNC	40	-4.328	0.000
	I-EN	117	-2.245	0.024
GET + NOUN	BNC	43	-4.247	0.000
	I-EN	92	-2.921	0.003
TAKE + NOUN	BNC	84	-3.138	0.001
	I-EN	145	-1.488	0.142

Table A-10:Mann-Whitney U Test for Noun-Noun collocation in BE and AE

Mann-Whitney U Test for Nouns	Corpora	U	Z	P
TIME + NOUN	BNC	14	-5.031	0.000
	I-EN	156	-1.190	0.242
WORK + NOUN	BNC	64	-3.679	0.000
	I-EN	137	-1.704	0.091
LIFE + NOUN	BNC	76	-3.354	0.001
	I-EN	123	-2.083	0.038
WAY + NOUN	BNC	71	-2.532	0.003
	I-EN	86	-2.015	0.017

Table A-11: Mann-Whitney U Test for Adjective-Noun collocations in BE and AE

Mann-Whitney U Test for Adjectives	Corpora	U	Z	P
NEW + NOUN	BNC	35	-4.463	0.000
	I-EN	89	-3.003	0.002
GOOD + NOUN	BNC	54	-3.949	0.000
	I-EN	127	-1.975	0.049
FIRST + NOUN	BNC	88	-3.030	0.002
	I-EN	114	-2.326	0.020
OTHER + NOUN	BNC	91	-2.948	0.003
	I-EN	112	-2.380	0.017

A.2 Appendix for the results of the corpus-based frequency analysis of lexical collocations produced by Arab EFLLs and NBESs

Table A-12: The normalised frequencies of the selected 200 *Verb* lemmas in the Arab EFLLs corpus

Verbs	NF	Verbs	NF	Verbs	NF	Verbs	NF	Verbs	NF	Verbs	NF
BE	43705	BELIEVE	737	STUDY	373	ACHIEVE	245	CONTAIN	180	SELL	140
HAVE	10865	TALK	730	ALLOW	364	PAY	245	BREAK	178	CONCLUDE	139
GO	7480	LEARN	728	SPEAK	350	SUFFER	243	TEND	175	OPEN	138
DO	6192	TRY	716	CLOSE	350	DEPEND	243	REQUIRE	173	CRY	138
MAKE	3208	HAPPEN	684	FORGET	340	HOPE	239	PRESENT	172	ADVISE	137
USE	2812	CHANGE	666	CHOOSE	339	CLAIM	239	VOTE	168	FIGHT	135
SEE	2807	CALL	660	INCLUDE	339	SEEM	239	BENEFIT	167	MANAGE	134
GET	2527	WATCH	631	DEVELOP	336	SAVE	239	COMMUNICATE	167	PUBLISH	133
TAKE	2303	STAY	627	READ	329	SET	238	QUIT	166	JOIN	133
THINK	1626	ACCORD	623	DIE	325	KILL	236	IMAGINE	164	ADD	132
SAY	1468	BUY	615	SUPPORT	317	ARGUE	236	PREVENT	164	HOLD	132
WANT	1439	VISIT	603	DISCUSS	312	RUN	228	PRODUCE	164	PREPARE	130
KNOW	1436	MENTION	550	IMPROVE	312	DESCRIBE	223	ATTRACT	163	EXPECT	128
GIVE	1383	KEEP	549	LET	297	PASS	221	BAN	163	PREFER	126
COME	1376	TRAVEL	532	ARRIVE	291	SLEEP	220	WEAR	162	WAKE	126
HELP	1282	AFFECT	527	WALK	287	GROW	216	DRINK	161	FORCE	126
LIKE	1232	LEAD	518	FOCUS	286	SMOKE	216	AVOID	158	OFFER	124
PLAY	1187	MEAN	511	EXPLAIN	283	COMPARE	215	RELATE	156	PLAN	124
SHOW	1171	WRITE	510	SIT	281	WAIT	211	EDUCATE	156	CONVINCE	124
WORK	1138	PUT	491	FINISH	271	REMEMBER	209	SHARE	156	HARM	124
FEEL	1055	ASK	489	TEACH	269	RAISE	209	FALL	153	ACT	123
START	1041	STOP	486	MOVE	267	CONTROL	209	WISH	153	WIN	123
SPEND	1026	DECIDE	473	FAIL	266	SEND	206	DEAL	152	COVER	120
BECOME	962	REACH	432	TREAT	266	SWIM	205	INVOLVE	149	CATCH	120
EAT	946	CREATE	424	PROVE	263	GAIN	204	RELAX	149	DENY	119
FIND	933	PROVIDE	422	BRING	259	HEAR	201	APPEAR	148	DECREASE	119
NEED	891	DRIVE	400	REDUCE	254	CARE	193	RELEASE	147	NOTICE	119
LIVE	814	FACE	397	FOLLOW	253	TURN	190	CARRY	145	APPLY	118
TELL	792	STATE	395	AGREE	253	BUILD	190	INTRODUCE	143	PROMOTE	115

LOOK	789	LOSE	395	MEET	249	BEGIN	190	SOLVE	143	CONNECT	115
CAUSE	784	INCREASE	391	UNDERSTAND	249	CONTINUE	183	EXPERIENCE	142	RELY	115
ENJOY	774	CONSIDER	378	REALIZE	247	ENTER	183	BURN	140	LIMIT	115
LOVE	747	LEAVE	373	END	245	BASE	183	ENCOURAGE	140	RETURN	114
										OCCUR	113
										IGNORE	113

Table A-13: The normalised frequencies of the selected 200 *Verb* lemmas in the NBESs Corpus

Verbs	NF	Verbs	NF	Verbs	NF	Verbs	NF	Verbs	NF	Verbs	NF
BE	36763	KEEP	436	SIT	254	SELL	185	FAIL	141	PULL	117
HAVE	11589	HOLD	416	SUGGEST	253	BREAK	184	SERVE	139	OPERATE	117
DO	4758	TURN	394	INCREASE	253	RETURN	183	DRIVE	139	AFFECT	117
SAY	2833	BRING	376	LOSE	251	ACCEPT	181	PREPARE	138	REGARD	114
GO	2013	LIKE	376	CHANGE	250	WAIT	180	OCCUR	138	THANK	113
MAKE	1868	HELP	368	OFFER	247	RAISE	179	REPRESENT	137	OBTAIN	113
GET	1855	SET	365	DEVELOP	242	CAUSE	178	KILL	132	COMPARE	113
SEE	1646	RUN	364	ADD	241	APPLY	175	EXIST	132	ACT	112
KNOW	1604	BEGIN	358	READ	237	COVER	175	DISCUSS	131	ANNOUNCE	112
TAKE	1546	WRITE	358	FALL	236	LIE	171	CLOSE	131	PUBLISH	111
THINK	1305	START	356	REMEMBER	228	REDUCE	171	DEAL	131	GONNA	110
COME	1301	PAY	339	SPEAK	224	BASE	170	ARGUE	130	MAINTAIN	110
GIVE	1150	LEAD	336	BUY	222	REPORT	167	PROVE	129	DEPEND	109
USE	1107	MOVE	328	RECEIVE	222	CONCERN	167	PRESENT	128	POINT	109
LOOK	935	PLAY	324	BUILD	215	EXPLAIN	166	MANAGE	128	ENCOURAGE	108
FIND	849	LIVE	307	SEND	215	WATCH	164	PLACE	127	INDICATE	108
WANT	781	LET	306	AGREE	214	CLAIM	163	INTRODUCE	127	EXPRESS	107
TELL	653	HEAR	303	GROW	214	CONTAIN	160	ENJOY	126	SUFFER	107
PUT	598	ALLOW	299	DECIDE	213	SUPPORT	160	PICK	125	AVOID	105
BECOME	588	BELIEVE	298	STOP	209	STAY	159	ENSURE	125	VISIT	105
LEAVE	588	INVOLVE	282	DESCRIBE	208	ESTABLISH	159	EAT	125	FORCE	104
WORK	587	HAPPEN	278	WIN	206	CUT	154	LOVE	124	DESIGN	104
MEAN	584	REMAIN	274	UNDERSTAND	204	CHOOSE	153	WEAR	124	FORGET	104
NEED	548	MEET	273	REACH	200	RISE	151	CATCH	124	TEND	104
SEEM	531	CONSIDER	269	DIE	195	JOIN	151	IMPROVE	123	TREAT	104
FOLLOW	521	CARRY	269	LEARN	195	BEAR	150	ENTER	122	FINISH	103

FEEL	519	EXPECT	268	DRAW	192	FACE	150	REFER	121	WONDER	103
INCLUDE	516	CONTINUE	268	CREATE	191	SUPPOSE	150	ARRIVE	121	REFLECT	100
ASK	513	PRODUCE	266	SPEND	191	ACHIEVE	149	RELATE	121	SAVE	100
SHOW	502	STAND	266	OPEN	189	SEEK	149	DETERMINE	120	REMOVE	100
PROVIDE	486	APPEAR	265	WALK	188	FORM	148	IDENTIFY	119	PROPOSE	99
TRY	472	REQUIRE	261	PASS	187	ACCORD	146	END	119	ADMIT	98
CALL	457	TALK	257	HOPE	186	WISH	145	PLAN	118	ASSUME	98
										REPLACE	97
										THROW	97

Table A-14: Extracted Verb-Noun Collocations with LL scores for the Arab EFLLs Corpus

Verb-Noun collocations	LL	Verb-Noun collocations	LL	Verb-Noun collocations	LL
BE SYMBOL	8.55	PLAY STATION	6.68	EAT SNACK	13.75
BE GUM	6.64	ALLOW STUDENT	8.43	EAT MEAT	13.58
BE CONSUMING	5.61	ALLOW PEOPLE	7.91	EAT BREAKFAST	12.97
BE NOTHING	5.51	REMAIN BIAS	5.02	*EAT ICECREAM	6.61
HAVE FUN	74.04	REQUIRE HOUSEHOLD	8.91	EAT JUNK	6.42
HAVE NOTHING	8.21	SIT TOGETHER	5.65	EAT RICE	5.23
DO ANYTHING	54.85	INCREASE VARIABILITY	7.55	LOVE MARRIAGE	22.52
DO SOMETHING	34.34	INCREASE BLOOD	6.13	LOVE ONE	6.47
DO EVERYTHING	11.92	INCREASE POPULATION	5.86	WEAR UNIFORM	22.44
DO SOMETHING	8.07	LOSE WEIGHT	59.81	CATCH FISH	5.33
DO NOTHING	8.04	DEVELOP NATION	39.56	THANK GOD	6.17
GO SHOPPING	35.75	DEVELOP COUNTRY	22.32	*THANK GAD	4.01
GO HOME	7.6	DEVELOP GALLSTONE	7.95	ENCOURAGE STUDENT	8.58
*GO TE (TO)	7.42	ADD MUSIC	14.95	*VISIT INDIA	6.78
*GO TP (TO)	6.8	READ HOLY	7.27	SAVE MONEY	11.5
*GO BAK	6.48	READ BOOK	6.79	SAVE TIME	8.25
MAKE PEOPLE	20.22	READ PROGRAMME	5.62	SAVE LIFE	5.88
MAKE SENSE	14.35	READ QURAN	5.28	REMOVE SEGREGATION	5.61
MAKE DECISION	8.7	SPEAK ENGLISH	10.52	DRIVE TEST	7.86
MAKE EXRCISE*	6.02	BUY CIGARETTE	11.16	LEARN OUTCOME	7.56
MAKE LIFE	5.77	BUY ANYTHING	10.54	PLAY GAME	50.24
MAKE REVISION	5.07	BUY GIFT	7.56	USE CELL	7.61
*KNOW ADAYS	10.13	SEND E-MAILS	8.81	WATCH TV	10.8
*KNOW WHATS	5.07	SEND EMAILS	6.56	WATCH FILM	10.56
*COME BACKE	6.64	STOP SMOKING	47.78	WATCH TELEVISION	9.79
GIVE BIRTH	8.28	LEARN INSTITUTION	34.72	WATCH CARTOON	7.66
GIVE STUDENT	7.72	LEARN PROCESS	5.1	STAY HOME	5.9
USE COMPUTER	84.61	SPEND TIME	6.39	CUT TREE	11.88
USE TECHNOLOGY	24.28	SPEND MONEY	5.32	JOIN COLLEGE	32.13
USE HERB	6.83	BREAK TIME	7.04	BEAR BABY	5.87
LOOK FORWARD	14.35	RETURN HOME	7.52	*DRIVE ACAR	6.7
FIND WAY	6.656	RETURN POLICY	5.65	KILL PEOPLE	9.67

FIND SOLUTION	5.48	RAISE AWARENESS	10.97	CLOSE FRIEND	5.38
WANT SEWAGE	6.8	RAISE FUND	6.58	ENJOY SWIMMING	10.64
WANT RETAINER	5.42	CAUSE PROBLEM	41.95	ENJOY SHOPPING	5.89
BECOME PART	15.81	CAUSE OBESITY	19.74	EAT DINNER	17.75
BECOME SLAVE	10.27	CAUSE AIR	13.43	EAT LUNCH	15.83
WORK MOM	9.72	CAUSE INJURY	13.16	PLAY FOOTBALL	230.2
INCLUDE MEAT	5.09	CAUSE VIOLENCE	9.54	PLAY FOOT	31.5
ASK SOMEONE	7.74	CAUSE HEART	8.54	*PLAY FOORBALL	12.68
*ASK ALLAH	6.74	CAUSE DISTORTION	5.72	PLAY SPORT	10.91
KEEP FIT	13.21	APPLY TAX	5.36	PAY ATTENTION	56.36
KEEP SILENCE	7.52	REDUCE SMOKING	9.54	PAY TAX	10.66
KEEP TRACK	5.61	REDUCE GREENHOUSE	6.11	PAY BILL	5.98
HELP PEOPLE	14.71	MOVE CASTLE	8.2	SET EMISSION	5.98
HELP STUDENT	12.08	START PLANNING	5.32	WRITE ESSAY	9.61
HELP SMOKER	9.54	START LIVING	5.32	WRITE POETRY	6.39
HELP SOMEONE	5.83				

Table A-15: Extracted Verb-Noun collocations with LL scores for the NBESs Corpus.

Verb-Noun collocations	LL	Verb-Nouncollocations	LL	Verb-Nouncollocations	LL	Verb-Nouncollocations	LL	Verb-Noun collocations	LL
BE SOMETHING	84.2	GIVE EFFECT	10.5	PAY WORK	26.57	SPEAK CLOCK	20.5	REDUCE AGENT	5.41
BE NOTHING	77.37	GIVE PRIMACY	9.59	PAY GAP	13.49	SPEAK ENGLISH	8.64	REDUCE DEFECT	30.01
BE EARNEST	21.8	GIVE IMPETUS	8.36	PAY MONEY	8.14	SEND LOCALIZATION	20.05	REDUCE YIELD	23.35
BE PART	12.97	GIVE DETAIL	8.01	PAY TAX	8.1	SEND SIGNAL	11.41	REDUCE UNCERTAINTY	19.37
BE ANATHEMA	10.97	GIVE REASON	7.77	PAY FEE	7.94	SEND INFORMATION	8.46	WATCH TELEVISION	39.84

HAVE ACCESS	89.61	GIVE MOMENT	7.5	PAY PERIOD	6.77	LEARN DISABILITY	207.18	STAY VISITOR	26.18
HAVE DIFFICULTY	51.18	GIVE WAY	7.2	PAY GUEST	6.36	LEARN PROCESS	68.86	CUT EDGE	21.57
HAVE NOTHING	46.28	GIVE NOTICE	6.23	MOVE AVERAGE	22.48	LEARN STYLE	49.73	CUT CORNER	18.8
HAVE IMPLICATION	34.14	GIVE TEMPERATURE	6.11	MOVE PART	6.37	LEARN EXPERIENCE	43.98	CUT FLOWER	16.14
HAVE CHILD	18.74	GIVE FEEDBACK	5.53	MOVE WATER	5.17	LEARN ORGANIZATION	42.66	CUT CONGESTION	11.66
HAVE TROUBLE	13.7	GIVE PEOPLE	5.53	PLAY FOOTBALL	14.42	LEARN THEORY	26.17	CUT THROAT	9.3
HAVE FORMULA	11.5	GIVE ARCHA0EOLOGIST	5.47	PLAY TENNIS	8.84	LEARN CYCLE	21.23	CUT POLLUTION	8.82
HAVE FAITH	8.09	GIVE INSTRUCTION	5.42	PLAY TEXT	7.67	LEARN CURVE	10.85	CUT FARE	7.75
HAVE FUN	7.56	USE EQUATION	39.56	ALLOW USER	26.25	LEARN SOMETHING	10.13	JOIN BELOW THRESHOLD	
HAVE RESOURCE	6.37	USE MATLAB	29.06	ALLOW PEOPLE	17	LEARN SITUATION	9.1	BEAR WITNESS	14.75
HAVE SOMETHING	5.03	USE LANGUAGE	23.43	ALLOW ABORTION	16.06	LEARN STRATEGY	8.79	BEAR DIAMETER	14.32
DO SOMETHING	85.52	USE HANSARD	15.19	ALLOW CUSTOMER	8.69	LEARN ANYTHING	8.39	DRIVE COMPONENT	93.01
DO ANYTHING	55.22	USE COMPUTER	13.57	ALLOW COMPARISON	8.17	SPEND TIME	37.55	DRIVE FORCE	77.89
DO NOTHING	49.9	*USE SOLIDWORKS	13.3	ALLOW RESEARCHER	6.11	SPEND MONEY	16.33	DRIVE SHAFT	23.27
DO THING	49.02	USE POLITICAL	11.13	ALLOW WOMEN	5.09	SPEND FUEL	6.78	CLOSE PROXIMITY	131.54
DO EVERYTHING	15.28	USE DATUM	7.77	REMAIN FAITHFUL	15.32	BREAK CROP	22.04	CLOSE RELATIONSHIP	101.21
DO MATHEMATICS	10.98	USE KEYWORDS	7.01	SIT HEIGHT	9.38	BREAK FREQUENCY	13.8	CLOSE INSPECTION	77.63
DO BUSINESS	10.22	LOOK TONIGHT	15.5	INCREASE DEMAND	59.29	BREAK CURFEW	7.55	CLOSE FRIEND	61.59
GO HOME	44.25	LOOK GLASS	14.72	INCREASE NUMBER	55.36	RETURN HOME	55.56	CLOSE CONTACT	60.01
GO HAND	9.77	LOOK INWARDS	7.57	INCREASE PRODUCTIVITY	50.81	RETURN SERIES	31.01	CLOSE SCRUTINY	32.27
GO BUST	8.62	FIND EVIDENCE	17.58	INCREASE PROFIT	31.99	RETURN AIRPORT/HOTEL	7.61	CLOSE ATTENTION	26.47

GO SHOPPING	8.39	FIND FAULT	7.2	INCREASE SALE	29.73	RAISE AWARENESS	96.27	CLOSE LOOK	25.36
GO CLANG	8.78	FIND LOCATION	5.66	INCREASE EFFICIENCY	28.72	RAISE QUESTION	37.79	CLOSE EXAMINATION	19.32
GO STRAWG*	7.78	WANT HUMANITY	9.14	INCREASE AWARENESS	28.58	RAISE CONCERN	20.98	CLOSE RELATIVE	18.48
GO BONG	7.78	WANT SOMETHING	6.16	INCREASE REVENUE	28.14	RAISE ISSUE	15.66	CLOSE ASSOCIATION	16.62
GO ROUND	7.78	BECOME PART	49.09	INCREASE PROFITABILITY	20.3	RAISE CAPITAL	12.67	CLOSE KIN	13.83
GO HAND-IN-HAND	6.94	BECOME COMMONPLACE	8.03	INCREASE PRODUCTION	14.25	RAISE PRICE	12.43	CLOSE COOPERATION	13.18
GO PING	5.69	BECOME LITERATE	7.35	INCREASE UNEMPLOYMENT	13.33	RAISE DOUBT	12.11	CLOSE MATCH	12.48
MAKE SENSE	197.1	WORK CLASS	470.8	INCREASE BIODIVERSITY	12.61	RAISE FUND	10.21	CLOSE QUARTER	9.83
MAKE DECISION	118.3	WORK PACKAGES	162.4	INCREASE CONSUMER	12.44	RAISE REVENUE	9.93	CLOSE MONITORING	9.49
MAKE USE	51.34	WORK ENVIRONMENT	68.53	INCREASE AMOUNT	11.78	RAISE TAX	6.86	CLOSE FRIENDSHIP	9.46
MAKE MISTAKE	49.67	WORK CAPITAL	37.2	INCREASE IMPORTANCE	11.62	RAISE CHILD	6.67	CLOSE NEIGHBOUR	9.22
MAKE PROFIT	27.37	WORK MEMORY	32.12	INCREASE TREND	11.38	CAUSE PROBLEM	62.77	CLOSE PARALLEL	9.08
MAKE REFERENCE	27.16	WORK CONDITION	32.06	INCREASE COMPETITION	11.31	CAUSE DAMAGE	24.45	CLOSE TIE	8.32
MAKE MONEY	24.55	WORK HOUR	28.71	INCREASE RISK	10.83	CAUSE HARM	21.94	CLOSE COLLABORATION	8.12
MAKE INTERFERENCE	23.89	WORK FORCE	28	INCREASE BREATHLESSNESS	10.56	CAUSE DIFFICULTY	18.7	CLOSE SUBSTITUTE	7.84
MAKE PROCESS	23.21	WORK EXPERIENCE	15.54	INCREASE RELIABILITY	10.44	CAUSE AGGLUTINATION	16.09	CLOSE LOCATION	7.62
MAKE COMPARISON	18.12	WORK PRACTICE	13.81	INCREASE PRESSURE	9.9	CAUSE CONFUSION	11.98	CLOSE ALLY	7.38
MAKE JUDGMENT	16.83	WORK ETHIC	10.63	INCREASE VALIDITY	9.35	CAUSE DEATH	10.89	CLOSE AGREEMENT	6.16
MAKE GENERALIZATION	10.94	WORK FLINT	10.54	INCREASE CAPACITY	8.25	CAUSE CONSTIPATION	9.64	CLOSE RELATION	5.93

MAKE ASSUMPTION	10.29	WORK SCHEDULE	10.12	INCREASE OUTPUT	7.76	CAUSE TROUBLE	7.37	EAT BEHAVIOUR	84.29
MAKE ALLOWANCE	10.01	WORK EXPLANATION	10.02	INCREASE FLEXIBILITY	7.56	CAUSE ERROR	7.13	EAT HABIT	54.44
MAKE PEOPLE	9.17	WORK FUNCTION	9.99	LOSE WEIGHT	59.56	CAUSE INJURY	6.97	EAT DISORDER	51.68
MAKE RESERVATION	6.33	WORK LOAD	7.77	LOSE FAITH	47.91	CAUSE DISRUPTION	6.91	EAT DEVIANCY	15.23
MAKE FRIEND	6.25	WORK BREAKDOWN	5.87	LOSE CONTROL	17.19	CAUSE INFECTION	6.47	EAT LIVER	15.1
MAKE PROGRESS	5.67	WORK PLACEMENT	5.04	LOSE ANTEMORTEM	15.53	CAUSE DISEASE	6.25	EAT PATHOLOGY	8.12
MAKE FUN	5.63	INCLUDE CONSIDERATION	77.65	LOSE CONSCIOUSNESS	14.6	CAUSE SATISFACTION	5.84	EAT PATTERN	7.19
MAKE ARENA	5.56	INCLUDE INFORMATION	6.43	LOSE SIGHT	13.49	CAUSE INTERFERENCE	5.66	LOVE AFFAIR	21.95
MAKE PURCHASE	5.38	INCLUDE THING	5.74	LOSE BUSINESS	10.03	APPLY FORCE	23.19	LOVE TRIANGLE	21.1
MAKE CONCESSION	5.32	ASK QUESTION	70.92	LOSE CONFIDENCE	9.95	APPLY LOAD	18.94	LOVE RELATIONSHIP	14.17
MAKE ADJUSTMENT	5.18	ASK PARTICIPANT	18.96	LOSE MONEY	9.05	APPLY TORQUE	13.49	LOVE SONG	12.79
MAKE IMPROVEMENT	5.08	KEEP TRACK	60.23	LOSE MARKET	9.01	APPLY TENSION	10.16	LOVE ELEGY	11.62
KNOW ALLERGY	48.96	KEEP PACE	25.9	LOSE EVERYTHING	8.73	REDUCE COST	79.99	LOVE ONE	11.59
KNOW SOMETHING	20.79	KEEP INFLATION	12.89	LOSE NOTHING	7.24	REDUCE SUGAR	27.96	LOVE POTION	10.55
KNOW NOTHING	15.62	KEEP RECORD	11.31	LOSE ENERGY	7.09	REDUCE RISK	24.43	LOVE POETRY	10.54
KNOW DRUG	10.05	KEEP COST	9.9	DEVELOP COUNTRY	957.8	REDUCE POVERTY	23.51	LOVE STOREY	8.01
KNOW EVERYTHING	9.33	HELP STUDENT	20.89	DEVELOP WORLD	106.3	REDUCE MORTALITY	21.3	LOVE SONNET	7.79
KNOW ANYTHING	8.15	HELP PEOPLE	17.65	DEVELOP NATION	52.34	REDUCE UNEMPLOYMENT	15.14	WEAR UNIFORM	26.35
KNOW TODAY	7.22	HELP LEARNER	7.3	DEVELOP ECONOMY	14.1	REDUCE INEQUALITY	13.98	WEAR MUFTI	23.55
KNOW CONCENTRATION	5.62	SET GOAL	16.72	DEVELOP BREAST	11.95	REDUCE EXPENDITURE	13.12	WEAR PATTERN	8.72

KNOW SOMEONE	5	SET SAIL	13.82	DEVELOP SOFTWARE	9.45	REDUCE DISTURBANCE	12.37	WEAR CLOTHES	8.19
COME ROUND	5.22	SET PORT	8.72	DEVELOP STRATEGY	9.42	REDUCE NITRATE	12.11	WEAR GLOVE	6.74
GIVE RISE	376.9	SET OBJECTIVE	5.97	ADD VALUE	33.25	REDUCE TRAFFIC	10.66	WEAR MAKE-UP	6.15
GIVE BIRTH	91.22	SET STANDARD	5.82	ADD MICRO-NUTRIENTS	24.55	REDUCE BLOOD	9.76	WEAR MARK	5.54
GIVE FORMULA	44.76	WRITE FORMULA	30.21	ADD FEATURE	21.91	REDUCE HEALTH	9.59	ENCOURAGE PEOPLE	40.66
GIVE INFORMATION	28.88	START POINT	245.9	ADD WEIGHT	9.41	REDUCE CONGESTION	8.39	ENCOURAGE INVESTMENT	13.35
GIVE ADVICE	28.32	START MATERIAL	40.94	ADD ITEM	8.75	REDUCE PAIN	8.25	ENCOURAGE STAFF	12.94
GIVE EVIDENCE	20.46	START POSITION	21.05	ADD PRODUCT	6.99	REDUCE ENERGY	7.25	ENCOURAGE TEAMWORK	9.67
GIVE PRIORITY	19.31	START SMOKING	6.56	ADD TAX	5.61	REDUCE PLASMA	7.13	ENCOURAGE STUDENT	6.78
GIVE TIME	15.91	PAY ATTENTION	129.1	READ COMPREHENSION	150.0	REDUCE DOSE	6.67	ENCOURAGE EMPLOYEE	6.22
GIVE CLUE	13.63	PAY EMPLOYMENT	33.14	SPEAK LANGUAGE	74.8	REDUCE DRAG	6.67	ENCOURAGE SHAREHOLDER	6.06
GIVE INSIGHT	12.91	PAY DIVIDEND	30.3	SPEAK DISCOURSE	32.56	REDUCE STRESS	6.66	SAVE MONEY	42.18
GIVE CREDIT	10.6	PAY TIME	29.47	SPEAK WORD	29.59	REDUCE ERROR	6.34	SAVE TIME	30.7
								SAVE COST	7.29

Left	KWIC	Right
The doctor suggested me a bed rest of more than 8 months. Those were the days when I was planning	to take admission	in university for graduation.
also that I will not bel able to pursue my studies. A very dear friend of mine asked me	to use computer	for pursuing your studies as you can get enrolled in an online learning course.
Meanwhile, those women prefer	to watch Oprah"s*	show as the only breather for them.
To many people, plastic bags are considered	to be ideal	when it comes to shopping because these bags are free and durable.
Her article explains how technological devices have and will continue	to impact schools	across the world.
throughout these past years making it much easier and efficient for us	to complete things	quickly, nevertheless, we still can live without them and human thinking and input is still used.
the administrator of the Autism Research Centre indicated that, "We can use computers	to teach emotion	recognition and to simplify communication by stripping out facial and vocal emotional expressions and slowing it down using email instead of face-to-face real-time modes" (Giulia Rhodes)
Also, to focus more in real life activities and	to perceive everything	that is artificial.
It has never been easier	to make friends	than it is right now.
Sometimes devices can be deadly when they distract a person. </s><s> A person is more likely	to get hit	by a car if they cross the street with their phones in their hands.
Certainly, plastic sacks should be recycled more vigorously and people should be more environmentally conscious	to avoid pollution	issues.
My parents would argue with me for hours just	to stop playing.	
First, exercise is vital	to maintaining health	, decreasing stress, relieving depression and anxiety, as well as improving one"s mood, sleep, and immune system.
Thus meaning, if one wants	to lose weight	or even meet specific fitness goals, one may need to exercise more often.
They would let their sons go	to get education	but not their daughters, away from financial issues.
It could lead	to reduced sense	of individuality is a big problem when it comes to cloning
from the region at the Forum for the Future in Tunisia, where all governments committed	to supporting progress	in gender equality.
Women rights in the Arab world As we all know, women are the other half of the world. </s><s> God created both gender	to make balance	in the world and as he said in the holy Quran "men are the protectors of the women," but unfortunately Arab understands this concept wrong.
This is a problem that we still didn"t solve, but some countries are trying to raise awareness and	to help people	to understand how education is so important.
However, according to Meris Lutz article, Arab World Failing Women in Gender Equality, the most appropriate course of action would be	to improve education	for women.
Women in the Middle East are considered	to be housewives	in the future; this result in them having less job opportunities …

Sonenshine suggests that it"s very important	to conduct programs	that can encourage women leadership and confidence in political life.
The article mainly suggests "the key is to convince countries that girls" education benefits not only women, but also the society as a whole, leading	to lower birth	rates and higher survival rates among mothers and children" (Lutz).
The key to close this gap is to convince people that woman will benefit the whole society. </s><s> And finally, to get rid from the poverty is	to let girls	study and have their own lifestyle.
but also their right to free expression and their right	to take part	in their country"s transition, only nine women legislators won seats…
The founders made this website	to help people	choose the right products that bring no harm to the person using them.
I am not sure if they gained a significant number of donors, but their comment section is empty which seems that no one cared	to give feedback	or express their opinion on donating.
Leonard claims that cosmetics companies feel free	to use chemicals	in their products because there have been no laws to stop them.
In order to achieve and accomplish your goals you need	to have devotion	and a strong commitment and take your work very seriously.
You are educating yourself self and developing your skills in order	to make something	beneficial to your country.
Women also don"t get appointed for the job because the interviewer thinks that they are going	to have children	, an investigation shown in "Nexis" tells that a women was send …
Another reason	to love work	is that it makes you feel more alive.
We have to work for every chance we get,	to earn something	very important which will help us in future and help us to work…
It doesn"t matter what your specific major is, as long as you have a college education, it will enable you	to develop skills	and abilities that will provide you with more opportunities …
I personally enjoy working: I try my best	to complete tasks	I"m given, find new jobs, and read about recent job trends so I can …
There are many factors that caused global warming	to take place	, and there are many signs that global warming is in fact happening …
The media generally uses persuasion	to compel individuals	to buy or do something.
The original iPhone was made so organic	to become inevitability	in one"s life.
Topic Sentence for Body 2: The ad which was about how	to make Ferrero	Rocher was the features that got the attraction to a lot of people.
The text was written in bold and in colors	to get peoples	attention and was said in an enthusiastic way.
When the drumroll starts, all of the teenagers in the ad are about	to do something	exciting that requires the watcher to hold their breath and let it…
and "The Unbelievable Game" show a lot more than that as a method	to seduce customers	in buying their product.
Bernard Beck claims that we approach our social assets	to find comfort	for our aspirations and reassurance and dreads.
Bernard beck talks about us, the people in this modern world on how we approach our social assets	to find comfort	for our aspirations and reassurance and dreads.
In conclusion, to prove yourself in society you don"t need to be athlete, artist or political, you just need	to use internet	to be fame.

They may be afraid of the choices that prisoners" will make and will most likely not want this law	to take place.	
Imprisonment is a way	to punish criminals	by putting them in jail for doing illegal crimes prohibited by law …
According to the definition, the main purpose of euthanasia is	to finish someone"s*	life provided the patient has no more hope of living anymore or …
Moreover, the way to perform euthanasia should be defined otherwise this will open the door for a legalized murder. </s><s> The way	to perform euthanasia	may be also medical as to give the patient a fast killing poison or …
In fact, governments in many countries want	to spread peace	for two reasons, security in the region, where citizens tend to be relaxed.
These friends have to deal with the side effects that a drug user can have such as mood swings and random aggression. They can also be negatively influenced	to use drugs	themselves.
by Mark Hendrickson, the author"s main point is to state that political elites took advantage of the hoax	to reap money	by selling carbon permits(13).
First of all, According to mark Hendrickson, The earth is supposed	to be warmer	which is good since we have emerged from an ice age and it …
From personal experience when I have gone	to buy games	my parents did not really ask much about the game and they did…
and also being able "to resolve different levels of gray" which she gave an example as if you were driving in a fog you are able	to recognize shapes	better.
All my life I have had to deal with some sort of curfew whether it be it high school to turn in by 10 or now in University	to be home	by midnight or spend the night somewhere else.
This brings to mind this one time when I was working in an architectural and consulting company and a client asked us	to send someone	to take measurements and propose a design for his house"s basement
Because it won"t do me any good if I cared about what people say. </s><s> Nothing is going	to satisfy people	these days, they will always say bad things and rumors about anyone.
The greatest concern in this regard appears	to be time	management.
Also, one has	to take measures	to ensure that large amounts of body fats do not accumulate by stabilizing blood sugar.
team work will be taught respect for one another in the spirit of sports and maintained in the workplace or schools also managing	to follow rules	just like in matches or games and making it their attitude and lifestyle.
They take the taxes from the raise prices of cigarettes	to create something	useful for the smokers to decrease smoking and then to stop it.
I asked my father, why is the US getting involved in other countries; he always said that "they create reasons	to enter countries	which they want to conquer".
I can simply tell you that it"s all about money. The US invaded Iraq	to protect Kuwait	, but what was the real reason for Iraq to invade Kuwait?
One of the 5% described it as teleporting from the real world were face	to face contact	occurs and social life gets built up into a virtual world where he …
Campaigns and awareness against domestic violence needs to spread vastly across the Middle East	to help victims	and inform citizens about the affects domestic violence holds against society.
and by the 1200"s it became very popular in the region because of Muslims not being allowed	to drink alcohol.	

Some of the games have a functions of voice chatting but I think it is not enough to make a person comfortable in face	to face conversations.	
The debates lack of scientific knowledge and resources are not enough	to prove anything.	
Using such methods can contribute to the reduction of global warming severe effects. </s><s> People, on a personal scale, can use hybrid or electric cars	to recover climate	change and stop further warming of the Earth.
I remember I had trouble the first month blending in with the other kids and trying not	to speak Arabic.	
Using media as part to steer people by announcement them. Media should be a strong	to convince people	and lead them to the right way for instance, TV show, brochure,…
I was sad at the beginning of the last semester because I have	to take IEP	, but then I realized it was beneficial to me.
Conclusion of essay: How this lesson affected you personally. Main ideas: The lesson I learnt,	to appreciate life	and to be grateful for everything I have.
I assumed that it was fine to miss a few classes and take tutoring classes in order	to keep track	of my studies.
Therefore, I will strive for a positive transformation in my attitude towards academic life. Moreover, I will try	to avoid anything	in life that pushes me towards negativity, and rather focus more ….
I gained confidence in myself that I actually can do something if I put my head into it, knew how	to handle duties	, and proved them wrong so they no longer can mock me …
I am very nice wonderfoul. CEPA 4 200605652 I like	to do alot*	of activetys , sports is one of them I like sport becos it make you ..
Than Thank you very much CEPA 3 200606986 In last weekend my Friends organise	to travel Ras-Alkhimah*	to see the places there , Wher we went there we ..
we first have to go to the starbooks cafe for the brecfast and than we go to the shopping and we don"t forget	to eat oar	langh there in any restorant in the shopping roud.
It was a great gift for me and it was my first mobile phone . I like	to play games	in it like tennis and puzzels game. a
We arrived to Cairo Airport at 2:00 pm . Firstly we went to the hotel	to take arest*	and to eat.
We visited churches and museums. </s><s> Before leaving I wanted	to do something	I always wanted to do.
I hug him, I was so happy to find him and my mother too. In another day, we went	to place wich*	called "Sunway lagon" it was a shopping center.
Name Noof Ebeed Grade 11 I like to be ll teacher and it is my future job. love	to be teacher	of math in the future because I want to sarvel my country.
I"m no go to Pares and no plae food pooland	to go scool*	ithe g no Mane noto cook no to sleping and nosemmingIm to holiday hape.
and then he come back to his family bringing his fishs and pearls to sold it	to get money	to his family.
I like writing story my family, life ago. </s><s> Last, help abook people and My	to use agood*	words.
I go to play football with boys and my mother and my sister go to ship	to buy water	and we was very happy .
When we feeling tired ,we staied on the coffee shop	to drank tea	, coffee or any thing we want .
school in this time to swim with me besaids we like	to play voleball*	togather we make tow thems the boys them and ..
Next I went with my friends	to play buding*	and beliardo.

My grandfather Asked me	to came because*	sleep in the tent to the desert .
but my mother rembered that she forgote	to put salt	in the rice.
I knew from the minute I got there that we were going	to have fun.	
In the XXX end it is advaiseble	to stop smoking	and start living, please let"s stop this cigarettes war and …
There are many ways	to stop smoking	like walking, chewing gum, drinking water or juice .... Etc.
I love	to write poetry	because I think it is a way of expressing our thoughts and feelings,…
How ever , I Like	to tack pictuers*	to keep it as a value things when I was taking pictures ..
when I want to watch film or programme and my little sisters and brothers want	to see carton*	, sometime, I cry because of this .

Figure A.2-1: Concordance lines for to+Verb-Noun collocations[1]

[1] The concordance lines show the students' exact words, unedited. Any spelling mistakes are kept for research purposes.

Table A-16: Extracted Verb-Adverb Collocations with LL scores for the Arab EFLLs corpus

Verb-Adverb collocations	LL	Verb-Adverb collocations	LL	Verb-Adverb collocations	LL	Verb-Adverb collocations	LL
BE~~VERY	990.39	BE ~~ ALONE	5.8	GET ~~ UP	6.62	ENJOY ~~ SO	17.86
BE~~NOT	290.18	BE ~~ FULLY	5.71	TAKE ~~ AWAY	9.39	ENJOY ~~ THERE	11.81
BE~~ SO	275.47	BE ~~ GENERALLY	5.59	TAKE ~~ IMMEDIATELY	7.34	TALK ~~ ABOUT	5.07
BE ~~ TOO	91.07	BE ~~ MAINLY	5.39	THINK ~~ TWICE	16.37	TRY ~~ HARD	10.7
BE ~~ REALLY	68	BE ~~ DIRECTLY	5.39	THINK ~~ DEEPLY	8.18	WATCH ~~ THERE	38.9
BE ~~ STILL	47	BE ~~ MOSTLY	5.32	SAY ~~ OTHERWISE	6.62	STAY ~~ LATE	16.59
BE ~~ ONLY	34.67	BE ~~ OBVIOUSLY	5.08	SAY ~~ NO	5.28	STAY ~~ ALONE	9.51
BE ~~ ALSO	29.86	BE ~~ ENVIROMENTALLY•	5.02	COME ~~ BACK	229.16	STAY ~~ AWAY	6.56
BE ~~ JUST	28.59	HAVE ~~ JUST	95.1	COME ~~ HOME	5.83	VISIT ~~ AGAIN	10.26
BE ~~ ALWAYS	27.31	HAVE ~~ EVER	88.08	LIKE ~~ SO	45.25	MENTION ~~ PREVIOUSLY	5.78
BE ~~ QUIET	25.73	HAVE ~~ RECENTLY	30.8	LIKE ~~ TOO	9.47	TRAVEL ~~ ABROAD	20.97
BE ~~ EXTREMELY	25.62	HAVE ~~ ALREADY	14.91	LIKE ~~ ELSE	6.42	AFFECT ~~ NEGATIVELY	9.02
BE ~~ FAR	15.52	HAVE ~~ NEVER	14.26	LIKE ~~ VERY	5.29	REACH ~~ THERE	5.17
BE ~~ TOTALLY	15.28	HAVE ~~ ALWAYS	7.33	PLAY ~~ TOGETHER	6.81	DRIVE ~~ FAST	19
BE ~~ ACTUALLY	15.13	HAVE ~~ TOO	7.02	WORK ~~ HARD	32.86	DRIVE ~~ CAREFULLY	14.08
BE ~~ ALREADY	12.5	HAVE ~~ VERY	6.22	WORK ~~ TOGETHER	9.55	DRIVE ~~ SPEEDILY	7.66
BE ~~ COMPLETELY	12.07	GO ~~ THERE	96.02	WORK ~~ EFFECTIVELY	6.41	INCREASE ~~ RAPIDLY	7.58
BE ~~ RAPIDLY	11.92	GO ~~ BACK	88.19	FEEL ~~ VERY	35.06	INCREASE ~~ EXPONENTIALLY	7.55
BE ~~ COMMONLY	10.48	GO ~~ ANYWHERE	15.06	FEEL ~~ SO	6.44	STUDY ~~ ABROAD	14.97
BE ~~ LEGALLY	10.12	GO ~~ AGAIN	14.67	SPEND ~~ TOO	9.37	STUDY ~~ HARD	7.13
BE ~~ INCREASEINGLY	10.12	DO ~~ NOT	867.19	SPEND ~~ VERY	5.86	DIE ~~ ANYWAY	13.31
BE ~~ NOW	9.77	DO ~~ NECESSARILY	11.78	SPEND ~~ ALONE	5.11	ARRIVE ~~ THERE	27.15
BE ~~ USUALLY	9.39	DO ~~ SO	8.58	BECOME ~~ TOO	22.14	ARRIVE ~~ SAFELY	5.93
BE ~~ RELATIVELY	9.05	MAKE ~~ SURE	5.43	BECOME ~~ VERY	9.6	SIT ~~ TOGETHER	22.99
BE ~~ EQUALLY	8.09	USE ~~ WISELY	12.85	BECOME ~~ SO	7.33	SIT ~~ STRAIGHT	6.35
BE ~~ HIGHLY	7.74	USE ~~ EXTENSIVELY	10.08	BECOME ~~ QUITE	5.17	FINISH ~~ QUICKLY	7.76

BE ~~ ALMOST	7.67	USE ~~ EXCESSIVELY	7.74	EAT ~~ MUCH	8.59	TEACH ~~ DIFFERENTLY	9.71
BE ~~ EASILY	7.57	USE ~~ REGULARLY	7.23	LIVE ~~ FAR	7.82	TREAT ~~ KINDLY	17.58
BE ~~ MUCH	7.14	SEE ~~ RECENTLY	43.38	LIVE ~~ ONCE	6.98	DEPEND ~~ MUCH	10.6
BE ~~ OFTEN	6.97	SEE ~~ THERE	16.46	LIVE ~~ TOGETHER	6.33	RUN ~~ AWAY	10.56
BE ~~ CLEARLY	6.33	GET ~~ EARLY	15.61	LIVE ~~ AWAY	6	PASS ~~ AWAY	14.63
BE ~~ TRULY	5.92	GET ~~ BACK	14.64	LIVE ~~ THERE	5.57	SLEEP ~~ WELL	5.37
SLEEP ~~ WELL	5.37	BREAK ~~ FAST	6.79	WAKE ~~EARLY	16.54	RELY ~~ TOTALLY	5.79
GROW ~~ UP	8.23	PRODUCE ~~ ENVIRONMENTALLY	5.42	WAKE ~~ UP	15.79	RETURN ~~ BACK	25.55
TURN ~~ UPSIDE	12	BAN ~~ DUE	7.02	ACT ~~ AGGRESSIVELY	14.17		
TURN ~~ DOWN	6.47	FALL ~~ ASLEEP	7.34	RELY ~~ MUCH	10.35		
BREAK ~~ DOWN	8.14	FALL ~~ DOWN	6.49	RELY ~~ TOO	8.09		

Table A-17: Extracted Verb-Adverb Collocations with LL scores for the NBESs Corpus

Verb-Adverb collocations	LL	Verb-Adverb collocations	LL	Verb-Adverb collocations	LL	Verb-Adverb collocations	LL
BE ~~ NOT	8590.97	BECOME ~~ INCREASEDLY	552.49	REMAIN ~~ REGARDLESS	5.94	ESTABLISH ~~ FIRMLY	5.23
BE ~~ ALSO	2709.61	BECOME ~~ PROGRESSIVELY	61.91	REMAIN ~~ QUITE	5.4	CUT ~~ BACK	38.1
BE ~~ VERY	2623.72	BECOME ~~ TOO	48.78	REMAIN ~~ COMPLETELY	5.22	CUT ~~ SHORT	9.02
BE ~~ STILL	1309.66	BECOME ~~ VERY	38.91	MEET ~~ REGULARLY	14.79	CUT ~~ DOWN	8.22
BE ~~ OFTEN	1123.37	BECOME ~~ SO	28.74	CONSIDER ~~ CAREFULLY	10.59	CUT ~~ ALTOGETHER	7
BE ~~ ONLY	916.66	BECOME ~~ EVEN	24.12	CONSIDER ~~ SEPARATELY	10.29	CHOOSE ~~ NOT	23.07
BE ~~ QUITE	717.73	BECOME ~~ EXTREMELY	21.9	CONSIDER ~~ ONLY	8.77	RISE ~~ SHARPLY	22.63
BE ~~ TOO	669.64	BECOME ~~ EVER	21.63	CONSIDER ~~ HIGHLY	7.81	RISE ~~ STEADILY	11.25

BE ~~ ALWAYS	598.57	BECOME ~~ ALMOST	20.39	CONSIDER ~~ HERE	7.17	RISE ~~ RAPIDLY	7.82
BE ~~ HIGHLY	549.33	BECOME ~~ WIDELY	11.75	CONSIDER ~~ VERY	6.04	RISE ~~ DRAMATICALLY	6.88
BE ~~ THEN	538.86	BECOME ~~ POLITICALLY	9.95	PRODUCE ~~ INTERNALLY	22.82	RISE ~~ AGAIN	6.75
BE ~~ NOW	517.71	BECOME ~~ SOCIALLY	9.34	PRODUCE ~~ VERY	11.73	JOIN ~~ TOGETHER	73.98
BE ~~ EXTREMELY	487.68	BECOME ~~ HIGHLY	8.84	PRODUCE ~~ MUCH	7.6	BEAR ~~ PREMATURELY	21.61
BE ~~ GENERALLY	421.42	BECOME ~~ SOMEWHAT	8.83	PRODUCE ~~ EFFICIENTLY	5.05	FACE ~~ DOWNWARD	8.05
BE ~~ MUCH	401.97	BECOME ~~ SLIGHTLY	8.52	STAND ~~ ALONE	62.39	SEEK ~~ WHOLEHEARTEDLY	6.4
BE ~~ JUST	397.24	BECOME ~~ COMMERCIALLY	6.9	STAND ~~ UNSTEADILY	9.25	SEEK ~~ ELSEWHERE	5.92
BE ~~ USUALLY	391.84	BECOME ~~ FULLY	6.88	STAND ~~ STRAIGHT	5.32	FAIL ~~ ADEQUATELY	9.03
BE ~~ REALLY	385.18	BECOME ~~ LARGELY	6.62	APPEAR ~~ QUITE	7.53	SERVE ~~ ONLY	13.23
BE ~~ RELATIVELY	379.74	BECOME ~~ FAR	6.18	APPEAR ~~ FREQUENTLY	5.68	SERVE ~~ PREDOMINANTLY	5.84
BE ~~ CLEARLY	369.15	BECOME ~~ GRADUALLY	5.78	APPEAR ~~ TOO	5.57	DRIVE ~~ NEGLIGENTLY	13.98
HAVE ~~ ALREADY	713.24	LEAVE ~~ ALONE	16.87	REQUIRE ~~ ONLY	8.81	DRIVE ~~ DANGEROUSLY	9.11
HAVE ~~ NOT	513.48	LEAVE ~~ BEHIND	15.82	TALK ~~ ABOUT	62.74	DRIVE ~~ FORWARD	8.43
HAVE ~~ ALSO	372.28	LEAVE ~~ COMPLETELY	10.46	TALK ~~ EXPRESSIVELY	34.29	DRIVE ~~ CAREFULLY	7.74
HAVE ~~ NEVER	323.09	LEAVE ~~ ONLY	7.83	TALK ~~ DIFFERENTLY	8.24	DRIVE ~~ UNDERGROUND	7.59
HAVE ~~ ALWAYS	232.1	LEAVE ~~ OVERNIGHT	6.76	SIT ~~ COMFORTABLY	59.47	DRIVE ~~ SOLELY	7.35
HAVE ~~ PREVIOUSLY	218.61	LEAVE ~~ CIRCULARLY	5.77	SIT ~~ DOWN	26.05	DRIVE ~~ AWAY	6.61
HAVE ~~ VERY	193.8	LEAVE ~~ MUCH	5.1	SIT ~~ BACK	8.94	OCCUR~~ ONLY	28.17
HAVE ~~ RECENTLY	155.78	WORK ~~ TOGETHER	180.15	SIT ~~ DIRECTLY	7.37	OCCUR ~~ FREQUENTLY	24.95
HAVE ~~ YET	112.37	WORK ~~ HARD	147.13	SUGGEST ~~ OTHERWISE	15.86	OCCUR ~~ TOGETHER	17.44
HAVE ~~ SIGNIFICANTLY	86.49	WORK ~~ WELL	80.48	SUGGEST ~~ EVEN	6.28	OCCUR ~~ NATURALLY	14.89

HAVE ~~ ONLY	70.49	WORK ~~ EFFICIENTLY	30.93	SUGGEST ~~ PERHAPS	5.76	OCCUR ~~ ONCE	8.64
HAVE ~~ SUCCESSFULLY	69.12	WORK ~~ EFFECTIVELY	30.69	INCREASE ~~ DRAMATICALLY	57.03	OCCUR ~~ SIMULTANEOUSLY	7.18
HAVE ~~ RELATIVELY	68.38	WORK ~~ CLOSELY	29.12	INCREASE ~~ LINEARLY	38.05	OCCUR ~~ GRADUALLY	7.04
HAVE ~~ TRADITIONALLY	60.99	WORK ~~ CORRECTLY	17.26	INCREASE ~~ EXPONENTIALLY	29.62	OCCUR ~~ THEN	6.81
HAVE ~~ MUCH	59.74	WORK ~~ PROPERLY	15.22	INCREASE ~~ SIGNIFICANTLY	27.65	OCCUR ~~ RAPIDLY	6.29
HAVE ~~ NOW	57.93	WORK ~~ COLLABORATIVELY	14.61	INCREASE ~~ RAPIDLY	16.65	OCCUR ~~ DUE	6.1
HAVE ~~ JUST	50.72	WORK ~~ OVERTIME	6.77	INCREASE ~~ SO	19.94	OCCUR ~~ MAINLY	5.47
HAVE ~~ LARGELY	47.61	WORK ~~ RATHER	6.52	INCREASE ~~ STEADILY	13.79	REPRESENT ~~ GRAPHICALLY	6.44
HAVE ~~ INCREASEDLY	46.39	WORK ~~ PERFECTLY	6.46	INCREASE ~~ SLIGHTLY	10.65	REPRESENT ~~ ALGEBRAICALLY	5.67
HAVE ~~ SLIGHTLY	45.03	WORK ~~ VERY	5.76	INCREASE ~~ GREATLY	10.22	EXIST ~~ INDEPENDENTLY	48.53
DO ~~ NOT	17133.5	WORK ~~ INDEPENDENTLY	5.62	INCREASE ~~ GRADUALLY	7.89	EXIST ~~ ONLY	14.15
DO ~~ SO	628.96	WORK ~~ SO	5.53	INCREASE ~~ SHARPLY	6.77	EXIST ~~ APART	10.28
DO ~~ NECESSARILY	372.35	WORK ~~APPROPRIATELY	5.14	INCREASE ~~ AGAIN	6.12	EXIST ~~ THEN	7.97
DO ~~ ALWAYS	102.64	WORK ~~ ALONE	5.02	INCREASE ~~ APPROXIMATELY	5.6	DISCUSS ~~ PREVIOUSLY	50.79
DO ~~ REALLY	80.25	NEED ~~ NOT	21.45	INCREASE ~~ SUBSTANTIALLY	5	DISCUSS ~~ BELOW	40.42
DO ~~ INDEED	54.73	NEED ~~ ONLY	7.85	LOSE ~~ FOREVER	16.2	DISCUSS ~~ HERE	23.6
DO ~~ WELL	43.21	NEED ~~ NECESSARILY	5.09	LOSE ~~ MUCH	6.23	DISCUSS ~~ ABOVE	21.92
DO ~~ JUST	40.95	SEEM ~~ QUITE	23.34	CHANGE ~~ DRAMATICALLY	31.5	DISCUSS ~~ FINALLY	9.11
DO ~~ EVEN	30.09	SEEM ~~ VERY	18.5	CHANGE ~~ SIGNIFICANTLY	15.4	DISCUSS ~~ BRIEFLY	9.02
DO ~~ AUTOMATICALLY	26.31	SEEM ~~ RATHER	15.79	CHANGE ~~ SLIGHTLY	8.33	DISCUSS ~~ FAR	8.91
DO ~~ ACTUALLY	23.48	SEEM ~~ LIKELY	14.04	CHANGE ~~ SO	7.67	DISCUSS ~~ BEFORE	6.42

DO ~~ HOWEVER	21.86	SEEM ~~ SOMEWHAT	13.49	CHANGE ~~ RAPIDLY	7.08	DISCUSS ~~ FULLY	5.4
DO ~~ EXPLICITLY	18.42	SEEM ~~ TOO	12.93	CHANGE ~~ GRADUALLY	6.51	CLOSE ~~ TOGETHER	7.23
DO ~~ LITTLE	18.1	SEEM ~~ SLIGHTLY	11.22	OFFER ~~ LITTLE	7.6	CLOSE ~~ ENOUGH	7.56
DO ~~ DIRECTLY	17.9	SEEM ~~ ENTIRELY	9.7	DEVELOP ~~ QUICKLY	8.26	DEAL ~~ EXCLUSIVELY	5.8
DO ~~ SIMPLY	16.2	SEEM ~~ ALMOST	9.32	DEVELOP ~~ ENOUGH	6.04	DEAL ~~ INAPPROPRIATELY	5.59
DO ~~ FULLY	11.5	SEEM ~~ SO	8.86	DEVELOP ~~ RAPIDLY	5.83	ARGUE ~~ CONVINCINGLY	18.01
DO ~~ EXACTLY	9.27	SEEM ~~ FAR	8.42	DEVELOP ~~ INDEPENDENTLY	5.43	ARGUE ~~ EVEN	10.26
DO ~~ MANUALLY	7.92	SEEM ~~ COMPLETELY	8.14	DEVELOP ~~ NORMALLY	5.32	ARGUE ~~ PERSUASIVELY	8.46
DO ~~ AWAY	7.66	SEEM ~~ FAIRLY	7.94	ADD ~~ TOGETHER	15.41	ARGUE ~~ INSTEAD	7.72
DO ~~ MERELY	7.28	SEEM ~~ NOT	7.89	ADD ~~ CUMULATIVELY	6.36	PROVE ~~ VERY	17.02
DO ~~ INITIALLY	7.03	SEEM ~~ PERFECTLY	6.52	READ ~~ ALOUD	25.29	PROVE ~~ OTHERWISE	11.14
DO ~~ ACTIVELY	6.79	SEEM ~~ THEN	6.02	FALL ~~ BACK	48.72	PROVE ~~ PARTICULARLY	5.77
DO ~~ COMPLETELY	5.73	SEEMS ~~ PARTICULARLY	5.44	FALL ~~ ASLEEP	35.13	PROVE ~~ CONCLUSIVELY	5.15
DO ~~ SIGNIFICANTLY	5.71	SEEMS ~~ HIGHLY	5.2	FALL ~~ SHORT	24.56	PRESENT ~~ HERE	22.69
DO ~~ ADEQUATELY	5.55	SEEMS ~~ LITTLE	5.13	FALL ~~ APART	20.71	MANAGE ~~ EFFECTIVELY	33.34
DO ~~ SUCCESSFULLY	5.49	FOLLOW ~~ OPTICALLY	11.14	FALL ~~ FAR	15.01	MANAGE ~~ CAREFULLY	15.33
DO ~~ READILY	5.26	FOLLOW ~~ EXACTLY	7.66	FALL ~~ DOWN	10.09	MANAGE ~~ APPROPRIATELY	9.17
SAY ~~ PREVIOUSLY	8.27	FOLLOW ~~ CLOSELY	7.53	FALL ~~ SIGNIFICANTLY	7.97	MANAGE ~~ WELL	5.96
GO ~~ BACK	229	FEEL ~~ ENOUGH	14.93	FALL ~~ AWAY	7.01	MANAGE ~~ EFFICIENTLY	5.47
GO ~~ FAR	164.3	FEEL ~~ STRONGLY	14.63	FALL ~~ BEHIND	6.08	MANAGE ~~ PROPERLY	5.22
GO ~~ AHEAD	78.59	FEEL ~~ VERY	10.47	FALL ~~ RAPIDLY	5.95	PLACE ~~ RANDOMLY	10.84
GO ~~ AWAY	26.31	FEEL ~~ EMOTIONALLY	8.25	SPEAK ~~ DIFFERENTLY	21.14	PLACE ~~ TOO	8.82
GO ~~ TOO	20.52	FEEL ~~ SO	8.05	SPEAK ~~ ALOUD	13.14	PLACE ~~ ANYWHERE	7.78
GO ~~ SO	18.74	FEEL ~~ TOO	6.04	SPEAK ~~ FLUENTLY	7.04	PLACE ~~ SO	5.55

GO ~~ STRAIGHT	16.47	FEEL ~~ QUITE	5.2	SPEAK ~~ FREELY	6.11	PLACE ~~ DIRECTLY	5.27
GO ~~ SOMEWAY	14.31	ASK ~~ IMPATIENTLY	7.09	SPEAK ~~ TENTATIVELY	5.94	PICK ~~ UP	10.71
GO ~~ WELL	12.71	SHOW ~~ BELOW	99.75	SPEAK ~~ CONFIDENTLY	5.6	EAT ~~ HEALTHILY	22.46
GO ~~ ABROAD	10.15	SHOW ~~ CLEARLY	19.87	SPEAK ~~ ABOUT	5.4	EAT ~~ MAINLY	12.79
GO ~~ SOMEWHERE	10.02	SHOW ~~ SCHEMATICALLY	13.3	BUY ~~ CHEAPLY	6.12	EAT ~~ AWAY	7.91
GO ~~ NOWHERE	8.96	SHOW ~~ GRAPHICALLY	12.64	BUILD ~~ LOCALLY	6.16	EAT ~~ TWICE	6.28
GO ~~ TOGETHER	8.19	SHOW ~~ DIAGRAMMATICALLY	12.16	SEND ~~ BACK	51.4	LOVE ~~ DEARLY	7.5
GO ~~ EVEN	7.73	SHOW ~~ VERY	7.98	SEND ~~ AWAY	13.4	IMPROVE ~~ GREATLY	6.9
GO ~~ WRONG	6.92	SHOW ~~ EVEN	7.34	SEND ~~ ABROAD	6.2	REFER ~~ BACK	82.57
GO ~~ SIGNIFICANTLY	5.91	SHOW ~~ JUST	6.81	AGREE ~~ WELL	7.58	REFER ~~ AS	14.04
GO ~~ DOWN	5.85	SHOW ~~ RELATIVELY	6.47	GROW ~~ RAPIDLY	56.46	RELATE ~~ DIRECTLY	14.86
MAKE ~~ SURE	173.65	SHOW ~~ ONLY	6.15	GROW ~~ FAST	12.95	RELATE ~~ BACK	9.8
MAKE ~~ VERY	28.77	SHOW ~~ MUCH	6.07	GROW ~~ STEADILY	11.22	RELATE ~~ SPECIFICALLY	6.88
MAKE ~~ EVEN	12.88	SHOW ~~ HERE	5.92	GROW ~~ ENORMOUSLY	11.08	DETERMINE ~~ EXPERIMENTALLY	8.6
MAKE ~~ MUCH	9.7	SHOW ~~ SIGNIFICANTLY	5.01	GROW ~~ DRAMATICALLY	9.13	DETERMINE ~~ GRAVIMETRICALLY	7.95
MAKE ~~ EXTREMELY	9.02	PROVIDE ~~ MUCH	8.04	GROW ~~ SLOWLY	9.09	DETERMINE ~~ EXACTLY	7.8
MAKE ~~ FAR	8.26	PROVIDE ~~ ENANTIOMERICALLY	6.62	GROW ~~ CONSIDERABLY	8.99	END ~~ FARTHEST	9.98
MAKE ~~ HERE	8.06	PROVIDE ~~ VERY	6.04	CONSIDER ~~ APPROXIMATELY	7.72	END ~~ HAPPILY	7.54
MAKE ~~ SO	5.44	TRY ~~ DESPERATELY	7.47	GROW ~~ INCREASEDLY	6.11	PLAN ~~ AHEAD	7.94
GET ~~ BACK	47.01	KEEP ~~ INDOORS	8.36	DECIDE ~~ NOT	8.11	PULL ~~ BACK	9.79
GET ~~ AWAY	46.07	KEEP ~~ AWAY	6.71	DESCRIBE ~~ BELOW	20.84	PULL ~~ TOGETHER	7.68
GET ~~ PROGRESSIVELY	26.94	KEEP ~~ DOWN	5.25	DESCRIBE ~~ ABOVE	12.48	PULL ~~ FORWARD	6.78
GET ~~ TOO	16.63	HOLD ~~ BACK	40.73	DESCRIBE ~~ PEJORATIVELY	7.04	OPERATE ~~ INDEPENDENTLY	13.59

GET ~~ TOGETHER	9.46	HOLD ~~ TOGETHER	38.43	DESCRIBE ~~ HERE	6.46	OPERATE ~~ EFFICIENTLY	13.31
GET ~~ WELL	5.55	TURN ~~ AWAY	35.65	DESCRIBE ~~ PREVIOUSLY	6.12	OPERATE ~~ EFFECTIVELY	9.83
SEE ~~ ALSO	59.82	TURN ~~ BACK	14.77	DESCRIBE ~~ AS	5.43	OPERATE ~~ FREELY	8.88
SEE ~~ BELOW	34.37	TURN ~~ UPSIDE	5.81	WIN ~~ BACK	9.93	OPERATE ~~ ABROAD	8.44
SEE ~~ CLEARLY	27.04	BRING ~~ ABOUT	137.93	UNDERSTAND ~~ FULLY	17.21	REGARD ~~ HIGHLY	5.61
SEE ~~ AS	24.31	BRING ~~ BACK	109.47	UNDERSTAND ~~ EXACTLY	10.1	REGARD ~~ HOLISTICALLY	5.52
SEE ~~ GENERALLY	8.82	BRING ~~ TOGETHER	100.89	REACH ~~ APPROXIMATELY	6.2	REGARD ~~ AS	5.11
SEE ~~ HERE	8.23	BRING ~~ FORWARD	38.31	REACH ~~ QUICKLY	5.36	OBTAIN ~~ EXPERIMENTALLY	15.27
SEE ~~ EXACTLY	7.02	BRING ~~ FORTH	15.18	DIE ~~ BRAVELY	12.65	COMPARE ~~ FAVOURABLY	22.71
KNOW ~~ EXACTLY	47.91	SET ~~ APART	48.69	DIE ~~ AWAY	5.63	ACT ~~ MORALLY	19.53
KNOW ~~ WELL	8.96	SET ~~ FORTH	30.4	DIE ~~ QUICKLY	6.4	ACT ~~ AUTONOMOUSLY	14.43
KNOW ~~ ENOUGH	5.9	SET ~~ ASIDE	24.63	DRAW ~~ TOGETHER	13.9	ACT ~~ UNILATERALLY	11.66
TAKE ~~ SERIOUSLY	63.98	SET ~~ UP	22.72	DRAW ~~ BACK	13.17	ACT ~~ ACCORDINGLY	10.5
TAKE ~~ AWAY	53.16	SET ~~ SO	6.93	DRAW ~~ HEAVILY	8.19	ACT ~~ DIFFERENTLY	10.25
TAKE ~~ LONG	19.66	RUN ~~ AWAY	39.34	CREATE ~~ WHOLLY	6.36	ACT ~~ APPROPRIATELY	7.45
TAKE ~~ TOO	17.77	RUN ~~ SMOOTHLY	17.52	SPEND ~~ TOGETHER	6.26	ACT ~~ INDIVIDUALLY	6.58
TAKE ~~ LITERALLY	17.08	RUN ~~ PARALLEL	9.13	SPEND ~~ MUCH	6.1	ACT ~~ FAIRLY	6.11
TAKE ~~ MUCH	9.22	RUN ~~ UP	8.72	SPEND ~~ TWICE	5.62	ACT ~~ SOLELY	5.54
TAKE ~~ TOGETHER	8.32	RUN ~~ AGAIN	8.31	SPEND ~~ FAR	5.56	DEPEND ~~ HEAVILY	50.56
TAKE ~~ BACK	7.46	RUN ~~ EFFICIENTLY	6.9	WALK ~~ AWAY	16.54	DEPEND ~~ LARGELY	32.88
TAKE ~~ APPROXIMATELY	5.39	RUN ~~ CORRECTLY	6.6	WALK ~~ COMFORTABLY	6.06	DEPEND ~~ PRIMARILY	20.13
TAKE ~~ DIRECTLY	5.05	RUN ~~ NORMALLY	5.45	PASS ~~ AWAY	12.09	DEPEND ~~ ONLY	15.87
THINK ~~ ABOUT	24.9	RUN ~~ DOWN	5.44	PASS ~~ STRAIGHT	11.7	DEPEND ~~ ENTIRELY	15.59
THINK ~~ DIFFERENTLY	6	RUN ~~ CONTINUOUSLY	5.16	PASS ~~ DIRECTLY	8.08	DEPEND ~~ GREATLY	13.6
COME ~~ BACK	74.26	WRITE ~~ ABOUT	6.7	PASS ~~ DOWN	5.19	DEPEND ~~ CRUCIALLY	9.45
COME ~~ TOGETHER	68.83	WRITE ~~ PHONETICALLY	5.88	SELL ~~ DIRECTLY	21.15	DEPEND ~~ MUCH	7.65
COME ~~ NOWHERE	32.01	START ~~ AGAIN	12.03	BREAK ~~ DOWN	118.11	DEPEND ~~ PARTLY	6.7

COME ~~ FORWARD	21.11	START ~~ IMMEDIATELY	7.14	BREAK ~~ AWAY	37.19	DEPEND ~~ COMPLETELY	6.21
COME ~~ AWAY	7.24	START ~~ AFRESH	6.56	BREAK ~~ EVEN	19.7	DEPEND ~~ MAINLY	5.79
GIVE ~~ BELOW	13.59	PAY ~~ BACK	19.98	BREAK ~~ LOOSE	5.88	DEPEND ~~ INDIVIDUALLY	5.12
GIVE ~~ MUCH	9.43	PAY ~~ ANNUALLY	9.43	RETURN ~~ BACK	12.64	POINT ~~ OUT	26.53
GIVE ~~ RELATIVELY	7.12	LEAD ~~ ASTRAY	27.46	RETURN ~~ PREMATURELY	5.37	POINT ~~ HERE	5.92
GIVE ~~ VERY	5.94	LEAD ~~ BACK	8.79	ACCEPT ~~ EVEN	6.08	POINT ~~ FORWARD	5.7
USE ~~ EXTENSIVELY	34.35	MOVE ~~ AWAY	263.49	WAIT ~~ DESIRABLY	5.05	POINT ~~ THEN	5.65
USE ~~ INTERCHANGEABLY	24.3	MOVE ~~ FORWARD	83.8	RAISE ~~ LOCALLY	6.43	EXPRESS ~~ DEONTICALLY	7.36
USE ~~ HERE	22.18	MOVE ~~ BACK	18.03	CAUSE ~~ STILL	47.91	EXPRESS ~~ ALGEBRAICALLY	6.14
USE ~~ INSTEAD	20	MOVE ~~ FREELY	14.89	APPLY ~~ DIRECTLY	7.08	EXPRESS ~~ COHERENTLY	5.61
USE ~~ EFFECTIVELY	15.9	MOVE ~~ QUICKLY	12.59	APPLY ~~ HERE	6.69	EXPRESS ~~ INFORMALLY	5.5
USE ~~ FREQUENTLY	9.54	MOVE ~~ TOGETHER	11.97	APPLY ~~ ONLY	6.34	SUFFER ~~ GREATLY	10.9
USE ~~ SPARINGLY	8.54	MOVE ~~ APART	8.83	APPLY ~~ WORLDWIDE	6.01	SUFFER ~~ TERRIBLY	6.41
USE ~~ SUCCESSFULLY	7.12	MOVE ~~ BACKWARD	7.83	APPLY ~~BACKWARD	5.12	SUFFER ~~ DISPROPORTIONATELY	5.77
USE ~~ ONLY	6.39	MOVE ~~ INDEPENDENTLY	6.74	APPLY ~~ BROADLY	5.12	SUFFER ~~ REASONABLY	5.72
USE ~~ APPROPRIATELY	5.77	MOVE ~~ RAPIDLY	5.53	LIE ~~ SOMEWHERE	16.18	SUFFER ~~ CHEMICALLY	5.38
USE ~~ COMMERCIALLY	5.69	PLAY ~~ VERY	13.04	LIE ~~ COMFORTABLY	5.01	VISIT ~~ REGULARLY	8.9
USE ~~ EXCLUSIVELY	5.32	PLAY ~~ QUIETLY	5.27	REDUCE ~~ CONSIDERABLY	11.56	VISIT ~~ FREQUENTLY	5.06
USE ~~ CLINICALLY	5.28	LIVE ~~ TOGETHER	71.67	REDUCE ~~ SHARPLY	8.55	DESIGN ~~ SPECIFICALLY	7.31
USE ~~ INTENSIVELY	5.03	LIVE ~~ ALONE	55	BASE ~~ SOLELY	13.67	TREAT ~~ DIFFERENTLY	24.51
LOOK ~~ BACK	61.97	LIVE ~~ LOCALLY	22.89	BASE ~~ ENTIRELY	13.1	TREAT ~~ HARSHLY	22.1
LOOK ~~ CLOSELY	35.61	LIVE ~~ NEARBY	14.42	BASE ~~ PRIMARILY	13.07	TREAT ~~ UNFAIRLY	17.15

LOOK ~~ FORWARD	35.27	LIVE ~~ INDEPENDENTLY	14.39	BASE ~~ PURELY	9.72	TREAT ~~ FAIRLY	13.06
LOOK ~~ BRIEFLY	12.12	LIVE ~~ WELL	6.27	BASE ~~ LOOSELY	5.9	TREAT ~~ EQUALLY	11.39
LOOK ~~ ELSEWHERE	10.85	LIVE ~~ HERE	5.32	REPORT ~~ FIRSTLY	6.49	TREAT ~~ BADLY	9.51
LOOK ~~ CAREFULLY	9.13	LET ~~ ALONE	107.55	CONCERN ~~ HERE	6.04	TREAT ~~ JUSTLY	6.03
LOOK ~~ AWAY	7.61	LET ~~ NOW	47.69	CONCERN ~~ ONLY	5.95	TREAT ~~ IDENTICALLY	5.97
FIND ~~ ELSEWHERE	18.94	LET ~~ JUST	6.14	CONCERN ~~ SOLELY	5.16	FINISH ~~ DIFFERENTLY	24.51
FIND ~~ VERY	15.53	HEAR ~~ CLEARLY	5.88	EXPLAIN ~~ AWAY	8.52	FINISH ~~ HARSHLY	22.1
FIND ~~ EXPERIMENTALLY	13.61	BELIEVE ~~ FIRMLY	6.75	EXPLAIN ~~ PREVIOUSLY	6.1	FINISH ~~ UNFAIRLY	17.15
FIND ~~ ONLY	10.19	INVOLVE ~~ GENETICALLY	7.12	EXPLAIN ~~ BELOW	5.6	FINISH ~~ FAIRLY	13.06
FIND ~~ HERE	6.28	HAPPEN ~~ AGAIN	10.95	CONTAIN ~~ PURELY	7.85	FINISH ~~ EQUALLY	11.39
FIND ~~ INCREASEDLY	6.03	HAPPEN ~~ QUICKLY	6.6	CONTAIN ~~ APPROXIMATELY	7.22	FINISH ~~ BADLY	9.51
WANT ~~ SUPPOSEDLY	6.16	REMAIN ~~ LARGELY	40.26	CONTAIN ~~ ONLY	5.81	FINISH ~~ JUSTLY	6.03
TELL ~~ EXACTLY	6.31	REMAIN ~~ RELATIVELY	35.96	STAY ~~ AHEAD	19.11	FINISH ~~ IDENTICALLY	5.97
PUT ~~ FORWARD	583.56	REMAIN ~~ FAIRLY	13.37	STAY ~~ OVERNIGHT	12.37	REFLECT ~~ BACK	23.86
PUT ~~ FORTH	114.6	REMAIN ~~ ALMOST	13.21	STAY ~~ AWAY	10.07	ADMIT ~~ INFORMALLY	6.63
PUT ~~ TOGETHER	79.99	REMAIN ~~ EVEN	10.32	STAY ~~ TOGETHER	7.37	THROW ~~ AWAY	11.28
PUT ~~ SIMPLY	21.26	REMAIN ~~ ESSENTIALLY	8.01	STAY ~~ INDOORS	6.29	THROW ~~ BACK	5.42
PUT ~~ DOWN	11.17	REMAIN ~~ FIRMLY	6.99	STAY ~~ THERE	5.31		
PUT ~~ BACK	8.53	REMAIN ~~ VERY	6.89	STAY ~~ FAIRLY	5.06		
PUT ~~ ASIDE	7.52	REMAIN ~~ HIGHLY	6.72	ESTABLISH ~~ WHOLLY	5.44		

Table A-18: The normalised frequencies of *Adjectives* in the Arab EFLLs corpus

Adjectives	NF	Adjectives	NF	Adjectives	NF	Adjectives	NF	Adjectives	NF	Adjectives	NF	Adjectives	NF
"many"	3084	"young"	390	"low"	214	"economic"	148	"close"	109	"clean"	83	"worth"	66
"good"	2822	"long"	386	"female"	214	"effective"	143	"go"	106	"entire"	83	"powerful"	66
"other"	2051	"hard"	379	"positive"	204	"warm"	142	"cultural"	105	"overall"	82	"live"	65
"bad"	1656	"easy"	372	"least"	200	"green"	142	"essential"	104	"interested"	81	"wild"	65
"more"	1293	"second"	371	"english"	200	"past"	142	"significant"	102	"exciting"	81	"beatiful"	65
"such"	1104	"due"	367	"hot"	191	"short"	142	"mobile"	102	"current"	80	"extra"	62
"different"	1082	"few"	338	"common"	190	"comfortable"	140	"emotional"	101	"rich"	78	"like"	62
"last"	1018	"huge"	331	"less"	187	"kuwaiti"	138	"afraid"	101	"unique"	77	"terrible"	62
"first"	997	"personal"	315	"clear"	182	"possible"	138	"likely"	101	"angry"	76	"helpful"	62
"happy"	988	"online"	305	"natural"	181	"aware"	133	"friendly"	100	"cheap"	76	"mental"	62
"big"	874	"real"	304	"very"	180	"medical"	130	"lovely"	100	"technological"	75	"married"	62
"important"	850	"whole"	301	"third"	177	"sick"	130	"popular"	99	"professional"	75	"sexual"	62
"new"	822	"strong"	292	"serious"	177	"normal"	129	"american"	97	"international"	75	"strict"	62
"most"	775	"sad"	291	"poor"	177	"enough"	126	"available"	97	"religious"	75	"marine"	61
"nice"	746	"successful"	286	"useful"	173	"special"	126	"violent"	94	"civil"	73	"hungry"	61
"plastic"	735	"political"	280	"logical"	173	"responsible"	126	"single"	94	"animal"	73	"wide"	61
"same"	670	"famous"	272	"fast"	172	"modern"	126	"valuable"	92	"beneficial"	72	"favorite"	59
"high"	654	"only"	269	"specific"	171	"ready"	121	"funny"	91	"general"	72	"suitable"	59
"great"	560	"major"	261	"interesting"	171	"public"	121	"active"	90	"private"	71	"financial"	59
"beautiful"	543	"negative"	256	"dependent"	170	"early"	121	"future"	90	"white"	71	"recent"	59
"human"	509	"full"	245	"smart"	168	"tired"	116	"traditional"	89	"cool"	71	"dead"	58
"much"	497	"several"	238	"amazing"	167	"grand"	116	"weak"	86	"renewable"	70	"commercial"	58
"old"	491	"true"	233	"simple"	166	"various"	115	"scientific"	86	"arabian"	70	"enjoyable"	58
"next"	468	"large"	231	"male"	164	"everyday"	115	"electronic"	86	"fresh"	70	"boring"	67
"own"	447	"certain"	231	"physical"	164	"final"	113	"basic"	86	"academic"	70	"colorful"	83
"social"	444	"little"	223	"video"	162	"late"	113	"creative"	85	"necessary"	68	"similar"	109
"main"	439	"free"	219	"cold"	156	"sure"	113	"black"	83	"top"	68	"favourite"	154
"small"	411	"right"	218	"safe"	154	"difficult"	110	"particular"	83	"educational"	67	"wrong"	218
"able"	390												

Table A-19: The normalised frequencies of *Adjectives* in the NBESs Corpus

Adjectives	NF	Adjectives	NF	Adjectives	NF	Adjectives	NF	Adjectives	NF	Adjectives	NF	Adjectives	NF
other	1,239.04	long	250.35	total	183.90	relevant	136.27	responsible	95.13	military	68.74	fresh	44.38
high	969.14	strong	246.75	old	182.46	easy	134.83	correct	95.13	fair	67.90	top	43.90
different	944.07	local	242.67	central	181.38	normal	134.47	additional	94.89	professional	66.10	immediate	43.42
social	930.15	general	242.43	personal	172.62	initial	133.87	civil	94.05	extra	65.26	safe	42.11
new	914.68	clear	241.00	wide	168.54	female	132.67	crucial	93.21	regular	63.82	little	41.63
good	737.26	american	234.64	direct	168.42	equal	131.23	active	91.41	broad	62.86	blue	39.59
important	657.97	specific	234.52	effective	167.46	full	130.39	bad	91.29	unlikely	62.26	dangerous	39.59
international	568.84	various	234.04	late	165.30	private	128.95	sexual	90.81	cold	60.70	quick	38.99
great	567.28	financial	233.08	successful	164.58	essential	128.23	aware	88.17	detailed	59.62	official	37.91
low	547.01	necessary	232.96	traditional	162.30	standard	125.60	regional	87.93	green	56.50	fine	37.67
large	517.38	natural	231.04	english	161.58	close	122.48	sufficient	86.01	wrong	56.50	happy	34.67
political	515.82	public	228.16	useful	161.10	male	119.84	annual	84.93	considerable	56.50	proper	34.07
economic	514.62	difficult	225.76	foreign	160.98	actual	119.24	special	84.45	technical	55.30	flat	33.11
human	510.78	positive	219.40	final	159.90	whole	119.00	apparent	80.25	deep	53.26	familiar	31.67

possible	488.11	individual	214.60	previous	157.86	practical	116.60	weak	79.65	cheap	52.66	realistic	31.55
small	417.09	physical	214.36	short	156.43	big	115.28	commercial	79.41	sure	49.30	mathematical	31.43
able	376.31	free	212.09	appropriate	155.59	primary	114.80	relative	78.93	interested	48.94	marginal	31.43
main	354.84	particular	211.73	potential	155.11	independent	114.32	formal	77.13	nuclear	48.58	racial	31.43
available	349.44	cultural	211.49	black	149.35	original	113.12	obvious	76.41	perfect	48.34	fat	31.19
significant	321.85	present	209.33	right	145.03	white	112.88	rural	76.29	heavy	47.62	excellent	30.59
similar	304.09	major	207.65	only	144.67	past	110.12	impossible	75.33	joint	47.14	hot	29.75
certain	300.73	real	203.09	complex	144.55	average	109.28	powerful	73.29	substantial	46.42	thermal	29.51
european	299.29	legal	198.05	recent	144.19	popular	109.04	reasonable	72.21	soft	45.82	sweet	29.51
common	287.78	environmental	195.53	medical	142.27	religious	107.72	unable	71.49	dark	45.82	analytical	29.51
likely	283.70	key	190.73	basic	142.03	hard	106.04	rich	70.78	valid	45.22	narrow	29.27
early	274.82	TRUE	190.13	domestic	139.63	complete	100.88	separate	70.66	dead	45.10	warm	25.67
national	274.70	simple	188.45	future	138.31	overall	98.61	light	70.54	due	44.74		
current	267.99	single	187.61	young	137.71	suitable	97.17	huge	69.34	expensive	44.50		
modern	264.39	poor	184.02	open	137.35	interesting	96.81	serious	69.22	slow	44.38		

Table A-20: Extracted Adjective-Noun Collocations with LL scores for the Arab EFLLs corpus

Adjective- Noun collocations	LL	Adjective-Noun collocations	LL	Adjective - Noun collocations	LL	Adjective-Noun collocations	LL
MANY ~~ THING	198	HIGH ~~ RATE	12.14	NEGATIVE ~~ SIDE	11.47	MODERN ~~ WORLD	6.53
MANY ~~ PEOPLE	127.32	HIGH ~~ COST	9.64	NEGATIVE ~~ WAY	9.42	MODERN ~~ AGE	6.19
MANY ~~ REASON	56.98	HIGH ~~ PRICE	7.48	NEGATIVE ~~ ASPECT	8.4	MODERN ~~ LIFE	6.16
MANY ~~ PLACE	47.02	HIGH ~~ EXERCISE	6.93	NEGATIVE ~~ INFLUENCE	8.4	MODERN ~~ TIME	5.44
MANY ~~ WAY	35.24	GREAT ~~ TIME	34.91	NEGATIVE ~~ BEHAVIOR*	7.8	PUBLIC ~~ SCHOOL	52.99
MANY ~~ PROBLEM	34.67	GREAT ~~ HOLIDAY	9.66	FULL ~~ MARK	19.01	PUBLIC ~~ TRANSPORT	32.77
MANY ~~ ANIMAL	23.99	GREAT ~~ INVENTION	7.57	SEVERAL ~~ REASON	17.04	PUBLIC ~~ UNIVERSITY	22.65
MANY ~~ ADVANTAGE	21.36	GREAT ~~ VALUE	6.51	TRUE ~~ MEANING	5.23	PUBLIC ~~ ARABIC	9.69
MANY ~~ DISADVANTAGE	16.29	GREAT ~~ DEAL	5.79	LARGE ~~ AMOUNT	46.06	PUBLIC ~~ EYE	9.51
MANY ~~ DISEASE	14.7	GREAT ~~ ROLE	5.23	LARGE ~~ NUMBER	42.36	PUBLIC ~~ PLACE	7.57
MANY ~~ KIND	14.22	GREAT ~~ DEBT	5.01	LARGE ~~ QUANTITY	20.45	PUBLIC ~~ PROPERTY	5.76
MANY ~~ YEAR	13.04	BEAUTIFUL ~~ PLACE	128.37	LARGE ~~ RANGE	7.88	PUBLIC ~~ SERVICE	5.01
MANY ~~ GAME	12.84	BEAUTIFUL ~~ THING	6.85	CERTAIN ~~ TYPE	13.96	EARLY ~~ AGE	53.64
MANY ~~ BENEFIT	11.61	BEAUTIFUL ~~ CITY	5.99	CERTAIN ~~ CONDITION	10.86	EARLY ~~ STAGE	16.74
MANY ~~ COUNTRY	11.26	BEAUTIFUL ~~ COUNTRY	5.37	CERTAIN ~~ BUDGET	6.25	EARLY ~~ MARRIAGE	12.99
MANY ~~ CASE	8.76	HUMAN ~~ BEING	51.28	CERTAIN ~~ CIRCUMSTANCE	6.11	EARLY ~~ MORNING	11.08

MANY ~~ DIFFERENCE	8.48	HUMAN ~~ RIGHT	30.74	CERTAIN ~~ LEVEL	6.05	EARLY ~~ PREGNANCY	6.06
MANY ~~ PLASE*	8.08	HUMAN ~~ BODY	21.37	CERTAIN ~~ CATEGORY	5.79	EARLY ~~ DEATH	5.81
MANY ~~ SONG	7.87	HUMAN ~~ ACTIVITY	20.29	LITTLE ~~ BROTHER	23.42	GRAND ~~ MOTHER NOT	93.72
MANY ~~ FIELD	7.54	HUMAN ~~ HEALTH	13.91	LITTLE ~~ SISTER	20.01	GRAND ~~ FATHER	52.33
MANY ~~ PICTURE	6.97	HUMAN ~~ QUALITY	12.45	LITTLE ~~ BIT	15.81	GRAND ~~ PARENT NOT	16.05
MANY ~~ INFORMATION	6.49	HUMAN ~~ LIFE	10.88	LITTLE ~~ GIRL	10.62	GRAND ~~ MATHER*	12.18
MANY ~~ TYPE	5.77	HUMAN ~~ RACE	10.47	LITTLE ~~ DETAIL	5.42	GRAND ~~ HOUSE	7.55
GOOD ~~ FILM	60.67	HUMAN ~~ MIND	6.59	FREE ~~ TIME	34.81	VARIOUS ~~ WAY	9.47
GOOD ~~ WAY	49	MUCH ~~ MONEY	40.62	FREE ~~ SPEECH	31.49	VARIOUS ~~ TYPE	6.98
GOOD ~~ HOLIDAY	41.72	MUCH ~~ FUN	11.74	FREE ~~ BAG	8.53	VARIOUS ~~ FIELD	5.41
GOOD ~~ FRIEND	39.81	MUCH ~~ TIME	11.08	FREE ~~ CHARGE	7.92	EVERYDAY ~~ LIFE	24.25
GOOD ~~ THING	26.01	MUCH ~~ DEPENDENCE	10.92	RIGHT ~~ SELF-EXPRESSION 0	12.04	FINAL ~~ EXAM	16.99
GOOD ~~ SOLUTION	18.69	MUCH ~~ COMPUTER	6.44	RIGHT ~~ THING	10.54	FINAL ~~ DAY	7.06
GOOD ~~ CHANCE	15.17	MUCH ~~ INFORMATION	6.18	RIGHT ~~ WAY	9.04	CLOSE ~~ DEBT	13.75
GOOD ~~ IDEA	15.58	OLD ~~ AGE	11.66	RIGHT ~~ ACTIVIST	5.3	CLOSE ~~ EYE	8.86
GOOD ~~ PLACE	12.6	OLD ~~ THING	7.32	WRONG ~~ POSTURE0	21.4	CLOSE ~~ FRIEND	5.38
GOOD ~~ MOVIE	11.34	OLD ~~ SYSTEM	6.78	LOW ~~ DIET0	18.26	ESSENTIAL ~~ PART	10.88
GOOD ~~ TIME	9.97	OLD ~~ GIRL	5.95	LOW ~~ INTENSITY	15.87	MOBILE ~~ PHONE	82.02
GOOD ~~ HABIT	9.18	OLD ~~ FORT	5.85	LOW ~~ BLOOD	14.31	EMOTIONAL ~~ GENERATION0	5.17

GOOD ~~ CHOICE	9.12	OLD ~~ LADY	5.38	LOW ~~ CARB 0	12.66	FRIENDLY ~~ MATERIAL	13.92
GOOD ~~ JOB	8.31	OLD ~~ DAY	5.07	LOW ~~ INCOME	12.41	FRIENDLY ~~ MATCH	6.4
GOOD ~~ SERVICE	8.28	NEXT ~~ DAY	175.74	LOW ~~ PRICE	9.79	LOVELY ~~ COUNTRY	8.66
GOOD ~~ OPTION	5.99	NEXT ~~ HOLIDAY	21.11	LOW ~~ RATE	9.46	LOVELY ~~ TIME	5.25
GOOD ~~ BEHAVIOR*	5.53	NEXT ~~ SUMMER	16.47	LOW ~~ PRESSURE	7.84	VIOLENT ~~ VIDEO 3B	32.95
OTHER ~~ HAND	264.3	NEXT ~~ FLIGHT	9.79	LOW ~~ POPULATION	7.49	VIOLENT ~~ GAME	30.28
OTHER ~~ PEOPLE	50.07	NEXT ~~ YEAR	8.93	LOW ~~ SUGAR	6.13	VIOLENT ~~ CONTENT	13.3
OTHER ~~ WORD	48.48	NEXT ~~ TIME	8.36	LOW ~~ SALARY	5.84	VIOLENT ~~ ACT	9.96
OTHER ~~ SIDE	27.16	NEXT ~~ GENERATION	7.27	LOW ~~ TRAINING	5.84	VIOLENT ~~ ROBBERY	7.42
OTHER ~~ THING	19.09	NEXT ~~ WEEK	5.7	LOW ~~ GRADE	5.49	VIOLENT ~~ VIDEOGAMES	6.2
OTHER ~~ COUNTRY	14.29	OWN ~~ HAND	7.66	LOW ~~ RECYCLING	5.33	VIOLENT ~~ MEDIA	5.89
OTHER ~~ DEVICE	12.69	OWN ~~ OPINION	7.48	LOW ~~ CARBOHYDRATE	5.24	SINGLE ~~ PARENT	67.34
OTHER ~~ WAY	9.43	OWN ~~ CHOICE	5.54	LOW ~~ CALORIE	5.07	SINGLE ~~ ADOPTION0	14.44
OTHER ~~ KIND	7.9	OWN ~~ CAR	5.47	FEMALE ~~ DRIVER	98.52	VALUABLE ~~ TIME	6.77
OTHER ~~ ALTERNATIVE	7.86	SOCIAL ~~ MEDIUM 00	153.67	FEMALE ~~ JUDGE	7.47	ACTIVE ~~ RECYCLING	6.33
OTHER ~~ ACTIVITY	6.91	SOCIAL ~~ NETWORKING	32.79	POSITIVE ~~ EFFECT	39.54	FUTURE ~~ GENERATION	10.7
OTHER ~~ ANIMAL	6.37	SOCIAL ~~ NETWORK	31.84	POSITIVE ~~ IMPACT	16.19	FUTURE ~~ JOB	9.48
OTHER ~~ MATERIAL	6.13	SOCIAL ~~ LIFE	31.19	POSITIVE ~~ INFLUENCE	9.44	FUTURE ~~ CAREER	7.03
BAD ~~ HOLIDAY	428.44	SOCIAL ~~ INTERACTION	30.84	POSITIVE ~~ ASPECT	9.44	FUTURE ~~ REFERENCE	6.68

BAD ~~ BREATH	13.24	SOCIAL ~~ SITE	23.61	POSITIVE ~~ SELF-ESTEEM	6.06	TRADITIONAL ~~ LEARNING	19.31
BAD ~~ EFFECT	35.09	SOCIAL ~~ APPLICATION	14.35	POSITIVE ~~ SIDE	5.44	TRADITIONAL ~~ CUSTOM	19.12
BAD ~~ HABIT	22.95	SOCIAL ~~ BEHAVIOR*	14.25	POSITIVE ~~ OUTCOME	5.3	TRADITIONAL ~~ METHOD	7.79
BAD ~~ THING	17.82	SOCIAL ~~ SKILL	14.16	LEAST ~~ RESULT	12.16	TRADITIONAL ~~ MARRIAGE	6.46
BAD ~~ WEEKEND	14.54	SOCIAL ~~ CIRCLE	12	LEAST ~~ HOUR	7.99	TRADITIONAL ~~ MEDICINE	5.23
BAD ~~ HOLIDY*	9.1	SOCIAL ~~ ISOLATION	10.27	HOT ~~ SUMMER	6.25	SCIENTIFIC ~~ RESEARCH	5.19
BAD ~~ LUCK	8.44	SOCIAL ~~ PRESSURE	8.56	COMMON ~~ ARGUMENT	6.88	ELECTRONIC ~~ BOOK	62.14
BAD ~~ VACATION 0	8.21	SOCIAL ~~ PROBLEM	6.86	LESS ~~ FAT	19.13	ELECTRONIC ~~ DEVICE	38.64
BAD ~~ NEWS	13.05	MAIN ~~ REASON	42.18	LESS ~~ ENERGY	7.66	ELECTRONIC ~~ COMMUNICATION	11.95
MORE ~~ TIME	11.77	MAIN ~~ SOURCE	19.39	LESS ~~ TIME	6.84	BASIC ~~ NEED	12.21
MORE ~~ FRUIT	9.73	MAIN ~~ CAUSE	17.61	LESS~~ WASTE	5.14	BASIC ~~ NECESSITY	7.76
MORE ~~ POWER	9.72	MAIN ~~ FACTOR	15.25	NATURAL ~~ GAS	20.38	BLACK ~~ MARKET	18.44
MORE ~~ HARM	8.79	MAIN ~~ POINT	10.04	NATURAL ~~ RESOURCE	13.76	PARTICULAR ~~ SUBJECT	5.93
MORE ~~ RESULT	8.65	MAIN ~~ AIM	8.09	NATURAL ~~ DISASTER	10.21	CLEAN ~~ HOUSE	7.67
MORE ~~ POLLUTION	7.66	MAIN ~~ PURPOSE	6.89	NATURAL ~~ HABITAT	10.21	CLEAN ~~ ENERGY	5.55
MORE ~~ PEOPLE	6.05	MAIN ~~ PRIORITY	5.65	NATURAL ~~ ECOSYSTEM	6.33	CLEAN ~~ MEDICINE 0	5.13
MORE ~~ ENERGY	5.88	MAIN ~~ TYPE	5.39	NATURAL ~~ VIEW 0	5.27	ENTIRE ~~ LIFE	5.12

MORE ~~ WATER	5.77	MAIN ~~ ROLE	5.12	NATURAL ~~ ENVIRONMENT	5.26	OVERALL ~~ QUANTITY	6.99
MORE ~~ FUN	5.69	YOUNG ~~ AGE	35.85	NATURAL ~~ BEAUTY	5	OVERALL ~~ PERFORMANCE	6.05
SUCH ~~ THING	20.43	YOUNG ~~ GENERATION	19.14	THIRD ~~ REASON	23.22	CURRENT ~~ GENERATION	14.53
SUCH ~~ CANCER	12	YOUNG ~~ CHILD	18.98	THIRD ~~ DAY	8.79	CURRENT ~~ SITUATION	8.54
SUCH ~~ SITUATION	10.12	YOUNG ~~ PEOPLE	17.26	THIRD ~~ PARTY	5.94	CURRENT ~~ CIRCUMSTANCE	7.17
DIFFERENT ~~ TYPE	32.6	YOUNG ~~ ADULT	13.04	SERIOUS ~~ PROBLEM	47.87	CURRENT ~~ STANDARD	6.16
DIFFERENT ~~ KIND	23.11	YOUNG ~~ BROTHER	7.43	SERIOUS ~~ RESULT	15.62	RICH ~~ PEOPLE	13.85
DIFFERENT ~~ NUTRIENT	20.84	YOUNG ~~ SISTER	7.13	SERIOUS ~~ ISSUE	10.01	RICH ~~ COUNTRY	5.82
DIFFERENT ~~ WAY	14.97	LONG ~~ TIME	97.13	SERIOUS ~~ THREAT	8.58	UNIQUE ~~ FEATURE	6.81
DIFFERENT ~~ COUNTRY	9.26	LONG ~~ PERIOD	50.79	SERIOUS ~~ DANGER	5.09	TECHNOLOGICAL ~~ DEVICE 0	41.98
DIFFERENT ~~ STYLE	8.34	LONG ~~ TERM	24.48	POOR ~~ PEOPLE	27.28	TECHNOLOGICAL ~~ ADVANCEMENT	14.42
DIFFERENT ~~ PLACE	6.99	LONG ~~ RUN	21.5	POOR ~~ POSTURE	9.65	TECHNOLOGICAL ~~ TOOL	7.82
DIFFERENT ~~ THING	5.98	LONG ~~ HOUR	15.91	POOR ~~ GRADE0	5.44	TECHNOLOGICAL ~~ ADVANCE	5.99
LAST ~~ SUMMER	313.09	LONG ~~ DISTANCE	9.94	USEFUL ~~ THING	14.16	PROFESSIONAL ~~ GAMING	15.89
LAST ~~ HOLIDAY	132.33	LONG ~~ TRIP	6.88	USEFUL ~~ ACTIVITY	5.3	PROFESSIONAL ~~ GAMERS	5.72
LAST ~~ YEAR	125.47	HARD ~~ DRIVE	17.03	FAST ~~ FOOD	145.41	PROFESSIONAL ~~ PLAYER	5.49

LAST ~~ WEEK	55.11	EASY ~~ WAY	31.25	FAST ~~ RESTAURANT	18.73	INTERNATIONAL ~~ STANDARD	6.82
LAST ~~ WEEKEND	47.39	EASY ~~ ACCESS	16.45	FAST ~~ COMPANY	14.11	INTERNATIONAL ~~ AIRPORT	5.85
LAST ~~ DAY	15.14	EASY ~~ VICTIM	6.64	FAST ~~ SHOP	8.48	INTERNATIONAL ~~ LAW	5.69
LAST ~~ MONTH	12.68	SECOND ~~ HAND	20.19	FAST ~~ INDUSTRY	7.12	RELIGIOUS ~~ PARTY	10.24
LAST ~~ VACATION	7.26	SECOND ~~ REASON	16.84	FAST ~~ CAR	6.85	CIVIL ~~ SERVICE	12.67
FIRST ~~ TIME	55.21	SECOND ~~ DAY	15.23	SPECIFIC ~~ TYPE	15.66	ANIMAL ~~ TESTING	43.34
FIRST ~~ THING	23.43	SECOND ~~ SMOKING	7.59	SPECIFIC ~~ SKILL	7.25	ANIMAL ~~ PET	21.06
FIRST ~~ PLACE	22.83	SECOND ~~ SMOKER	6.54	SPECIFIC ~~ GOAL	5.65	ANIMAL ~~ EXPERIMENTATION	15.86
FIRST ~~ DAY	20.73	FEW ~~ MINUTE	29.09	DEPENDENT ~~ COMPUTER	121.68	ANIMAL ~~ RIGHT	6.46
FIRST ~~ WEEK	12.73	FEW ~~ YEAR	15.8	DEPENDENT ~~ TECHNOLOGY	13.06	ANIMAL ~~ EXPERIMENT	6.28
FIRST ~~ STEP	9.62	FEW ~~ DAY	11.19	SMART ~~ PHONE	161.46	ANIMAL ~~ SPECIE	6.19
FIRST ~~ SIDE	8.89	FEW ~~ HOUR	7.23	SMART ~~ DEVICE	43.26	GENERAL ~~ STORY	5.43
FIRST ~~ YEAR	6.53	FEW ~~ LINE	6.18	SMART ~~ CAR	9.95	PRIVATE ~~ SCHOOL	55.25
BIG ~~ BROTHER	18.88	FEW ~~ DECADE	5.93	SMART ~~ DRIVING	8.33	PRIVATE ~~ UNIVERSITY	22.78
BIG ~~ PROBLEM	16.28	FEW ~~ PEOPLE	5.23	SMART ~~ DETRIMENT	8.31	PRIVATE ~~ INSTITUTES	7.73
BIG ~~ HOTEL	14.97	HUGE ~~ AMOUNT	28.42	AMAZING ~~ COUNTRY	5.78	PRIVATE ~~ INFORMATION	6.74
BIG ~~ PICTURE	10.85	HUGE ~~ IMPACT	24.06	SIMPLE ~~ THING	8.96	PRIVATE ~~ SECTOR	6.61
BIG ~~ CAR	10.74	HUGE ~~ IMPROVEMENT	8.32	SIMPLE ~~ STEP	5.15	WHITE ~~ BLOOD	13.28

BIG ~~ HOUSE	10.04	HUGE ~~ INFLUENCE	6.34	MALE ~~ DRIVER	42.67	WHITE ~~ CELL	12.45
BIG ~~ TREE	9.25	HUGE ~~ ROLE	5.98	MALE ~~ FEMALE	20.73	WHITE ~~ MATTER	8.44
BIG ~~ ROLE	8.5	HUGE ~~ HEADACHE	5.61	PHYSICAL ~~ ACTIVITY	52.85	RENEWABLE ~~ ENERGY	124.83
BIG ~~ CITY	7.67	HUGE ~~ NUMBER	5.6	PHYSICAL ~~ SCIENCE	28.93	RENEWABLE ~~ SOURCE	28.93
BIG ~~ DIFFERENCE	7.57	HUGE ~~ PART	5.34	PHYSICAL ~~ EFFECT	12.74	RENEWABLE ~~ RESOURCE	17.71
BIG ~~ PARK	7.01	PERSONAL ~~EXPERIENCE	27.55	PHYSICAL ~~ PROBLEM	11.58	RENEWABLE ~~ GENERATOR	12.24
BIG ~~ SISTER	5.96	PERSONAL ~~ COMPUTER	21.1	PHYSICAL ~~ HEALTH	9.28	RENEWABLE ~~ POWER	7.18
IMPORTANT ~~ THING	45.69	PERSONAL ~~ OPINION	14.1	PHYSICAL ~~ EXERCISE	8.38	FRESH ~~ AIR	42.18
IMPORTANT ~~ ASPECT	12.54	PERSONAL ~~ CHOICE	11.9	PHYSICAL ~~ FEATURE	7.42	FRESH ~~ FOOD	34.3
IMPORTANT ~~ ROLE	11.8	PERSONAL ~~ LIFE	11.79	PHYSICAL ~~ DISABILITY	6.16	FRESH ~~ MEAT	9.24
IMPORTANT ~~ FACTOR	9.93	PERSONAL ~~ FREEDOM	9.63	PHYSICAL ~~ DECLINE	5.24	ACADEMIC ~~ PERFORMANCE	44.94
IMPORTANT ~~ PART	9.72	PERSONAL ~~ REVOLUTION	5.48	PHYSICAL ~~ APPEARANCE	5.24	ACADEMIC ~~ SUCCESS	5.66
IMPORTANT ~~ INVENTION	9.7	PERSONAL ~~ CAPACITY	5.32	VIDEO ~~ GAME	380.21	TOP ~~ PRIORITY	6.81
IMPORTANT ~~ POINT	9.51	PERSONAL ~~ INTERVIEW	5.32	VIDEO ~~ PLAYER	9.49	BORING ~~ FILM	6
IMPORTANT ~~ MAKER	5.96	ONLINE ~~ SHOPPING	25.74	VIDEO ~~ GAMING	6.46	LIVE ~~ LIFE	13.38
NEW ~~ SYSTEM	28.43	ONLINE ~~ SHOP BAWE	25.4	VIDEO ~~ MAKING	5.94	WILD ~~ ANIMAL	91.23
NEW ~~ VOTING	21.48	ONLINE ~~ SHOPPER	16.45	VIDEO ~~ ADDICTION	5.03	WILD ~~ LIFE	7.24

NEW ~~ GENERATION	21.4	ONLINE ~~ SITE	8.8	SAFE ~~ WAY	6.59	BEATIFUL ~~ PLACE*	17.38
NEW ~~ TECHNOLOGY	15.94	ONLINE ~~ LIBRARY	7.76	SAFE ~~ ENVIRONMENT	5.63	BEATIFUL ~~ COUNTRY*	8.61
NEW ~~ FRIEND	15.03	ONLINE ~~ DICTIONARY	5.56	FAVOURITE ~~ POSSESSION	118.94	LIKE ~~ FILM	16.11
NEW ~~ RESTAURANT	13.75	ONLINE ~~ DEVICE	5.42	FAVOURITE ~~ SONG	13.05	LIKE ~~ FOOD	10.99
NEW ~~ THING	10.41	ONLINE ~~ COURSE	5.37	ECONOMIC ~~ DEVELOPMENT	12.29	LIKE ~~ MOVIE	9.86
NEW ~~ INVENTION	9.73	ONLINE ~~ WEBSITES	5.2	ECONOMIC ~~ GROWTH	9.54	LIKE ~~ LION	8.52
NEW ~~ DEVICE	6.24	REAL ~~ WORLD	99.15	ECONOMIC ~~ BENEFIT	9.43	LIKE ~~ PARK	8.34
MOST ~~ PLACE	59.71	REAL ~~ LIFE	51.03	EFFECTIVE ~~ METHOD	10.97	LIKE ~~ HOLIDAY	7.52
MOST ~~ THING	48.01	REAL ~~ TRUTH	23.44	GREEN ~~ HOUSE	24.81	TERRIBLE ~~ RESULT	10.32
MOST ~~ PEOPLE	46.16	REAL ~~ ISSUE	10.18	GREEN ~~ TREE	22.9	TERRIBLE ~~ ACCIDENT	5.01
MOST ~~ CASE	11.41	REAL ~~ CONVERSATION	7.68	GREEN ~~ AREA	21.57	MENTAL ~~ HEALTH	11.86
MOST ~~ DWELLER /	9.56	WHOLE ~~ WORLD	20.08	GREEN ~~ EMISSION	12.21	MENTAL ~~ PROBLEM	9.98
MOST ~~ COUNTRY	8.48	WHOLE ~~ FAMILY	12.51	GREEN ~~ LAND	11.28	MENTAL ~~ ILLNESS	8.76
MOST ~~ INVENTION	7.93	WHOLE ~~ DAY	8.99	GREEN ~~ TEA	11.07	MENTAL ~~ ISSUE	8.65
MOST ~~ RESULT	7.44	WHOLE ~~ LIFE	8.54	GREEN ~~ PARK	9.08	MENTAL ~~ CAPACITY	6.32
NICE ~~ PLACE	45.52	STRONG ~~ BODY	11.03	GREEN ~~ GRASS	7.11	MARRIED ~~ PARENTS	6.05
NICE ~~ TIME	33.69	STRONG ~~ EVIDENCE	10.42	GREEN ~~ HILL	7.11	SEXUAL ~~ EXPLOITATION	10.02
NICE ~~ DAY	11.29	STRONG ~~ SYSTEM	5.79	GREEN ~~ MOUNTIN*	6.25	SEXUAL ~~ GESTURE	10.02

NICE ~~ HOLIDAY	10.24	STRONG ~~ STRUCTURE	5.38	GREEN ~~ GAS	5.76	SEXUAL ~~ ORIENTATION	10.02
NICE ~~ COUNTRY	9.21	SUCCESSFUL ~~ MARRIAGE	7.04	GREEN ~~ PLANT	5.3	SEXUAL ~~ ATTRACTION	8.11
NICE ~~ WEEKEND	7.91	SUCCESSFUL ~~ LIFE	5.83	PAST ~~ YEAR	21.33	SEXUAL ~~ CONTENT	7.07
NICE ~~ WEATHER	6.81	SUCCESSFUL ~~ CAREER	5.58	PAST ~~ DECADE	17.85	SEXUAL ~~ BEHAVIOR*	5.41
PLASTIC ~~ BAG	1417.08	POLITICAL ~~ RIGHT	10.96	SHORT ~~ TIME	12.81	STRICT ~~ DIET	18.31
PLASTIC ~~ SACK	265.17	POLITICAL ~~ SCENE	10.46	SHORT ~~ PERIOD	11.39	STRICT ~~ TEACHER	13.96
PLASTIC ~~ SURGERY	61.15	POLITICAL ~~ INSIDER	9.3	SHORT ~~ SENTENCE	7.41	STRICT ~~ RULE	8.9
PLASTIC ~~ MATERIAL	10.63	POLITICAL ~~ CORRUPTION	7.61	SHORT ~~ RESOLUTION	6.48	MARINE ~~ LIFE	44.37
PLASTIC ~~ PRODUCT	8.72	POLITICAL ~~ PARTY	5.99	SHORT ~~ RANGE	5.53	MARINE ~~ ANIMAL	27.4
PLASTIC ~~ WASTE	8.21	POLITICAL ~~ PRISONER	5.3	POSSIBLE ~~ SOLUTION	9.97	MARINE ~~ MAMMAL	20.55
SAME ~~ TIME	65.71	FAMOUS ~~ ACTOR	17.99	MEDICAL ~~ MARIJUANA	75.67	MARINE ~~ WILDLIFE	9.69
SAME ~~ SEX	22.13	FAMOUS ~~ ARTIST	8.91	MEDICAL ~~ TREATMENT	18.92	MARINE ~~ SPECIE	9.41
SAME ~~ MARRIAGE	12.61	FAMOUS ~~ COMPANY	7.13	MEDICAL ~~ DRUG	12.04	MARINE ~~ ECOSYSTEM	6.64
SAME ~~ THING	12.31	FAMOUS ~~ PLACE	6.55	MEDICAL ~~ CONDITION	8.62	MARINE ~~ CREATURE	6.15
SAME ~~ GENDER	10.13	FAMOUS ~~ PEOPLE	5.04	MEDICAL ~~ ADVICE	8.01	WIDE ~~ VARIETY	11.23
SAME ~~ LEVEL	8.74	ONLY ~~ WAY	17.43	MEDICAL ~~ STORAGE	6.01	WIDE ~~ RANGE	9.9
SAME ~~ ARTICLE	8.6	ONLY ~~ ORGAN	6.69	MEDICAL ~~ SCIENCE	5.83	FAVORITE ~~ SHOW*	7.23
SAME ~~ ROOM	7.93	MAJOR ~~ PROBLEM	22.18	NORMAL ~~ LIFE	15.03	FAVORITE ~~ TV*	6.44

SAME ~~ AMOUNT	5.79	MAJOR ~~ ISSUE	17.07	NORMAL ~~ PEOPLE	7.63	FAVORITE ~~ POSSESSION*	5.94
HIGH ~~ SCHOOL	83.02	MAJOR ~~ MUSIC	6.96	ENOUGH ~~ RESTAURANT	10.49	SUITABLE ~~ PARTNER	14.24
HIGH ~~ INTENSITY	50.42	MAJOR ~~ ROLE	6.54	ENOUGH ~~ TIME	8.45	FINANCIAL ~~ RESPONSIBILITY	6.15
HIGH ~~ LEVEL	44.77	MAJOR ~~ FLOOD	5.99	ENOUGH ~~ EVIDENCE	7.26	FINANCIAL ~~ PROBLEM	5.8
HIGH ~~ MARK	19.71	MAJOR ~~ PART	5.91	ENOUGH ~~ SEAT	5.02	RECENT ~~ YEAR	9.17
HIGH ~~ EDUCATION	14.38	MAJOR ~~ CAUSE	5.64	MODERN ~~ SOCIETY	16.79	RECENT ~~ TIME	7.79
HIGH ~~ GRADUATE	12.18	NEGATIVE ~~ EFFECT	147.48	MODERN ~~ DAY	14.23	RECENT ~~ STUDY	6.51
ENJOYABLE ~~ THING	6.09	NEGATIVE ~~ IMPACT	82.7	MODERN ~~ TECHNOLOGY	8.62		

Table A-21: Extracted Adjective-Noun Collocations with LL scores for the NBESs Corpus

Adjective- Noun collocations	LL	Adjective-Noun collocations	LL	Adjective - Noun collocations	LL	Adjective-Noun collocations	LL
OTHER ~~ HAND	1978.25	PARTICULAR ~~ COUNTRY	14.42	NORMAL ~~ PLOT	14.55	OBVIOUS ~~ CHOICE	6.84
OTHER ~~ WORD	590.56	PARTICULAR ~~ LANGUAGE	12.65	NORMAL ~~ BOWEL	13.23	OBVIOUS ~~ SIGN	5.95
OTHER ~~ FACTOR	423.81	PARTICULAR ~~ SEGMENT	12.56	NORMAL ~~ TEMPERATURE	13.03	OBVIOUS ~~ BENEFIT	5.86
OTHER ~~ COUNTRY	231.59	PARTICULAR ~~ TIME	12.55	NORMAL ~~ DISTANCE	12.3	OBVIOUS ~~ WAY	5.62

OTHER ~~ PEOPLE	147.57	PARTICULAR ~~ TASK	11.67	NORMAL ~~ LIFE	11.07	RURAL ~~ AREA	217.59
OTHER ~~ MEMBER	139.56	CULTURAL ~~ DIFFERENCE	167.06	NORMAL ~~ SURVIVAL	10.93	RURAL ~~ COMMUNITY	123.51
OTHER ~~ AREA	108.24	CULTURAL ~~ HERITAGE	60.93	INITIAL ~~ CINDITION	70.37	RURAL ~~ BUSINESS	78.67
OTHER ~~ SIDE	100.28	CULTURAL ~~ DIVERSITY	53.58	INITIAL ~~ STAGE	70.11	RURAL ~~ DEVELOPMENT	29.7
OTHER ~~ CAUSE	97.52	CULTURAL ~~ BACKGROUND	52.97	INITIAL ~~ POSITION	40.15	RURAL ~~ ECONOMY	21.09
OTHER ~~ THING	90.45	CULTURAL ~~ CONTEXT	47.45	INITIAL ~~ INVESMENT	34.89	RURAL ~~ POPULATION	20.46
OTHER ~~ SPECIE	85.45	CULTURAL ~~ TRADITION	42.85	INITIAL ~~ VELOCITY	32.51	RURAL ~~ SECTOR	19.72
OTHER ~~ VARIABLE	83.02	CULTURAL ~~ IDENTITY	39.03	INITIAL ~~ OUTLAY	27.38	RURAL ~~ FOX	19
OTHER ~~ GROUP	81.03	CULTURAL ~~ VALUE	36.5	INITIAL ~~ VALUE	21.01	RURAL ~~ EXTENSION	18.01
OTHER ~~ METHOD	80.96	CULTURAL ~~ RELATIVISM	33.75	INITIAL ~~ CONCENTRATION	19.91	RURAL ~~ BOUNDARY	16.48
OTHER ~~ ASPECT	73.18	CULTURAL ~~ MECCA	31.05	INITIAL ~~ ASYMMETRY	18.94	RURAL ~~ LAYER	13.99
OTHER ~~ WAY	72.04	CULTURAL ~~ DIMENSION	29.05	INITIAL ~~ LEVEL	17.08	RURAL ~~ ENTERPRISE	13.53
OTHER ~~ TYPE	71.86	CULTURAL ~~ WEB	28.82	INITIAL ~~ DESIGN	16.99	RURAL ~~ PEOPLE	9.61
OTHER ~~ SOURCE	59.16	CULTURAL ~~ NORM	23.3	INITIAL ~~ CONSONANT	16.97	RURAL ~~ SCIENTIST	8.29
OTHER ~~ FORM	57.89	CULTURAL ~~ INFLUENCE	22.92	INITIAL ~~ COST	16.88	RURAL ~~ LANDSCAPE	8.22
OTHER ~~ CHARACTER	53.93	CULTURAL ~~ HISTORY	22.14	INITIAL ~~ READING	16.35	RURAL ~~ POVERTY	8.02

HIGH ~~ LEVEL	1342.73	CULTURAL ~~ CAPITAL	19.26	INITIAL ~~ HYPOTHESIS	14.7	RURAL ~~ SETTING	7.96
HIGH ~~ QUALITY	394.95	CULTURAL ~~ SUPERIORITY	18.84	INITIAL ~~ EVALUATION	13.83	RURAL ~~REGION	7.58
HIGH ~~ RATE	373.48	CULTURAL ~~ REVOLUTION	18.23	INITIAL ~~ ASSESSMENT	13.63	RURAL ~~ SOCIETY	6.7
HIGH ~~ TEMPERATURE	298.61	CULTURAL ~~ ASPECT	15.14	INITIAL ~~ INCOME	13.37	IMPOSSIBLE ~~ TASK	19.12
HIGH ~~ PRICE	253.1	CULTURAL ~~ FACTOR	14.61	INITIAL ~~ SPECIFICATION	13.28	IMPOSSIBLE ~~ EVERYTHING	7.76
HIGH ~~ DEGREE	196.09	PRESENT ~~ COMPLAINT	192.87	INITIAL ~~ TREATMENT	13.27	POWERFUL ~~ TOOL	103.48
HIGH ~~ COST	158.55	PRESENT ~~ ILLNESS	99.52	INITIAL ~~ PHASE	13.17	POWERFUL ~~ FORCE	24.35
HIGH ~~ RISK	141.9	PRESENT ~~ DAY	58.99	FEMALE ~~ EXPATRIATE	127.78	POWERFUL ~~ POSITION	18.77
HIGH ~~ VALUE	131.46	PRESENT ~~ SYMPTOM	43.81	FEMALE ~~ CHARACTER	92.25	POWERFUL ~~ IMAGE	18.08
HIGH ~~ STANDARD	128.37	PRESENT ~~ SITUATION	27.8	FEMALE ~~ EDUCATION	51.93	POWERFUL ~~ STATE	15.19
HIGH ~~ CONCENTRATION	120	PRESENT ~~ CASE	22.35	FEMALE ~~ INJECTING	50.38	POWERFUL ~~ COMPUTER	12.29
HIGH ~~ FREQUENCY	114.19	PRESENT ~~ STUDY	18.71	FEMALE ~~ SEXUALITY	48.97	POWERFUL ~~ NATION	9.9
HIGH ~~ SPEED	113.25	PRESENT ~~ PROBLEM	17.29	FEMALE ~~ CHIMPANZEE	44.48	POWERFUL ~~ ALLY	9.42
HIGH ~~ PROPORTION	111.85	PRESENT ~~ VALUE	16.48	FEMALE ~~ SPEECH	37.47	POWERFUL ~~ CHARACTER	8.45
HIGH ~~ WAGE	110.59	PRESENT ~~ MOMENT	15.16	FEMALE ~~ SUBORDINATION	28.58	POWERFUL ~~ MUSCLE	7.99
HIGH ~~ STATUS	101.48	PRESENT ~~ TABLE	13.93	FEMALE ~~ COMMUNITY	25.04	POWERFUL ~~ EMOTION	7.54

HIGH ~~ PRESSURE	96.12	PRESENT ~~ VALUE	12.8			POWERFUL ~~ MAN	7.41
HIGH ~~ PERCENTAGE	96.05	PRESENT ~~ TIME	12.1	FEMALE ~~ VAMPIRE	21.03	REASONABLE ~~ DOUBT	49.1
HIGH ~~ UNEMPLOYMENT	91.45	PRESENT ~~ DIFFICULTY	10.19	FEMALE ~~ STUDENT	20.42	REASONABLE ~~ MAN	40.32
HIGH ~~ ENERGY	91.05	PRESENT ~~ OVERVIEW	9.64	FEMALE ~~ VOICE	20	REASONABLE ~~ AMOUNT	29.86
DIFFERENT ~~ TYPE	486.99	PRESENT ~~ PAPER	8.98	FEMALE ~~ PROTAGONIST	19.86	REASONABLE ~~ PRICE	18.15
DIFFERENT ~~ WAY	307.28	PRESENT ~~ READER	7.33	FEMALE ~~ ROLE	19.13	REASONABLE ~~ ESTIMATE	17.01
DIFFERENT ~~ GROUP	133.25	PRESENT ~~ SCENARIO	7.18	FEMALE ~~ GENDER	18.31	REASONABLE ~~ CARE	16.83
DIFFERENT ~~ LEVEL	132.41	PRESENT ~~ VERB	7.02	FEMALE ~~ BEHAVIOUR	16.16	REASONABLE ~~ CONFIDENCE	14.7
DIFFERENT ~~ CULTURE	115.62	PRESENT ~~ RESULT	6.44	FEMALE ~~ PRATICIPATION	15.83	REASONABLE ~~ PERSON	13.91
DIFFERENT ~~ ASPECT	103.52	MAJOR ~~ PROBLEM	96.62	FEMALE ~~ TRANSCRIPT	15.71	REASONABLE ~~ EXPECTATION	13.51
DIFFERENT ~~ METHOD	96.16	MAJOR ~~ FACTOR	86.43	FEMALE ~~ EMPLOYEE	15.54	REASONABLE ~~ ASSUMPTION	12.3
DIFFERENT ~~ APPROACH	94.97	MAJOR ~~ DIFFERENCE	80.09	EQUAL ~~ OPPORTUNITY	66.07	REASONABLE ~~ STEP	11.58
DIFFERENT ~~ COUNTRY	90.11	MAJOR ~~ ROLE	70.44	EQUAL ~~ RIGHT	28.17	REASONABLE ~~ LEVEL	10.76
DIFFERENT ~~ KIND	88.76	MAJOR ~~ PART	45.25	EQUAL ~~ FOOTING	24.85	REASONABLE ~~ OPTION	10.71
DIFFERENT ~~ FORM	79.65	MAJOR ~~ CAUSE	44.41	EQUAL ~~ IMPORTANCE	24.72	REASONABLE ~~ ESTIMATION	9.18
DIFFERENT ~~ PERSPECTIVE	75.6	MAJOR ~~ ISSUE	43.18	EQUAL ~~ NUMBER	23.52	REASONABLE ~~ SIZE	8.3

DIFFERENT ~~ SPECIE	70.94	MAJOR ~~ CHANGE	41.72	EQUAL ~~ ACCESS	21.91	REASONABLE ~~ TIME	8.07
DIFFERENT ~~ STAGE	66.71	MAJOR ~~ ADVANTAGE	36.01	EQUAL ~~ VARIANCE	17.36	REASONABLE ~~ CONCLUSION	7.68
DIFFERENT ~~ STYLE	63.71	MAJOR ~~ SOURCE	35.57	EQUAL ~~ CHANCE	15.94	REASONABLE ~~ RANGE	6.08
DIFFERENT ~~ PART	52.63	MAJOR ~~ CONCERN	35.52	EQUAL ~~ DISTRIBUTION	15.11	REASONABLE ~~ COST	6.07
DIFFERENT ~~ AREA	45.31	MAJOR ~~ CONTRIBUTOR	33.91	EQUAL ~~ FORMULA	13.81	REASONABLE ~~ GROUND	6.05
DIFFERENT ~~ PEOPLE	45	MAJOR ~~ IMPACT	30.04	EQUAL ~~ VALUE	12.21	RICH ~~ COUNTRY	55.16
DIFFERENT ~~ TIME	42.85	MAJOR ~~ CHALLENGE	29.78	EQUAL ~~ COST	11.88	RICH ~~ PICTURE	21.38
DIFFERENT ~~ THING	41.92	MAJOR ~~ PLAYER	27.83	EQUAL ~~ SIZE	10.99	RICH ~~ NATION	19.4
SOCIAL ~~ CLASS	579.22	MAJOR ~~ THEME	25.42	EQUAL ~~ PAY	10.25	RICH ~~ SOURCE	13.19
SOCIAL ~~ RESPONSIBILITY	366.08	MAJOR ~~ CITY	22.6	EQUAL ~~ PROPORTION	9.74	RICH ~~ DIVERSITY	12.75
SOCIAL ~~ INTERACTION	220.67	MAJOR ~~ DRAWBACK	22.33	EQUAL ~~ AMOUNT	9.72	RICH ~~ BURIAL	10.85
SOCIAL ~~ STRUCTURE	211.53	MAJOR ~~ COMPONENT	21.03	EQUAL ~~ LENGTH	9.7	RICH ~~ SOIL	10.32
SOCIAL ~~ GROUP	166.35	MAJOR ~~ STAKEHOLDER	18.27	EQUAL ~~ BASIS	9.2	RICH ~~ BIODIVERSITY	10.15
SOCIAL ~~ SCIENCE	149.24	REAL ~~ WORLD	338.31	EQUAL ~~PROBABILITY	8.36	RICH ~~ FOOD	6.94
SOCIAL ~~ STATUS	138.01	REAL ~~ LIFE	188.64	EQUAL ~~ WEIGHT	7.59	SEPARATE ~~ ENTITY	55
SOCIAL ~~ RELATION	132.97	REAL ~~ WAGE	171.19	FULL ~~ EMPLOYMENT	89.79	SEPARATE ~~ PERSONALITY	15.68
SOCIAL ~~ FACILITATION	114.92	REAL ~~ RATE	72.9	FULL ~~ POTENTIAL	81.91	SEPARATE ~~ CATEGORY	14.96

SOCIAL ~~ CONTRACT	108.82	REAL ~~ SITUATION	47.96	FULL ~~ TIME	52.41	SEPARATE ~~ OCCASION	14.88
SOCIAL ~~ CONTEXT	106.44	REAL ~~ EXCHANGE	42.24	FULL ~~ RANGE	40.86	SEPARATE ~~ SCIENCE	14.28
SOCIAL ~~ LIFE	104.57	REAL ~~ INTEREST	40.06	FULL ~~ UNDERSTANDING	38.83	SEPARATE ~~ SPHERE	13.43
SOCIAL ~~ CAPITAL	96.04	REAL ~~ POSSIBILITY	38.25	FULL ~~ ADVANTAGE	34.14	SEPARATE ~~ AREA	10.74
SOCIAL ~~ POLICY	95.84	REAL ~~ THREAT	33.92	FULL ~~ LOAD	31.48	SEPARATE ~~ REALM	10.58
SOCIAL ~~ MOVEMENT	86.46	REAL ~~ TIME	31.61	FULL ~~ PICTURE	30.3	SEPARATE ~~ SECTION	10.47
SOCIAL ~~ REALITY	85.17	REAL ~~ ESTATE	20.75	FULL ~~ CAPACITY	28.04	SEPARATE ~~ REGRESSION	10.41
SOCIAL ~~ INCLUSION	78.19	REAL ~~ INCOME	19.87	FULL ~~ EXTENT	26.69	SEPARATE ~~ GROUP	9.24
SOCIAL ~~ FORM	76.13	REAL ~~ PROBLEM	18.36	FULL ~~ TERM	19.74	SEPARATE ~~ ENTRY	8.62
SOCIAL ~~ IDENTITY	75.54	REAL ~~ VALUE	18.19	FULL ~~ EXAMINATION	17.4	SEPARATE ~~ SYSTEM	7.77
SOCIAL ~~ MOBILITY	71.54	REAL ~~ DANGER	17.3	FULL ~~ HARDENING	15.95	SEPARATE ~~ CLASS	5.33
NEW ~~ TECHNOLOGY	332.96	REAL ~~ EIGNVALUE	16.58	FULL ~~ ORBIT-FORM	15.85	SEPARATE ~~ PART	5.3
NEW ~~ ENTRANT	273.08	REAL ~~ VECTOR	13.12	FULL ~~ SCALE	15.61	LIGHT ~~ INTENSITY	73.93
NEW ~~ PRODUCT	227.24	REAL ~~ BALANCE	11.55	FULL ~~ LISTING	15.54	LIGHT ~~ RAIL	64.81
NEW ~~ IDEA	126.07	REAL ~~ OUTPUT	11.52	FULL ~~ VERSION	15.15	LIGHT ~~ SOURCE	51.99
NEW ~~ DILEMMAS	103.82	LEGAL ~~ SYSTEM	227.25	FULL ~~ WORKING	14.41	LIGHT ~~ BULB	45.9
NEW ~~ MARKET	95.8	LEGAL ~~ PROFESSION	114.65	FULL ~~ RATIONALITY	14.4	LIGHT ~~ STIMULUS	36.01
NEW ~~ INVENTION	84.34	LEGAL ~~ ADVICE	88.74	FULL ~~ CIRCLE	13.6	LIGHT ~~ MICROSCOPE	34.29
NEW ~~ MILLENNIUM	83.25	LEGAL ~~ RELATION	87.33	PRIVATE ~~ SECTOR	324.16	LIGHT ~~ CHAIN	30.5
NEW ~~ SYSTEM	82.89	LEGAL ~~ FRAMEWORK	58.9	PRIVATE ~~ SPHERE	188.05	LIGHT ~~ DRINKER	30.49

NEW ~~ WORD	78.49	LEGAL ~~ IMPLICATION	53.27	PRIVATE ~~ ASYLUM	66.23	LIGHT ~~ MODULATOR	18.35
NEW ~~ FORM	71.94	LEGAL ~~ RIGHT	50	PRIVATE ~~ PROPERTY	56.82	LIGHT ~~ WAVE	17.32
NEW ~~ APPROACH	62.08	LEGAL ~~ OBLIGATION	46.07	PRIVATE ~~ MADHOUSE	50.08	LIGHT ~~ SPOT	17.25
NEW ~~ GENERATION	58.95	LEGAL ~~ PROTECTION	36.48	PRIVATE ~~ LIFE	49.92	LIGHT ~~ WAVELENGTH	14.56
NEW ~~ MEMBER	56.36	LEGAL ~~ INSTRUMENT	30.16	PRIVATE ~~ DOMAIN	48.16	LIGHT ~~ CONE	13.58
NEW ~~ MOVEMENT	51.2	LEGAL ~~ PLURALISM	28.14	PRIVATE ~~ OWNERSHIP	45.75	LIGHT ~~ INCIDENT	12.88
NEW ~~ SPECIE	46.39	LEGAL ~~ PRINCIPLE	27.9	PRIVATE ~~ COMPANY	35.89	LIGHT ~~ REACTION	11.84
NEW ~~ CONCEPT	42.86	LEGAL ~~ REQUIREMENT	27.63	PRIVATE ~~ INVESTMENT	31.72	LIGHT ~~ EXPOSURE	10.75
NEW ~~ SPECIES	42.64	LEGAL ~~ STATUS	26.63	PRIVATE ~~ SCHOOL	27.67	LIGHT ~~ MICROSCOPY	10.08
NEW ~~ PEER	41.66	LEGAL ~~ PARENT	25.12	PRIVATE ~~ ENTERPRISE	26.94	LIGHT ~~ SHINE	9.84
NEW ~~ DEVELOPMENT	41.51	LEGAL ~~ POSITION	23.12	PRIVATE ~~ KEY	26.2	LIGHT ~~ ENERGY	8.34
GOOD ~~ SERVICE	257.58	LEGAL ~~ PROCEEDING	22.99	PRIVATE ~~ LANGUAGE	24.59	LIGHT ~~ SCATTERING	8.27
GOOD ~~ FIT	232.17	LEGAL ~~ PERCENTAGE	22.29	PRIVATE ~~ CONVERSATION	21.62	HUGE ~~ AMOUNT	118.76
GOOD ~~ QUALITY	202	LEGAL ~~ ORDER	22.22	PRIVATE ~~ CULT	19.08	HUGE ~~ NUMBER	31.31
GOOD ~~ WAY	201.8	LEGAL ~~ CONSEQUENCE	21.97	PRIVATE ~~ AGENT	16.32	HUGE ~~ IMPACT	27.04
GOOD ~~ EXAMPLE	198.8	ENVIRONMENTAL ~~ PROTECTION	176.02	PRIVATE ~~ CAR	15.1	HUGE ~~ EFFECT	18.74

GOOD ~~ UNDERSTANDING	148.03	ENVIRONMENTAL ~~ IMPACT	148.31	PRIVATE ~~ INVESTOR	14.75	HUGE ~~ SUCCESS	17.76
GOOD ~~ PRACTICE	144.83	ENVIRONMENTAL ~~ SUSTAINABILITY	129.01	PRIVATE ~~ RIGHT	13.16	HUGE ~~ INCREASE	17.11
GOOD ~~ INTEREST	99.41	ENVIRONMENTAL ~~ FACTOR	119.95	ESSENTIAL ~~ PART	49.85	HUGE ~~ PROFIT	17.07
GOOD ~~ CHOICE	75.15	ENVIRONMENTAL ~~ ISSUE	104.54	ESSENTIAL ~~ ELEMENT	46.78	HUGE ~~ LOSS	17
GOOD ~~ PERFORMANCE	70.54	ENVIRONMENTAL ~~ CONDITION	101.04	ESSENTIAL ~~ NUTRIENT	27.53	HUGE ~~ DEBT	15.99
GOOD ~~ OPTION	54.05	ENVIRONMENTAL ~~ POLICY	84.25	ESSENTIAL ~~ FEATURE	21.1	HUGE ~~ RANGE	15.65
GOOD ~~ METHOD	52.11	ENVIRONMENTAL ~~ LAW	80.54	ESSENTIAL ~~ COMPONENT	17.92	HUGE ~~ INVESTMENT	14
GOOD ~~ INDICATOR	51.85	ENVIRONMENTAL ~~ MANAGEMENT	62.5	ESSENTIAL ~~ OIL	12.63	HUGE ~~ SUM	12.73
GOOD ~~ MANAGEMENT	50.9	ENVIRONMENTAL ~~ DEGRADATION	53.94	ESSENTIAL ~~ NATURE	11.33	HUGE ~~ INFLUENCE	12.5
GOOD ~~ THING	48.98	ENVIRONMENTAL ~~ CONCERN	49.81	ESSENTIAL ~~ AMINO	10.4	HUGE ~~ POTENTIAL	11.88
GOOD ~~ IDEA	48.48	ENVIRONMENTAL ~~ LEGISLATION	44.86	ESSENTIAL ~~ ROLE	10.27	HUGE ~~ VARIETY	8.2
GOOD ~~ RESULT	48.01	ENVIRONMENTAL ~~ REGULLATION	35.47	ESSENTIAL ~~ REQUIREMEMENT	9.97	HUGE ~~ MARKET	8.17
GOOD ~~ APPROXIMATION	46.93	ENVIRONMENTAL ~~ EFFECT	29.34	ESSENTIAL ~~ ORDER	9.92	HUGE ~~ PROBLEM	8.16
GOOD ~~ CONDITION	46.13	ENVIRONMENTAL ~~ CHANGE	26.12	ESSENTIAL ~~ FACTOR	9.44	HUGE ~~ DEBATE	7.92
GOOD ~~ RELATIONSHIP	43.92	ENVIRONMENTAL ~~ POLLUTION	26.1	ESSENTIAL ~~ ACID	6.86	HUGE ~~ CAPITAL	7.34
IMPORTANT ~~ ROLE	418.69	ENVIRONMENTAL ~~ STANDARD	25.22	ESSENTIAL ~~ ASPECT	6.55	HUGE ~~ GAP	7.22

IMPORTANT ~~ FACTOR	373.53	ENVIRONMENTAL ~~ ARCHAEOLOGY	22.62	ESSENTIAL ~~ TOOL	6.33	SERIOUS ~~ THREAT	82.74
IMPORTANT ~~ ASPECT	209.4	ENVIRONMENTAL ~~ PROBLEM	20.2	ESSENTIAL ~~ CHARACTERISTIC	5.52	SERIOUS ~~ PROBLEM	74.42
IMPORTANT ~~ PART	182.39	ENVIRONMENTAL ~~ DAMAGE	15.95	STANDARD ~~ DEVIATION	481.18	SERIOUS ~~ CRIME	34.18
IMPORTANT ~~ ISSUE	116.31	KEY ~~ FACTOR	146.91	STANDARD ~~ ERROR	106.47	SERIOUS ~~ IMPLICATION	29.15
IMPORTANT ~~ ELEMENT	70.73	KEY ~~ ISSUE	89.74	STANDARD ~~ LIVING	75.18	SERIOUS ~~ HARM	25.61
IMPORTANT ~~ THING	51.77	KEY ~~ INFORMATION	84.53	STANDARD ~~ SOLUTION	53.33	SERIOUS ~~ OFFENSE	23.57
IMPORTANT ~~ POINT	50.22	KEY ~~ ROLE	71.25	STANDARD ~~ CARE	40.1	SERIOUS ~~ ISSUE	23.24
IMPORTANT ~~ FEATURE	41.81	KEY ~~ FEATURE	65.82	STANDARD ~~ CURVE	38.95	SERIOUS ~~ INJURY	19.51
IMPORTANT ~~ SOURCE	40.46	KEY ~~ ELEMENT	65.16	STANDARD ~~ FORM	30.28	SERIOUS ~~ CONSEQUENCE	17.52
IMPORTANT ~~ IMPLICATION	37.34	KEY ~~ POINT	57.77	STANDARD ~~ CHOCOLATE	30.15	SERIOUS ~~ DISEASE	17.22
IMPORTANT ~~ DETERMINANT	31.18	KEY ~~ PLAYER	54.29	STANDARD ~~ METHOD	16.92	SERIOUS ~~ COMPLICATION	15.47
IMPORTANT ~~ TOOL	31.17	KEY ~~ STAKEHOLDER	45.35	STANDARD ~~ SODIUM	15.06	SERIOUS ~~ DAMAGE	15.29
IMPORTANT ~~ CONSIDERATION	24.28	KEY ~~ ASPECT	40.06	STANDARD ~~ SET	14.58	SERIOUS ~~ RISK	13.97
IMPORTANT ~~ DISTINCTION	24.13	KEY ~~ CONCEPT	35.61	STANDARD ~~ OVALBUMIN	14.57	SERIOUS ~~ CONSIDERATION	12.28
IMPORTANT ~~ CONTRIBUTION	23.89	KEY ~~ AREA	32.65	STANDARD ~~ SERVICE	13.88	SERIOUS ~~ VIOLATION	12.14
IMPORTANT ~~ THEME	19.21	KEY ~~ THEME	29.71	STANDARD ~~ PROCEDURE	12.76	SERIOUS ~~ EFFECT	12.11

IMPORTANT ~~ COMPONENT	19.12	KEY ~~ SUCCESS	29.09	STANDARD ~~ WIDTH	11.96	SERIOUS ~~ ILLNESS	11.82
IMPORTANT ~~ REASON	18.83	KEY ~~ INDICATOR	25.77	STANDARD ~~ PROOF	11.5	SERIOUS ~~ FLAW	11.46
IMPORTANT ~~ DECISION	17.4	KEY ~~ CRYPTOGRAPHY	19.62	STANDARD ~~ HYDROXIDE	11.35	SERIOUS ~~ DIFFICULTY	10.82
INTERNATIONAL ~~ RELATION	419.69	KEY ~~ COMPONENT	18.71	STANDARD ~~ REVUE	11.13	SERIOUS ~~ THOUGHT	8.53
INTERNATIONAL ~~ LAW	415.73	KEY ~~ DETERMINANT	17.14	STANDARD ~~ PRACTICE	10.72	MILITARY ~~ FORCE	62.77
INTERNATIONAL ~~ TRADE	172.96	KEY ~~ VARIABLE	17.1	STANDARD ~~ PROFILE	10.16	MILITARY ~~ EXPENDITURE	54.71
INTERNATIONAL ~~ SYSTEM	155.79	KEY ~~ PRINCIPLE	15.95	CLOSE ~~ PROXMITY	142.51	MILITARY ~~ COMMANDER	47.19
INTERNATIONAL ~~ INSTITUTION	125.82	TRUE ~~ NATURE	29.29	CLOSE ~~ RELATIONSHIP	138.33	MILITARY ~~ INTERVENTION	30.94
INTERNATIONAL ~~ COMMUNITY	115.61	TRUE ~~ FEELING	24.59	CLOSE ~~ INSPECTION	77.63	MILITARY ~~ CAPABILITY	24.82
INTERNATIONAL ~~ POLITICS	103.98	TRUE ~~ VALUE	23.82	CLOSE ~~ FRIEND	69.81	MILITARY ~~ SERVICE	24.8
INTERNATIONAL ~~ ORGANIZATION	97.18	TRUE ~~ IDENTITY	23.44	CLOSE ~~ CONTACT	60.01	MILITARY ~~ POWER	22.52
INTERNATIONAL ~~ MARKET	92.98	TRUE ~~ MINIMA	22.6	CLOSE ~~ SCRUTINY	32.27	MILITARY ~~ CAMPAIGN	18.02
INTERNATIONAL ~~ SOCIETY	88.33	TRUE ~~ PREMISS	19.62	CLOSE ~~ ATTENTION	26.47	MILITARY ~~ DICTATORSHIP	17.88
INTERNATIONAL ~~ ASSIGNMENT	75.05	TRUE ~~ INTENTION	19.44	CLOSE ~~ LOOK	25.36	MILITARY ~~ SUPERIORITY	12.96
INTERNATIONAL ~~ ECONOMY	70.54	TRUE ~~ BELIEF	17.19	CLOSE ~~ RELATIVE	22.63	MILITARY ~~ SECURITY	12.83
INTERNATIONAL ~~ STUDENT	67.41	TRUE ~~ MEANING	15.05	CLOSE ~~ EYE	20.58	MILITARY ~~ EQUIPMENT	12.6

INTERNATIONAL ~~ COOPERATION	62.95	TRUE ~~ SELF	13.99	CLOSE ~~ EXAMINATION	19.32	MILITARY ~~ THREAT	12.49
INTERNATIONAL ~~ RIGHT	54.67	TRUE ~~ STATEMENT	12.16	CLOSE ~~ MATCH	18.91	MILITARY ~~ VICTORY	12.39
INTERNATIONAL ~~ REGIME	44.23	TRUE ~~ SOLUTION	11.08	CLOSE ~~ ASSOCIATION	18.8	MILITARY ~~ STRENGTH	12.22
INTERNATIONAL ~~ LEVEL	43.07	TRUE ~~ PICTURE	10.25	CLOSE ~~ TIE	17.62	MILITARY ~~ OPERATION	12.13
INTERNATIONAL ~~ STANDARD	37.06	TRUE ~~ ESSENCE	9.07	CLOSE ~~ COORPERATION	15.35	MILITARY ~~ CONFLICT	11.28
INTERNATIONAL ~~ ARENA	31.63	TRUE ~~ COGNITION	8.83	CLOSE ~~ KIN	13.83	MILITARY ~~ CONQUEST	10.33
INTERNATIONAL ~~ BUSINESS	29.22	TRUE ~~ THOUGHT	7.69	CLOSE ~~ GAP	12.99	MILITARY ~~ ALLIANCE	9.52
GREAT ~~ DEAL	494.8	TRUE ~~ RELATIVE	7.47	CLOSE ~~ MONITORING	11.94	MILITARY ~~ SPENDING	9.02
GREAT ~~ IMPORTANCE	162.34	TRUE ~~ CONCLUSION	6.27	CLOSE ~~ FRIENDSHIP	11.91	FAIR ~~ TRAIL	97.78
GREAT ~~ NUMBER	119.06	TRUE ~~ LOVE	6.25	CLOSE ~~ LOCATION	11.29	FAIR ~~ VIEW	91.97
GREAT ~~ EXTENT	93.47	TRUE ~~ CITIZEN	5.32	MALE ~~ FEMALE	215.21	FAIR ~~ TRADE	52.12
GREAT ~~ POWER	75.52	SIMPLE ~~ MODEL	24.08	MALE ~~ DOMINANCE	65.39	FAIR ~~ HEARING	51.71
GREAT ~~ EMPHASIS	74.58	SIMPLE ~~ WAY	23.76	MALE ~~ BREADWINNER	35.79	FAIR ~~ CHOCOLATE	23.36
GREAT ~~ DETAIL	69.96	SIMPLE ~~ STRUCTURE	23.76	MALE ~~ COUNTERPART	33.45	FAIR ~~ VALUE	21.62
GREAT ~~ DEGREE	65.95	SIMPLE ~~ METHOD	23.02	MALE ~~ DOMINATION	31.98	FAIR ~~ AMOUNT	19.52
GREAT ~~ AMOUNT	64.72	SIMPLE ~~ LOW-PASS	22.27	MALE ~~ ATHLETE	28.29	FAIR ~~ REPRESENTATION	16.65
GREAT ~~ IMPACT	56.76	SIMPLE ~~ IDEA	21.87	MALE ~~ PROTAGNIST	22.51	FAIR ~~ SHARE	14.31

GREAT ~~ UNDERSTANDING	51.01	SIMPLE ~~ TERM	21.8	MALE ~~ CHARACTER	19.96	FAIR ~~ DISTRIBUTION	10
GREAT ~~ EFFECT	44.29	SIMPLE ~~ FORM	21.3	MALE ~~ TRANSCRIPT	18.43	FAIR ~~ COMPARISON	7.97
GREAT ~~ VARIETY	43.79	SIMPLE ~~ EXTENSION	20.3	MALE ~~ SEX	17.4	FAIR ~~ COMPETITION	5.99
GREAT ~~ DEPTH	39.67	SIMPLE ~~ REGRESSION	15.96	MALE ~~ PROFESSION	17.4	FAIR ~~ TEST	5.22
GREAT ~~ SUCCESS	39.54	SIMPLE ~~ EXAMPLE	14.75	MALE ~~ MODEL	16.33	PROFESSIONAL ~~ ENGINEER	58.47
GREAT ~~ EXPECTATION	35.07	SIMPLE ~~ TASK	14.35	MALE ~~ LION	14.52	PROFESSIONAL ~~ MEDICINE	29.06
GREAT ~~ RISK	33.32	SIMPLE ~~ CALCULATION	14.12	MALE ~~ SPEAKER	13.88	PROFESSIONAL ~~ BODY	22.77
GREAT ~~ DISTANCE	32.88	SIMPLE ~~ ALGEBRA	12.73	MALE ~~ BEHAVIOUR	12.85	PROFESSIONAL ~~ PRACTICE	17.04
GREAT ~~ CONCERN	32.26	SIMPLE ~~ FACT	12.37	MALE ~~ SUPERIORITY	12.67	PROFESSIONAL ~~ SKILL	15.89
GREAT ~~ SIGNIFICANCE	31.15	SIMPLE ~~ FILTER	11.98	MALE ~~ BIAS	11.81	PROFESSIONAL ~~ MANNER	15.15
LOW ~~ LEVEL	448.45	SIMPLE ~~ DESCRIPTION	11.69	MALE ~~ DOCTOR	11.62	PROFESSIONAL ~~ PRACTITIONER	13.02
LOW ~~ COST	311.08	SIMPLE ~~ DIAGRAM	11.31	MALE ~~ BODY	11.31	PROFESSIONAL ~~ TRANSLATOR	12.05
LOW ~~ CLASS	257.32	SIMPLE ~~ PROFILE	11.31	MALE ~~ NIGHTMARE	10.79	PROFESSIONAL ~~ PRIVILEGE	11.68
LOW ~~ PRICE	251.47	SIMPLE ~~ SENTENCE	10.65	ACTUAL ~~ VALUE	58.34	PROFESSIONAL ~~ SERVICE	8.98
LOW ~~ RATE	210.93	SINGLE ~~ CAUSE	83.7	ACTUAL ~~ COST	29.06	PROFESSIONAL ~~ LIFE	8.88
LOW ~~ TEMPERATURE	128.95	SINGLE ~~ MARKET	45.06	ACTUAL ~~ FACT	27.31	PROFESSIONAL ~~ SOLIDER	7.53
LOW ~~ VALUE	97.46	SINGLE ~~ CYLINDER	40.05	ACTUAL ~~ RESULT	20.53	PROFESSIONAL ~~ ORGANIZATION	7.51

LOW ~~ INCOME	94.55	SINGLE ~~ STRAND	26.54	ACTUAL ~~ PRICE	17.05	PROFESSIONAL ~~ CONDUCT	6.79
LOW ~~ DENSITY	85.75	SINGLE ~~ CRYSTAL	26.2	ACTUAL ~~ PERFORMANCE	16.7	PROFESSIONAL ~~ HEALTH	6.5
LOW ~~ ENERGY	79.85	SINGLE ~~ PERSON	25.88	ACTUAL ~~ DRIFT	16.41	PROFESSIONAL ~~ TEAM	6.19
LOW ~~ CONCENTRATION	69.89	SINGLE ~~ STRIP	22.28	ACTUAL ~~ DATUM	16.02	EXTRA ~~ COST	45.94
LOW ~~ WAGE	68.85	SINGLE ~~ WORD	18.89	ACTUAL ~~ HARM	14.62	EXTRA ~~ HOUR	25.06
LOW ~~ ESTEEM	67.11	SINGLE ~~ CURRENCY	17.68	ACTUAL ~~ INFLATION	12.9	EXTRA ~~ POUND	19.28
LOW ~~ RISK	58.42	SINGLE ~~ UNIT	17.12	ACTUAL ~~ SITUATION	12.89	EXTRA ~~ PHTOPSING	18.9
LOW ~~ STATUS	56.17	SINGLE ~~ ENTITY	15.75	ACTUAL ~~ BIAS	12	EXTRA ~~ DIMENSION	16.87
LOW ~~ FREQUENCY	51.53	SINGLE ~~ GENE	14.27	ACTUAL ~~ CONTENT	11.78	EXTRA ~~ EFFORT	15.46
LOW ~~ PRESSURE	49.86	SINGLE ~~ PAIR	14.23	ACTUAL ~~ NUMBER	11.65	EXTRA ~~ TIME	15.44
LOW ~~ FILTER	49.5	SINGLE ~~ MOLECULE	14.23	ACTUAL ~~ TEST	11	EXTRA ~~ CONE	13.92
LOW ~~ PASS	47.95	SINGLE ~~ LOCUS	14.01	ACTUAL ~~ EXPEREINCE	10.81	EXTRA ~~ INFORMATION	13.15
LOW ~~ CONTENT	46.56	SINGLE ~~ INSTANCE	13.98	ACTUAL ~~ INSERTION	9.88	EXTRA ~~ FEATURE	12.35
LARGE ~~ NUMBER	562.04	SINGLE ~~ ROBOT	13.85	ACTUAL ~~ PRACTICE	9.47	EXTRA ~~ MONEY	11.79
LARGE ~~ AMOUNT	504.09	SINGLE ~~MECHANISM	13.23	ACTUAL ~~ EVENT	9.08	EXTRA ~~ UNIT	11.5
LARGE ~~ SCALE	247.81	SINGLE ~~ BLINDING	12.13	ACTUAL ~~ PROFIT	8.98	EXTRA ~~ WEEK	9.93
LARGE ~~ PROPORTION	157.67	SINGLE ~~ PHOTON	11.75	WHOLE ~~ PROCESS	77.68	EXTRA ~~ INCOME	9.92
LARGE ~~ EXTENT	100.66	POOR ~~ COUNTRY	128.22	WHOLE ~~ SYSTEM	71.39	EXTRA ~~ STAFF	9.51

LARGE ~~ QUANTITY	95.91	POOR ~~ QUALITY	98.1	WHOLE ~~ WORLD	41.52	EXTRA ~~ CAPACITY	8.26
LARGE ~~ DAM	81.36	POOR ~~ CONDITION	58.1	WHOLE ~~ RANGE	27.95	EXTRA ~~ PAYMENT	7.46
LARGE ~~ AREA	80.06	POOR ~~ PERFORMANCE	56.39	WHOLE ~~ POPULATION	26.99	EXTRA ~~ REVENUE	7.32
LARGE ~~ PART	79.5	POOR ~~ HEALTH	45.29	WHOLE ~~ PROJECT	25.85	EXTRA ~~ MARK	6.79
LARGE ~~ RANGE	68.66	POOR ~~ FARMER	34.84	WHOLE ~~ COMMUNITY	19.46	EXTRA ~~ RESOURCE	6.65
LARGE ~~ SIZE	66.78	POOR ~~ DIET	34.15	WHOLE ~~ STOREY	18.03	REGULAR ~~ BASIS	96.01
LARGE ~~ GROUP	66	POOR ~~ HOUSEHOLD	27	WHOLE ~~ SOCIETY	14.51	REGULAR ~~ INTERVAL	69.1
LARGE ~~ COMPANY	63.56	POOR ~~ RELIEF	26.91	WHOLE ~~ EDIFICE	14.15	REGULAR ~~ MONITORING	38.74
LARGE ~~ SAMPLE	54.62	POOR ~~ NUTRITION	25.48	WHOLE ~~ GROUP	12.98	REGULAR ~~ MEETING	29.18
LARGE ~~ BREVILLE	42.72	POOR ~~ APPETITE	23.57	WHOLE ~~ TEAM	12.8	REGULAR ~~ CONTACT	25.56
LARGE ~~ MARKET	41.42	POOR ~~ PEOPLE	23.04	WHOLE ~~ MATRIX	12.24	REGULAR ~~ MEDICATION	23.55
LARGE ~~ CORPORATION	39.55	POOR ~~ WOMAN	19.75	WHOLE ~~ GENOME	11.83	REGULAR ~~ PATTERN	23.05
LARGE ~~ FARM	38.53	POOR ~~ PRESERVATION	18.9	WHOLE ~~ BODY	11.76	REGULAR ~~ EXERCISE	20.99
LARGE ~~ VOLUME	36.43	POOR ~~ FAMILY	16.76	WHOLE ~~ DAY	11.47	REGULAR ~~ HEXAGON	20.22
LARGE ~~ FIRM	35.91	POOR ~~ SERVICE	16.43	WHOLE ~~ LIFE	11.35	REGULAR ~~ REVUE	19.21
POLITICAL ~~ POWER	191.01	POOR ~~ SKILL	15.53	WHOLE ~~ SAMPLE	10.62	REGULAR ~~ BLOOD	15.45
POLITICAL ~~ SYSTEM	176.43	POOR ~~ COMMUNICATION	15.31	WHOLE ~~ SILAGE	9.29	REGULAR ~~ ROUTINE	15.27
POLITICAL ~~ PARTY	160.54	POOR ~~ HARVEST	13.64	WHOLE ~~ SET	9.25	REGULAR ~~ VERB	13.41
POLITICAL ~~ COMMUNITY	98.65	POOR ~~AERATION	12.35	PRACTICAL ~~ APPLICATION	68.06	REGULAR ~~ VISIT	12.12

POLITICAL ~~ ECONOMY	90.08	TOTAL ~~ COST	208.41	PRACTICAL ~~ TESTING	49.15	REGULAR ~~ RHYME	9.52
POLITICAL ~~ SCIENTIST	89.95	TOTAL ~~ NUMBER	141.43	PRACTICAL ~~ REASON	43.46	REGULAR ~~ CUSTOMER	8.8
POLITICAL ~~ STABILITY	75.13	TOTAL ~~ ENERGY	97.22	PRACTICAL ~~ IMPLICATION	31.25	REGULAR ~~ SCHEME	6.65
POLITICAL ~~ OBLIGATION	73.32	TOTAL ~~ ASSET	65.14	PRACTICAL ~~ WISDON	27.4	REGULAR ~~ OBSERVATION	6.31
POLITICAL ~~ CULTURE	62.43	TOTAL ~~ ACIDITY	54.88	PRACTICAL ~~ SESSION	27.34	BROAD ~~ RANGE	71.06
POLITICAL ~~ PHILOSOPHY	58.23	TOTAL ~~ HEAD	53.08	PRACTICAL ~~ BENEFIT	25.6	BROAD ~~ SPECTRUM	44.72
POLITICAL ~~ INSTABILITY	57.67	TOTAL ~~ DRIFT	51.68	PRACTICAL ~~ RESULT	24.74	BROAD ~~ SENSE	32
POLITICAL ~~ IDEOLOGY	57.01	TOTAL ~~ OUTPUT	51.12	PRACTICAL ~~ DESIGN	19.13	BROAD ~~ PEAK	26.46
POLITICAL ~~ AGENDA	56.73	TOTAL ~~ REPLACEMENT	48.48	PRACTICAL ~~ DIFFICULTY	18.29	BROAD ~~ CATEGORY	23.4
POLITICAL ~~ CLIMATE	52.28	TOTAL ~~ REVENUE	43.81	PRACTICAL ~~ SCHEDULE	16.04	BROAD ~~ BASE	20.85
POLITICAL ~~ SPHERE	44.21	TOTAL ~~ SALE	43.6	PRACTICAL ~~ SOLUTION	12.08	BROAD ~~ DEFINITION	16.5
POLITICAL ~~ PROCESS	40.99	TOTAL ~~ EXPENDITURE	43.27	PRACTICAL ~~ HANDBOOK	11.3	BROAD ~~ GENERALIZATION	13.45
POLITICAL ~~ SITUATION	37.55	TOTAL ~~ AMOUNT	70.91	PRACTICAL ~~ ISSUE	11.23	BROAD ~~ SCOPE	12.65
POLITICAL ~~ ORGANIZATION	36.64	TOTAL ~~ INTAKE	33.96	PRACTICAL ~~ USE	11.01	BROAD ~~ ACRE	12.13
POLITICAL ~~ PARTICIPATION	36.31	TOTAL ~~ VOLUME	30.41	PRACTICAL ~~ PROBLEM	10.88	BROAD ~~ TERM	10.94
POLITICAL ~~ LEADER	36.09	TOTAL ~~ PRODUCTIVITY	30.27	PRACTICAL ~~ HANDOUT	10.34	BROAD ~~ AUDIENCE	10.89

ECONOMIC ~~ GROWTH	577.79	TOTAL ~~ POPULATION	29.9	PRACTICAL ~~ TERM	10.18	BROAD ~~ ENVIRONMENT	9.93
ECONOMIC ~~ DEVELOPMENT	319.63	TOTAL ~~ COUNT	22.42	PRACTICAL ~~ PURPOSE	10.03	BROAD ~~ PICTURE	9.21
ECONOMIC ~~ POLICY	162.72	TOTAL ~~ AREA	21.85	PRACTICAL ~~ ADVICE	9.22	BROAD ~~ AGRICULTURE	7.97
ECONOMIC ~~ COOPERATION	156.78	TOTAL ~~ LOSS	20.29	BIG ~~ BHANG	138.25	BROAD ~~ CONTEXT	7.22
ECONOMIC ~~ INTEGRATION	117.86	OLD ~~ AGE	86.05	BIG ~~ PROBLEM	46.84	BROAD ~~ FIELD	7.15
ECONOMIC ~~ ACTIVITY	90.96	OLD ~~ MAN	64.77	BIG ~~ IMPACT	33.12	BROAD ~~ ISSUE	5.19
ECONOMIC ~~ BENEFIT	84.01	OLD ~~ ADULT	55.5	BIG ~~ PICTURE	27.02	BROAD ~~ PRINCIPLE	5.08
ECONOMIC ~~ GLOBALISATION	73.67	OLD ~~ FISH	48.32	BIG ~~ DIFFERENCE	24.97	UNLIKELY ~~ ASCTIES	13.32
ECONOMIC ~~ REFORM	71.24	OLD ~~ LADY	34.51	BIG ~~ SPENDER	24.94	UNLIKELY ~~ DIAGNOSIS	9.48
ECONOMIC ~~ RECESSION	68.04	OLD ~~ PEOPLE	34.31	BIG ~~ PART	22.03	UNLIKELY ~~ EVENT	5.28
ECONOMIC ~~ PERFORMANCE	58.23	OLD ~~ GENERATION	23.33	BIG ~~ BUSINESS	20.85	COLD ~~ FRONT	77.93
ECONOMIC ~~ FACTOR	57.43	OLD ~~ CHILD	19.12	BIG ~~ THREAT	17.74	COLD ~~ WATER	42.81
ECONOMIC ~~ THEORY	56.13	OLD ~~ GIRL	18.38	BIG ~~ MARKET	17.29	COLD ~~ WAR	40.78
ECONOMIC ~~ CRISIS	54.19	OLD ~~ BABY	18.1	BIG ~~ CHALLENGE	17.23	COLD ~~ AIR	25.26
ECONOMIC ~~ LOSS	46.75	OLD ~~ TRADITION	16	BIG ~~ ADVANTAGE	16.06	COLD ~~ WEATHER	20.42
ECONOMIC ~~ RECOVERY	44.55	OLD ~~ SYSTEM	15.98	BIG ~~ PLAYER	13.98	COLD ~~ ADAPTATION	19.31

ECONOMIC ~~ HISTORY	38.44	OLD ~~ WORKER	15.4	BIG ~~ COMPANY	12.27	COLD ~~ MATTER	18.47
ECONOMIC ~~ STRUCTURE	36.78	OLD ~~ BOY	15.36	BIG ~~ PROFIT	11.36	COLD ~~ TEMPERATURE	11.94
ECONOMIC ~~ LIBERALIZATION	35.16	OLD ~~ TEXT	14.92	BIG ~~ SIZE	10.86	COLD ~~ CLIMATE	10.94
ECONOMIC ~~ SECURITY	31.3	OLD ~~ MALE	14.3	BIG ~~ HOUSE	9.81	COLD ~~ ENVIRONMENT	9.41
HUMAN ~~ RIGHT	909.5	OLD ~~ WOMAN	12.97	BIG ~~ ISSUE	9.77	DETAILED ~~ ANALYSIS	129.39
HUMAN ~~ BEING	365.24	OLD ~~ WORD	12.95	BIG ~~ EFFECT	8.57	DETAILED ~~ DESCRIPTION	104.79
HUMAN ~~ RESOURCE	302.04	OLD ~~ SON	12.91	BIG ~~ BRAND	7.23	DETAILED ~~ INFORMATION	66.07
HUMAN ~~ NATURE	177.18	OLD ~~ COURT	12.57	PRIMARY ~~ SCHOOL	193.47	DETAILED ~~ PLAN	25.47
HUMAN ~~ ACTIVITY	143.84	CENTRAL ~~ BANK	230.27	PRIMARY ~~ CARE	105.5	DETAILED ~~ ACCOUNT	21.98
HUMAN ~~ CAPITAL	141.55	CENTRAL ~~ GOVERNMENT	133.58	PRIMARY ~~ QUALITY	78.61	DETAILED ~~ UNDERSTANDING	18.22
HUMAN ~~ LIFE	108.38	CENTRAL ~~ CYANOSIS	87.06	PRIMARY ~~ SOURCE	76.43	DETAILED ~~ EXPLANATION	15.49
HUMAN ~~ BEHAVIOUR	103.15	CENTRAL ~~ PATCH	61.31	PRIMARY ~~ CONCERN	37.15	DETAILED ~~ SCHEMATIC	14.86
HUMAN ~~ VIOLATION	87.8	CENTRAL ~~ ROLE	57.35	PRIMARY ~~ COMPLETION	32.36	DETAILED ~~ STUDY	13.45
HUMAN ~~ BODY	87.17	CENTRAL ~~ THEME	49.07	PRIMARY ~~ OBJECTIVE	28.45	DETAILED ~~ DISCUSSION	11.95
HUMAN ~~ ERROR	70.81	CENTRAL ~~ DETERMINER	45.86	PRIMARY ~~ FUNCTION	26.1	DETAILED ~~ OBSERVATION	10.54
HUMAN ~~ MANAGEMENT	69.53	CENTRAL ~~ BANKER	41.05	PRIMARY ~~ COMMODITY	25.45	DETAILED ~~ BREAKDOWN	9.53
HUMAN ~~ HEALTH	63.24	CENTRAL ~~ CHARACTER	35.64	PRIMARY ~~ TUMOUR	22.65	DETAILED ~~ REVUE	9.08
HUMAN ~~ SKIN	56.09	CENTRAL ~~ AXIS	34.59	PRIMARY ~~ AIM	18.81	DETAILED ~~ INSIGHT	8.06

HUMAN ~~ POPULATION	46.23	CENTRAL ~~ SYSTEM	28.98	PRIMARY ~~ CAREGIVER	18.75	DETAILED ~~ PLANNING	7.26
HUMAN ~~ ACTION	43.74	CENTRAL ~~ EXCUTIVE	23.39	PRIMARY ~~ PURPOSE	17.81	DETAILED ~~ DESIGN	7.05
HUMAN ~~ MIND	42.93	CENTRAL ~~ IMPORTANCE	22.98	PRIMARY ~~ ROLE	17.52	DETAILED ~~ RESEARCH	6.58
HUMAN ~~ BRAIN	42.09	CENTRAL ~~ ADMINSTRATION	22.63	PRIMARY ~~ REASON	16.76	DETAILED ~~ EXAMINATION	6.51
HUMAN ~~ RELATION	40.11	CENTRAL ~~ TENET	22.58	PRIMARY ~~ LABOUR	16.74	DETAILED ~~ INVESTIGATION	6.18
HUMAN ~~ SOCIETY	37.02	CENTRAL ~~ AUTHORITY	20.83	PRIMARY ~~ MARKET	16.54	DETAILED ~~ DATUM	6.08
POSSIBLE ~~ CAUSE	88.23	CENTRAL ~~ EUROPE	18.67	PRIMARY ~~ GIVER	16.23	GREEN ~~ BELT	64.29
POSSIBLE ~~ REASON	80.8	CENTRAL ~~ PART	15.9	PRIMARY ~~ CHILD	15.4	GREEN ~~ MANURE	63.55
POSSIBLE ~~ SOLUTION	69.27	CENTRAL ~~ SIERRA	15.65	PRIMARY ~~ GOAL	15.14	GREEN ~~ WASTE	30.31
POSSIBLE ~~ EXPLANATION	67.47	CENTRAL ~~ PROTAGONIST	13.16	INDEPENDENT ~~ VARIABLE	218.37	GREEN ~~ HUE	27.32
POSSIBLE ~~ COMBINATION	37.22	PERSONAL ~~ EXPERIENCE	106.78	INDEPENDENT ~~ SCHOOL	53.84	GREEN ~~ TEA	26.28
POSSIBLE ~~ WAY	31.42	PERSONAL ~~ IDENTITY	72.4	INDEPENDENT ~~ SECTOR	42.75	GREEN ~~ VEGETABLE	16.16
POSSIBLE ~~ SOURCE	25.2	PERSONAL ~~ PRONOUN	67.48	INDEPENDENT ~~ STATE	21.46	GREEN ~~ SPUTUM	13.07
POSSIBLE ~~ ERROR	19.49	PERSONAL ~~ CHARACTERISTIC	51.63	INDEPENDENT ~~ INSTITUTION	14.67	GREEN ~~ HOUSE	12.54
POSSIBLE ~~ WORLD-STATE	17.53	PERSONAL ~~ JUDGMENT	40.4	INDEPENDENT ~~ SAMPLE	10.45	GREEN ~~ SOLUTION	11.78
POSSIBLE ~~ WORLD	17.5	PERSONAL ~~ INTEREST	38.39	INDEPENDENT ~~ REVUE	9.96	GREEN ~~ LIGHT	9.99

POSSIBLE ~~ CLAIM	16.56	PERSONAL ~~ INFORMATION	35.44	INDEPENDENT ~~ PERSON	8.55	GREEN ~~ LEAF	9.42
POSSIBLE ~~ EFFECT	16.17	PERSONAL ~~ RELATIONSHIP	34.28	INDEPENDENT ~~ EDUCATION	8.53	GREEN ~~ PROTEIN	8.06
POSSIBLE ~~ DIAGNOSIS	14.07	PERSONAL ~~ COMPUTER	28.59	INDEPENDENT ~~ ORGANIZATION	7.51	GREEN ~~ COLOUR	7.67
POSSIBLE ~~ WORLD-STATES	13.86	PERSONAL ~~ RESPONSIBILITY	25.49	INDEPENDENT ~~ LIVING	7.2	GREEN ~~ REVOLUTION	7.57
POSSIBLE ~~ ANSWER	13.56	PERSONAL ~~ DIPLOMACY	23.12	INDEPENDENT ~~ BANK	6.28	GREEN ~~ CROP	7.38
POSSIBLE ~~ STRUCTURE	13.05	PERSONAL ~~ OPINION	21.93	INDEPENDENT ~~ DECISION	6.2	GREEN ~~ LINE	6.74
POSSIBLE ~~ IMPLICATION	12.76	PERSONAL ~~ GLORY	21.91	INDEPENDENT ~~ WOMAN	6.17	GREEN ~~ ISSUE	6.21
POSSIBLE ~~ ROUTE	11.02	PERSONAL ~~ COMMUNICATION	21.63	INDEPENDENT ~~ LIFE	6.08	WRONG ~~ DECISION	13.15
POSSIBLE ~~ RISK	9.87	PERSONAL ~~ CARE	18.45	INDEPENDENT ~~ INDIVIDUAL	5.77	WRONG ~~ WAY	11.92
POSSIBLE ~~ OUTCOME	9.22	PERSONAL ~~ SELLING	17.28	INDEPENDENT ~~ THOUGHT	5.4	WRONG ~~ ACTION	5.89
SMALL ~~ AMOUNT	233.41	PERSONAL ~~ PREFERENCE	17.13	INDEPENDENT ~~ RESEARCHER	5.3	WRONG ~~ THING	5.13
SMALL ~~ SCALE	170.57	PERSONAL ~~ LIFE	15.7	INDEPENDENT ~~ TEMPREATURE	5.23	CONSIDERABLE ~~ AMOUNT	118.56
SMALL ~~ NUMBER	159.92	PERSONAL ~~ AUTONOMY	15.64	INDEPENDENT ~~ EXISTENCE	5.12	CONSIDERABLE ~~ VARIATION	19.16
SMALL ~~ SIZE	130.64	PERSONAL ~~ LEVEL	14.73	ORIGINAL ~~ TEXT	34.49	CONSIDERABLE ~~ DIFFERENCE	18.94
SMALL ~~ FIRM	127.93	WIDE ~~ RANGE	805.53	ORIGINAL ~~ JEANS	30.92	CONSIDERABLE ~~ NUMBER	15.83
SMALL ~~ COMPANY	75.21	WIDE ~~ VARIETY	181.96	ORIGINAL ~~ POSITION	30.01	CONSIDERABLE ~~ COST	11.87

SMALL ~~ SAMPLE	68.96	WIDE ~~ CONTEXT	75.59	ORIGINAL ~~ ITALIC	23.86	CONSIDERABLE ~~ DEBATE	10.69
SMALL ~~ INTESTINE	63.19	WIDE ~~ AUDIENCE	58.56	ORIGINAL ~~ MESSAGE	22.24	CONSIDERABLE ~~ DISCRETION	9.87
SMALL ~~ GROUP	62.2	WIDE ~~ SCOPE	49.09	ORINGINAL ~~ CAR	22.16	CONSIDERABLE ~~ EFFORT	8.9
SMALL ~~ BOWEL	61.62	WIDE ~~ COMMUNITY	38.75	ORIGINAL ~~ SOURCE	20.13	CONSIDERABLE ~~ IMPACT	8.71
SMALL ~~ BUSINESS	60.84	WIDE ~~ SOCIETY	27.33	ORIGINAL ~~ SPECIFICATION	19.21	CONSIDERABLE ~~ SCOPE	7.78
SMALL ~~ MAMMAL	54.55	WIDE ~~ SCALE	19.45	ORIGINAL ~~ SAMPLE	18.94	CONSIDERABLE ~~ TIME	7.78
SMALL ~~ CHANGE	47.8	WIDE ~~ DISCRETION	18.99	ORIGINAL ~~ SIN	18.44	CONSIDERABLE ~~ CONTRIBUTION	7.6
SMALL ~~ PROPORTION	43.49	WIDE ~~ SPREAD	15.44	ORIGINAL ~~ IMAGE	17.68	CONSIDERABLE ~~ EXTENT	7.51
SMALL ~~ BREVILLE	37.4	WIDE ~~ SHOT	13.73	ORIGINAL ~~ LENGTH	14.83	CONSIDERABLE ~~ BENEFIT	6.6
SMALL ~~ FARMER	37.13	WIDE ~~ PICTURE	13.39	ORIGINAL ~~ QUESTION	14.27	CONSIDERABLE ~~ WEIGHT	6.16
SMALL ~~ OBSTRUCTION	36.97	WIDE ~~ PERSPECTIVE	11.99	ORIGINAL ~~ DESIGN	13.25	CONSIDERABLE ~~ INCREASE	6.15
SMALL ~~ QUANTITY	35.52	WIDE ~~ AREA	11.75	ORIGINAL ~~ PRODUCTION	13.32	CONSIDERABLE ~~ POWER	6.13
SMALL ~~ POPULATION	33.96	WIDE ~~ DISTRIBUTION	11.5	ORIGINAL ~~ FROWARD	12.23	CONSIDERABLE ~~ INVESMENT	6.05
SMALL ~~ DELIVERY	33.39	WIDE ~~ ISSUE	11.3	ORIGINAL ~~ VIRUS	11.81	CONSIDERABLE ~~ DEGREE	5.85
AVAILABLE ~~ URL	103.4	WIDE ~~ ARRAY	10.99	ORIGINAL ~~ HYPOTHESIS	11.33	CONSIDERABLE ~~ INTEREST	5.85
AVAILABLE ~~ RESOURCE	40.16	WIDE ~~ COVERAGE	10.43	ORIGINAL ~~ MYTH	11.26	TECHNICAL ~~ SKILL	35.79
AVAILABLE ~~ DATUM	15.62	WIDE ~~ NETWORK	10.38	ORIGINAL ~~ IDEA	10.6	TECHNICAL ~~ KNOWLEDGE	19.25
AVAILABLE ~~ EVIDENCE	13.16	WIDE ~~ SPECTRUM	8.99	WHITE ~~ MAN	54.25	TECHNICAL ~~ PAPER	18.34

AVAILABLE ~~ EVERYONE	11.77	DIRECT ~~ TECHNIQUE	82.48	WHITE ~~ WINE	52.31	TECHNICAL ~~ REGULATION	17.18
AVAILABLE ~~ SPACE	11.36	DIRECT ~~ INVESTMENT	80.57	WHITE ~~ LUPIN	47	TECHNICAL ~~ EXPERTISE	17.14
AVAILABLE ~~ INFORMATION	8.57	DIRECT ~~ EFFECT	77.28	WHITE ~~ SERVANT	35.22	TECHNICAL ~~ PROGRESS	16.78
AVAILABLE ~~ FEEDBACK	7.75	DIRECT ~~ CONTACT	74.82	WHITE ~~ NOISE	33.52	TECHNICAL ~~ ASSISTANCE	16.32
AVAILABLE ~~ SOCIOLOGY	7.45	DIRECT ~~ INVOLVEMENT	71.71	WHITE ~~ LIGHT	30.13	TECHNICAL ~~ DETAIL	15.55
AVAILABLE ~~ MATERIAL	7.43	DIRECT ~~ MARKETING	54.14	WHITE ~~ SUPERMACIST	26.24	TECHNICAL ~~ KNOW-HOW	14.68
AVAILABLE ~~ OPTION	7.3	DIRECT ~~ CONTRAST	37.78	WHITE ~~ PAPER	24.54	TECHNICAL ~~ PUBLISHING	14.09
AVAILABLE ~~ ANYONE	6.86	DIRECT ~~ HEATING	37.38	WHITE ~~ POWDER	22.05	TECHNICAL ~~ ASPECT	13.72
AVAILABLE ~~ NUTRIENT	6.76	DIRECT ~~ SPEECH	35.35	WHITE ~~ CELL	21.43	TECHNICAL ~~ CHALLENGE	11.05
AVAILABLE ~~ CONSUMER	6.45	DIRECT ~~ CONSEQUENCE	33	WHITE ~~ MATTER	21.18	TECHNICAL ~~ DEPARTMENT	10.76
AVAILABLE ~~ SERVER	6.45	DIRECT ~~ COMPARISON	31.91	WHITE ~~ FISH	20.68	TECHNICAL ~~ BARRIER	10.43
AVAILABLE ~~ FARMER	5.79	DIRECT ~~ ACTION	31.45	WHITE ~~ MALE	19.54	TECHNICAL ~~ COMMUNITY	9.29
AVAILABLE ~~ CAPACITY	5.29	DIRECT ~~ LINK	27.04	WHITE ~~ MESTIZO	18.29	TECHNICAL ~~ SUPPORT	9.28
SIGNIFICANT ~~ DIIFERENCE	359.09	DIRECT ~~ EMPLOYEE	24.45	WHITE ~~ COLLAR	17.91	TECHNICAL ~~ EFFICIENCY	9.17
SIGNIFICANT ~~ IMPACT	153.2	DIRECT ~~ RESULT	23.37	WHITE ~~ POPULATION	17.66	TECHNICAL ~~ TERM	8.55
SIGNIFICANT ~~ EFFECT	150.02	DIRECT ~~ MILK	20.57	WHITE ~~ SHIRT	16.7	TECHNICAL ~~ DIFFICULTY	7.03

SIGNIFICANT ~~ CHANGE	96.37	DIRECT ~~ CONFLICT	17.77	WHITE ~~ BLOOD	16.15	TECHNICAL ~~ PROBLEM	6.43
SIGNIFICANT ~~ ROLE	86.75	DIRECT ~~ RELATIONSHIP	17.47	WHITE ~~ CLASS	14.7	DEEP ~~ FOUCS	111.23
SIGNIFICANT ~~ IMPROVEMENT	67.7	DIRECT ~~ IMPACT	17.19	WHITE ~~ PLASTER	14.59	DEEP ~~ UNDERSTANDING	71.9
SIGNIFICANT ~~ AMOUNT	53.49	DIRECT ~~ CORRELATION	16.71	PAST ~~ YEAR	303.56	DEEP ~~ THROMBOSIS	35.79
SIGNIFICANT ~~ FACTOR	48.05	EFFECTIVE ~~ WAY	93.13	PAST ~~ HISTORY	248.67	DEEP ~~ STRUCTURE	31.78
SIGNIFICANT ~~ CORRELATION	38.43	EFFECTIVE ~~ COMMUNICATION	77.35	PAST ~~ DECADE	180.51	DEEP ~~ VEIN	25.59
SIGNIFICANT ~~ CONTRIBUTION	34.49	EFFECTIVE ~~ MANAGEMENT	38.1	PAST ~~ MEDICAL	99.72	DEEP ~~ MEANING	16.2
SIGNIFICANT ~~ FAMILY	32.37	EFFECTIVE ~~ MARKETING	33.11	PAST ~~ EXPERIENCE	71.39	DEEP ~~ SEA	14.12
SIGNIFICANT ~~ HISTORY	30.73	EFFECTIVE ~~ STRATEGY	32.6	PAST ~~ MONTH	70.8	DEEP ~~ SHOT	12.94
SIGNIFICANT ~~ PROPORTION	29.94	EFFECTIVE ~~ TOOL	27.49	PAST ~~ PRESENT	40.56	DEEP ~~ X-RAY	12.75
SIGNIFICANT ~~ INCREASE	29.92	EFFECTIVE ~~ METHOD	23.95	PAST ~~ PERFORMANCE	38.81	DEEP ~~ EXPOSURE	11.7
SIGNIFICANT ~~ REDUCTION	26.01	EFFECTIVE ~~ MEANS	20.62	PAST ~~ WEEK	36.18	DEEP ~~ WATER	11.48
SIGNIFICANT ~~ EVIDENCE	24.66	EFFECTIVE ~~ TREATMENT	16.18	PAST ~~ EVENT	30.69	DEEP ~~ LEVEL	9.75
SIGNIFICANT ~~ PART	24.36	EFFECTIVE ~~ PLANNING	15.91	PAST ~~ PARTICIPLE	24.63	DEEP ~~ ROOT	9.14
SIGNIFICANT ~~ FINDING	23.98	EFFECTIVE ~~ GOVERNANCE	14.86	PAST ~~ CENTURY	18.82	DEEP~~ SOIL	8.34
SIGNIFICANT ~~ NUMBER	23.55	EFFECTIVE ~~ LEADERSHIP	14.46	PAST ~~ FORM	15.82	DEEP ~~ INSIGHT	8.12

SIGNIFICANT ~~ INFLUENCE	19.23	EFFECTIVE ~~ COMMUINCATOR	12.69	PAST ~~ SOCIETY	12.82	DEEP ~~ ION	7.84
SIMILAR ~~ WAY	76.79	EFFECTIVE ~~ SERVICE	11.86	PAST ~~ TENSE	12.38	DEEP ~~ RED	7.51
SIMILAR ~~ PATTERN	64.73	EFFECTIVE ~~ VACCINE	10.99	PAST ~~ YIELD	10.88	DEEP ~~ CONNECTION	6.93
SIMILAR ~~ VEIN	37.9	EFFECTIVE ~~ PRESSURE	10	PAST ~~ INJUSTICE	10.18	DEEP ~~ LAYER	6.69
SIMILAR ~~ SITUATION	36.92	EFFECTIVE ~~ SYSTEM	9.91	PAST ~~ VERB	7.75	DEEP ~~ SPACE	5.95
SIMILAR ~~ RESULT	35.51	EFFECTIVE ~~ USE	9.88	PAST ~~ DATUM	7.17	CHEAP ~~ LABOUR	49.59
SIMILAR ~~ MANNER	34.3	EFFECTIVE ~~ POLICY	9.54	PAST ~~ VEGETATION	6.2	CHEAP ~~ OPTION	19.32
SIMILAR ~~ SIZE	26.94	EFFECTIVE ~~ RESPONSE	9.29	AVERAGE ~~ RATE	50.69	CHEAP ~~ PRICE	17.9
SIMILAR ~~ FASHION	25.95	LATE ~~ STAGE	124.5	AVERAGE ~~ VALUE	46.63	CHEAP ~~ FOOD	17.5
SIMILAR ~~ PROBLEM	24.24	LATE ~~ TH	96.41	AVERAGE ~~ COST	33.48	CHEAP ~~ IMPORT	16.23
SIMILAR ~~ CHARACTERISTIC	19.79	LATE ~~ CENTURY	93.3	AVERAGE ~~ PRICE	30.13	CHEAP ~~ FLIGHT	13.34
SIMILAR ~~ TREND	18.51	LATE ~~ BRONZE	61.82	AVERAGE ~~ GROWTH	27.86	CHEAP ~~ MATERIAL	11.83
SIMILAR ~~ EFFECT	17.31	LATE ~~ DATE	56.44	AVERAGE ~~ TIME	27.05	CHEAP ~~ ALTERNATIVE	11.05
SIMILAR ~~ PRODUCT	14.61	LATE ~~ AGE	51.23	AVERAGE ~~ EXPENDITURE	26.56	CHEAP ~~ COST	8.67
SIMILAR ~~ PROPERTY	14.18	LATE ~~ LIFE	36.32	AVERAGE ~~ PERSON	26.03	CHEAP ~~ EGG	8.47
SIMILAR ~~ TECHNIQUE	14.06	LATE ~~ AGES	26.61	AVERAGE ~~ TEMPREATURE	25.77	CHEAP ~~ GOOD	6.8
SIMILAR ~~ STRUCTURE	11.86	LATE ~~ PERIOD	23.52	AVERAGE ~~ SIZE	24.73	CHEAP ~~ RATE	5.18
SIMILAR ~~ INTEREST	10.67	LATE ~~ MIDDLE	21.56	AVERAGE ~~ VELOCITY	23.65	SURE ~~ MERRY-MACABRE	9.83

SIMILAR ~~ POSITION	10.44	LATE ~~ DEVELOPMENT	20.94	AVERAGE ~~ LENGTH	23.1	INTERESTED ~~ PARTY	25.8
SIMILAR ~~ SYMPTOM	10.36	LATE ~~ YEAR	19.88	AVERAGE ~~ INCOME	22.37	INTERESTED ~~ INTERESTEDPARTICIPATING	13.9
SIMILAR ~~ BACKGROUND	10.22	LATE ~~ TECHNOLOGY	15.57	AVERAGE ~~ AGE	21.82	INTERESTED ~~ READER	5.26
CERTAIN ~~ EXTENT	250.32	LATE ~~ TEEN	13.41	AVERAGE ~~ INFLATION	20.53	NUCLEAR ~~ WEAPON	114.68
CERTAIN ~~ DEGREE	126.6	LATE ~~ PREGNANCY	12.5	AVERAGE ~~ NUMBER	20.4	NUCLEAR ~~ POWER	53.22
CERTAIN ~~ AMOUNT	91.27	LATE ~~ TIME	11.84	AVERAGE ~~ EXTENSION	18.09	NUCLEAR ~~ DNA	51.93
CERTAIN ~~ AREA	84.66	LATE ~~ SECTION	11.23	AVERAGE ~~ LEVEL	16.72	NUCLEAR ~~ SPIN	31.61
CERTAIN ~~ LEVEL	67.71	LATE ~~ SPRING	9.71	AVERAGE ~~ LOAD	15.85	NUCLEAR ~~ WASTE	30.62
CERTAIN ~~ ASPECT	59.53	LATE ~~ PART	9.67	AVERAGE ~~ MARK	14.61	NUCLEAR ~~ FISSION	26.77
CERTAIN ~~ CIRCUMSTANCE	58.39	LATE ~~ FLASH	9.39	POPULAR ~~ CULTURE	51.48	NUCLEAR ~~ FAMILY	26.43
CERTAIN ~~ TYPE	58	SUCCESSFUL ~~ IMPLEMENTATION	53.65	POPULAR ~~ SUPPORT	44.16	NUCLEAR ~~ ENERGY	20.96
CERTAIN ~~ CONDITION	47.61	SUCCESSFUL ~~ MARKETING	28.5	POPULAR ~~ OPINION	37.63	NUCLEAR ~~ GLASS	18.11
CERTAIN ~~ GROUP	43.12	SUCCESSFUL ~~ STORY	26.48	POPULAR ~~ MOVEMENT	23.26	NUCLEAR ~~ FORCE	17.48
CERTAIN ~~ WAY	37.6	SUCCESSFUL ~~ BRAND	26.34	POPULAR ~~ BELIEF	20.98	NUCLEAR ~~ FUEL	17.43
CERTAIN ~~ SITUATION	35.23	SUCCESSFUL ~~ OUTCOME	24.97	POPULAR ~~ DISCOURSE	15.67	NUCLEAR ~~ BOMB	12.34
CERTAIN ~~ CHARACTERISTIC	34.58	SUCCESSFUL ~~ STRATEGY	24.91	POPULAR ~~ TOURIST	13.8	NUCLEAR ~~ CAPABILITY	12.13
CERTAIN ~~ SPECIE	25.59	SUCCESSFUL ~~ PRESIDENT	22.31	POPULAR ~~ CONSCIOUSNESS	11.33	NUCLEAR ~~ PROGRAMME	11.91

CERTAIN ~~ POINT	24.76	SUCCESSFUL ~~ MANAGER	18	POPULAR ~~ CHILD	11.19	NUCLEAR ~~ GENE	11.63
CERTAIN ~~ WORD	23.65	SUCCESSFUL ~~ BUSINESS	15.73	POPULAR ~~ ATTRACTION	11.06	NUCLEAR ~~ GENOME	10.36
CERTAIN ~~ TRAIT	23.65	SUCCESSFUL ~~ COMPLETION	15.62	POPULAR ~~ DESTINATION	9.81	NUCLEAR ~~ WAR	10.33
CERTAIN ~~ FEATURE	19.91	SUCCESSFUL ~~ LEARNING	13.47	POPULAR ~~ MUSIC	9.61	NUCLEAR ~~ CRISIS	10.1
CERTAIN ~~ TIME	19.37	SUCCESSFUL ~~ OPERATION	12.91	POPULAR ~~ REBELLION	9.31	NUCLEAR ~~ FRACTION	9.9
CERTAIN ~~ LIMITATION	16.18	SUCCESSFUL ~~ COMPANY	12.57	POPULAR ~~ RESORT	8.94	NUCLEAR ~~ ARM	9.06
COMMON ~~ LAW	439.51	SUCCESSFUL ~~ APPLICATION	10.35	POPULAR ~~ GENRE	8.47	PERFECT ~~ COMPETITION	96.05
COMMON ~~ SENSE	112.58	SUCCESSFUL ~~ MANAGEMENT	9.23	POPULAR ~~ SPORT	8.03	PERFECT ~~ SUBSTITUTE	20.12
COMMON ~~ FEATURE	78.6	SUCCESSFUL ~~ LAUNCH	8.54	POPULAR ~~ CATEGORY	7.93	PERFECT ~~ EXAMPLE	14.06
COMMON ~~ GROUND	62.42	SUCCESSFUL ~~ ENTITY	7.52	POPULAR ~~ IMAGE	7.51	PERFECT ~~ SENSE	10.87
COMMON ~~ ANCESTOR	55.94	SUCCESSFUL ~~ INTEGRATION	7.47	POPULAR ~~ THEME	7.22	PERFECT ~~ MONITORING	9.99
COMMON ~~ CURRENCY	53.77	SUCCESSFUL ~~ DEVELOPMENT	7.31	POPULAR ~~ METHOD	6.26	PERFECT ~~ CONVERGENCE	9.6
COMMON ~~ CAUSE	48.99	SUCCESSFUL ~~ TEAM	6.77	RELIGIOUS ~~ MOVEMENT	114.22	PERFECT ~~ KNOWLEDGE	8.59
COMMON ~~ INTEREST	45.22	TRADITIONAL ~~ VIEW	66.14	RELIGIOUS ~~ BELIEF	113.74	PERFECT ~~ MODEL	8.15
COMMON ~~ FACTOR	39.17	TRADITIONAL ~~ METHOD	57.21	RELIGIOUS ~~ PRACTICE	76.83	PERFECT ~~ BEING	6.88
COMMON ~~ PRACTICE	38.36	TRADITIONAL ~~ COMPETITOR	45.72	RELIGIOUS ~~ FESTIVAL	43.26	PERFECT ~~ FIT	6.57

COMMON ~~ IDENTITY	36.19	TRADITIONAL ~~ APPROACH	43.05	RELIGIOUS ~~ SIGNIFICANCE	29.41	PERFECT ~~ CODE	6.24
COMMON ~~ THEME	32.35	TRADITIONAL ~~ FORM	38.96	RELIGIOUS ~~ OUTLOOK	24.48	PERFECT ~~ FORM	5.79
COMMON ~~ LAWYER	30.71	TRADITIONAL ~~ PRACTICE	31.84	RELIGIOUS ~~ CEREMONY	20.24	PERFECT ~~ WORLD	5.36
COMMON ~~ GOAL	30.4	TRADITIONAL ~~ ROLE	25.27	RELIGIOUS ~~ THINKER	18.52	HEAVY ~~ DRINKER	81.01
COMMON ~~ RULE	30.31	TRADITIONAL ~~ ROMAN	21.41	RELIGIOUS ~~ FREEDOM	18.31	HEAVY ~~ CHAIN	60.23
COMMON ~~ MARKET	24.1	TRADITIONAL ~~ CULTURE	21.28	RELIGIOUS ~~ GROUP	17.94	HEAVY ~~ BURDEN	43.36
COMMON ~~ HUMANITY	22.95	TRADITIONAL ~~ THEORY	20.82	RELIGIOUS ~~ TEXT	15.41	HEAVY ~~ RAINFALL	36.05
COMMON ~~ PROBLEM	21.32	TRADITIONAL ~~ SENSE	20.41	RELIGIOUS ~~ DOUBT\	13.26	HEAVY ~~ INDUSTRY	22.85
COMMON ~~ METHOD	21.26	TRADITIONAL ~~ PRODUCT	19.18	RELIGIOUS ~~ OBSERVANCE	12.74	HEAVY ~~ BUYER	22.09
COMMON ~~ ENEMY	19.72	TRADITIONAL ~~ MODEL	19.13	RELIGIOUS ~~ RITUAL	12.5	HEAVY ~~ RAIN	18.48
LIKELY ~~ CAUSE	111.69	TRADITIONAL ~~ VALUE	18.63	RELIGIOUS ~~ CELEBRATION	11.07	HEAVY ~~ WATCHER	14.64
LIKELY ~~ DIAGNOSIS	40.95	TRADITIONAL ~~ SONNET	16.31	RELIGIOUS ~~ PRINCIPLE	10.87	HEAVY ~~ RELIANCE	13.84
LIKELY ~~ SOURCE	10.64	TRADITIONAL ~~ HIERARCHY	15.97	RELIGIOUS ~~ INSTITUTION	10.68	HEAVY ~~ CRITICISM	12.46
LIKELY ~~ FELICITY	10.33	TRADITIONAL ~~ COMMUNITY	15.67	RELIGIOUS ~~ COMMUNITY	10.47	HEAVY ~~ FINE	11.68
LIKELY ~~ EXPLANATION	7.81	TRADITIONAL ~~ WAY	15.14	RELIGIOUS ~~ AUTHORITY	10.23	HEAVY ~~ PRESSURE	11.25
EARLY ~~ STAGE	370.53	TRADITIONAL ~~ PUB	13.68	RELIGIOUS ~~ LAW	9.82	HEAVY ~~ METAL	10.79

EARLY ~~ CENTURY	188.35	TRADITIONAL ~~ NOTION	13.62	HARD ~~ WORK	35.43	HEAVY ~~ USER	10.76
EARLY ~~ PERIOD	102.26	USEFUL ~~ TOOL	182.65	HARD ~~ BRAND	23.18	HEAVY ~~ EMPHASIS	10.21
EARLY ~~ AGE	70.75	USEFUL ~~ INFORMATION	46.72	HARD ~~ COPY	22.27	HEAVY ~~ GRAZING	9.78
EARLY ~~ CHILDHOOD	64.71	USEFUL ~~ INSIGHT	17.76	HARD ~~ TIME	19.5	HEAVY ~~ SOIL	9.15
EARLY ~~ ENGLAND	62.17	USEFUL ~~ SOURCE	17.72	HARD ~~ WATER	17.05	HEAVY ~~ LOSS	7.84
EARLY ~~ TH	53.04	USEFUL ~~ METHOD	14.76	HARD ~~ BITE	16.69	HEAVY ~~ INTERVENTION	7.13
EARLY ~~ YEAR	42.39	USEFUL ~~ TECHNIQUE	12.7	HARD ~~ DRIVE	16.46	HEAVY ~~ INVESTMENT	6.47
EARLY ~~ PHASE	32.9	USEFUL ~~ WAY	8.57	HARD ~~ SPHERE	15.74	JOINT ~~ VENTURE	199.29
EARLY ~~ DEVELOPMENT	29.52	USEFUL ~~ PURPOSE	8.52	HARD ~~ HAMMER	12.32	JOINT ~~ ENTERPRISE	31.21
EARLY ~~ EMBYRO	28.14	USEFUL ~~ INDICATOR	6.19	HARD ~~ EVIDENCE	7.43	JOINT ~~ PAIN	26.06
EARLY ~~ WORK	25.32	USEFUL ~~ LIFE	5.97	HARD ~~ INDEX	7.1	JOINT ~~ SPACE	17.39
EARLY ~~ EUROPE	24.16	USEFUL ~~ DEFINITION	5.38	COMPLETE ~~ TASK	89.87	JOINT ~~ SIGNIFICANCE	16.81
EARLY ~~ REFORMATION	23.49	USEFUL ~~ KNOWLEDGE	5.16	COMPLETE ~~ ESSAYS	75.55	JOINT ~~ DECISION	11.76
EARLY ~~ FLOCK	22.84	FOREIGN ~~ INVESTMENT	250.8	COMPLETE ~~ QUESTIONNAIRE	51.61	JOINT ~~ REPLACEMENT	9.6
EARLY ~~ DATE	22.17	FOREIGN ~~ POLICY	207.12	COMPLETE ~~ ESSAY	30.69	JOINT ~~ DISEASE	7.15
EARLY ~~ MORNING	21.34	FOREIGN ~~ EXCHANGE	151.52	COMPLETE ~~ TILING	28.4	SUBSTANTIAL ~~ AMOUNT	54.41
EARLY ~~ HOMINID	20.68	FOREIGN ~~ ACCENT	127.27	COMPLETE ~~ PICTURE	20.89	SUBSTANTIAL ~~ CONTRIBUTION	16

EARLY ~~ DAY	19.64	FOREIGN ~~ RESERVE	118.77	COMPLETE ~~ PROJECT	18.22	SUBSTANTIAL ~~ NUMBER	13.32
EARLY ~~ RETIREMENT	19.38	FOREIGN ~~ CURRENCY	100.82	COMPLETE ~~ CONTRAST	13.95	SUBSTANTIAL ~~ LOSS	12.13
NATIONAL ~~ CUISINE	150.4	FOREIGN ~~ LANGUAGE	94.42	COMPLETE ~~ UNDERSTANDING	13.15	SUBSTANTIAL ~~ EVIDENCE	11.53
NATIONAL ~~ IDENTITY	128.74	FOREIGN ~~ INVESTOR	86.53	COMPLETE ~~ ASSEMBLY	13.11	SUBSTANTIAL ~~ COST	9.09
NATIONAL ~~ INTEREST	126.89	FOREIGN ~~ MARKET	75.08	COMPLETE ~~ SEQUENCE	12.21	SUBSTANTIAL ~~ CHANGE	8.59
NATIONAL ~~ GOVERNMENT	126.18	FOREIGN ~~ CAPITAL	68.28	COMPLETE ~~ GENOME	12.08	SUBSTANTIAL ~~ DIFFERENCE	8.03
NATIONAL ~~ LEVEL	84.54	FOREIGN ~~ AID	60.8	COMPLETE ~~ SENTENCE	10.57	SUBSTANTIAL ~~ BENEFIT	7.11
NATIONAL ~~ INCOME	74.32	FOREIGN ~~ COUNTRY	42.1	COMPLETE ~~ CONTROL	10.1	SUBSTANTIAL ~~ RESEARCH	5.89
NATIONAL ~~ SECURITY	60.07	FOREIGN ~~ SUBSIDIARY	39.37	COMPLETE ~~ MOISTURE	10.04	SUBSTANTIAL ~~ INCREASE	5.43
NATIONAL ~~ BOUNDARY	51.76	FOREIGN ~~ RELIGION	39.23	COMPLETE ~~ DESCRIPTION	9.69	SUBSTANTIAL ~~ DEMAND	5.11
NATIONAL ~~ COURT	50.52	FOREIGN ~~ TRADE	35.99	COMPLETE ~~ SOLUTION	9.6	SOFT ~~ NON-TENDER	103.03
NATIONAL ~~ ECONOMY	49.6	FOREIGN ~~ COMPETITION	26.63	COMPLETE ~~ FAILURE	8.85	SOFT ~~ TISSUE	99.84
NATIONAL ~~ PRIDE	43.62	FOREIGN ~~ COMPANY	24.88	COMPLETE ~~ SKELTON	8.56	SOFT ~~ DRINK	47.8
NATIONAL ~~ BORDER	35.59	FOREIGN ~~ FIRM	23.39	COMPLETE ~~ ELIMINATION	8.53	SOFT ~~ FRUIT	44.41
NATIONAL ~~ CULTURE	33.13	FOREIGN ~~ RELATION	20.92	OVERALL ~~ PERFORMANCE	66.43	SOFT ~~ INJURY	31.67
NATIONAL ~~ LAW	31.22	FOREIGN ~~ DEBT	18.6	OVERALL ~~ EFFICIENCY	37.23	SOFT ~~ CHEESE	23.13

NATIONAL ~~ POLICY	28.34	FINAL ~~ PRODUCT	95.86	OVERALL ~~ PROJECT	27.46	SOFT ~~ WATER	15.78
NATIONAL ~~ AUTHORITY	27.69	FINAL ~~ STAGE	80.29	OVERALL ~~ SIGNIFICANCE	22.39	SOFT ~~ BRAND	14.72
NATIONAL ~~ SOVEREIGNTY	25.27	FINAL ~~ VALUE	78.85	OVERALL ~~ COST	21.86	SOFT ~~ APPROACH	14.46
NATIONAL ~~ CONSCIOUSNESS	23.25	FINAL ~~ SECTION	39.47	OVERALL ~~ EFFECT	21.82	SOFT ~~ POWER	12.78
NATIONAL ~~ GOVERNANE	21.67	FINAL ~~ DECISION	39	OVERALL ~~ SIZE	20.53	SOFT ~~ TENDER	12.22
NATIONAL ~~ PARK	21.35	FINAL ~~ RESULT	36.66	OVERALL ~~ STRATEGY	18.59	SOFT ~~ TEXTURE	11.96
CURRENT ~~ ASSET	155.73	FINAL ~~ POSITION	35.06	OVERALL ~~ PICTURE	17.77	SOFT ~~ COLOUR	10.32
CURRENT ~~ SITUATION	132.71	FINAL ~~ LINE	33.23	OVERALL ~~ SUCCESS	16.27	SOFT ~~ LAW	10.22
CURRENT ~~ TREATMENT	96.61	FINAL ~~ OUTCOME	30.07	OVERALL ~~ AIM	15.5	SOFT ~~ MARKET	7.11
CURRENT ~~ LIABILITY	95.06	FINAL ~~ STANZA	25.95	OVERALL ~~ SYSTEM	14.66	DARK ~~ MATTER	334.1
CURRENT ~~ SYSTEM	71.79	FINAL ~~ CONSONANT	25.68	OVERALL ~~ EQUATION	13.93	DARK ~~ SIDE	17.57
CURRENT ~~ RATIO	70.89	FINAL ~~ ACCEPTOR	24.88	OVERALL ~~ PROFIT	13.64	DARK ~~ SPOT	13.48
CURRENT ~~ LEVEL	43.63	FINAL ~~ SCENE	24.31	OVERALL ~~ TREND	12.54	DARK ~~ GRAIN	13.02
CURRENT ~~ SALARY	38.2	FINAL ~~ REPORT	21.72	OVERALL ~~ DESIGN	10.51	DARK ~~ ORANGE	12.24
CURRENT ~~ STATE	38.01	FINAL ~~ SOLUTION	21.01	OVERALL ~~ PROFITABILLITY	10.16	DARK ~~ FRINGE	11.03
CURRENT ~~ MEDICATION	35.32	FINAL ~~ DIVIDEND	20	OVERALL ~~ RESULT	9.79	DARK ~~ COLOUR	7.3
CURRENT ~~ INCOME	35.09	FINAL ~~ TEMPREATURE	15.59	OVERALL ~~ SCORE	9.49	DARK ~~ AREA	5.32
CURRENT ~~ MARKET	29.9	FINAL ~~ PART	15.47	OVERALL ~~ OUTPUT	9.48	VALID ~~ ARGUMENT	45.46

CURRENT ~~ TREND	27.32	FINAL ~~ EXPERIMENT	15.46	SUITABLE ~~ MATERIAL	47.87	VALID ~~ CONTRACT	14.36
CURRENT ~~ ACCOUNT	27.21	FINAL ~~ POINT	14.49	SUITABLE ~~ METHOD	18.36	VALID ~~ POINT	11.73
CURRENT ~~ CRISIS	25.81	PREVIOUS ~~ YEAR	142.19	SUITABLE ~~ CANDIDATE	13.57	VALID ~~ CRITICISM	7.5
CURRENT ~~ FLOW	23.94	PREVIOUS ~~ SECTION	125.31	SUITABLE ~~ SECTION	11.65	VALID ~~ CONSIDERATION	6.6
CURRENT ~~ POSITION	23.25	PREVIOUS ~~ STUDY	123.85	SUITABLE ~~ SITE	10.04	DEAD ~~ BODY	23.1
CURRENT ~~ ADMISSION	23.2	PREVIOUS ~~ EXPERIENCE	81.81	SUITABLE ~~ LOCATION	7.71	DEAD ~~ TIME	21.41
CURRENT ~~ VOLTAGE	22.73	PREVIOUS ~~ RESEARCH	72.34	SUITABLE ~~ OPTION	7.7	DEAD ~~ EMPEROR	13.62
CURRENT ~~ PRICE	21.97	PREVIOUS ~~ WORK	39.27	SUITABLE ~~ CONDITION	7.56	DEAD ~~ LINK	12.41
MODERN ~~ SOCIETY	276.98	PREVIOUS ~~ HISTORY	32.38	SUITABLE ~~ MARKETING	7.27	DEAD ~~ BOYFRIEND	10.46
MODERN ~~ HUMAN	96.73	PREVIOUS ~~ INVESTIGATION	26.44	SUITABLE ~~ HABITAT	7.1	DEAD ~~ ANCESTOR	7.96
MODERN ~~ DAY	89.54	PREVIOUS ~~ FILM	22.87	SUITABLE ~~ VALUE	6.98	DEAD ~~ END	6.98
MODERN ~~ ENVIRONMENTALISM	77.77	PREVIOUS ~~ LITERATURE	22.77	SUITABLE ~~ MIX	6.59	DEAD ~~ DRUNK	6.44
MODERN ~~ WORLD	71.89	PREVIOUS ~~ PARAGRAPH	22.58	SUITABLE ~~ FOOD	5.93	DEAD ~~ FATHER	5.83
MODERN ~~ ENGLAND	68.37	PREVIOUS ~~ SURGERY	21.57	SUITABLE ~~ SYSTEM	5.48	DUE ~~ LACK	55.31
MODERN ~~ PERIOD	58.55	PREVIOUS ~~ PERIOD	18.07	SUITABLE ~~ SOLUTION	5.42	DUE ~~ PROCESS	39.31
MODERN ~~ MEDICINE	57.35	PREVIOUS ~~ DECADE	16.95	SUITABLE ~~ LEVEL	5.33	DUE ~~ SAFEGUARD	26.29

MODERN ~~ TECHNOLOGY	55.03	PREVIOUS ~~ KNOWLEDGE	16.06	INTERESTING ~~ POINT	34.12	DUE ~~ DILIGENCE	16.05
MODERN ~~ LIFE	50.63	PREVIOUS ~~ ONE	15.84	INTERESTING ~~ FEATURE	31.85	DUE ~~ CHANCE	14.21
MODERN ~~ STATE	39.69	PREVIOUS ~~ EXPERIMENT	13.52	INTERESTING ~~ INSIGHT	28.52	DUE ~~ INACCURACY	10.95
MODERN ~~ SCIENCE	27.17	PREVIOUS ~~ CENTURY	13.07	INTERESTING ~~ RESULT	20.93	DUE ~~ FRICTION	8.98
MODERN ~~ AGE	25.25	PREVIOUS ~~ STROKE	11.9	INTERESTING ~~ QUESTION	17	DUE ~~ LOSS	8.53
MODERN ~~ ERA	23.03	PREVIOUS ~~ GENERATION	11.85	INTERESTING ~~ DISCUSSION	12.66	DUE ~~ GRAVITY	7.87
MODERN ~~ EUROPE	21.21	SHORT ~~ TERM	513.85	INTERESTING ~~ THING	11.58	DUE ~~ CONTAMINATION	6.38
MODERN ~~ TECHNIQUE	21.1	SHORT ~~ PERIOD	179.35	INTERESTING ~~ EXAMPLE	10.48	DUE ~~ COURSE	5.73
MODERN ~~ CAPITALISM	20.24	SHORT ~~ RUN	128.19	INTERESTING ~~ STORY	10.12	EXPENSIVE ~~ ITEM	12.02
MODERN ~~ AUDIENCE	17.38	SHORT ~~ TIME	110.46	INTERESTING ~~ EXTENSION	8.81	EXPENSIVE ~~ PRODUCT	10.21
MODERN ~~ AGRICULTURE	16.74	SHORT ~~ DISTANCE	49.21	INTERESTING ~~ FACT	8.11	EXPENSIVE ~~ LUXURY	6.9
MODERN ~~ CULTURE	16.62	SHORT ~~ WAVELENGTH	40.11	INTERESTING ~~ FINDING	7.6	EXPENSIVE ~~ MATERIAL	6.61
LONG ~~ TERM	978.88	SHORT ~~ MEMORY	34.27	INTERESTING ~~ STUDY	7.08	EXPENSIVE ~~ BREAD-MAKERS	6.31
LONG ~~ PERIOD	299.83	SHORT ~~ DURATION	27.91	INTERESTING ~~ TOPIC	6.54	EXPENSIVE ~~ DRUG	5.23
LONG ~~ RUN	293.86	SHORT ~~ LIFE	23.18	INTERESTING ~~ ASPECT	5.68	SLOW ~~ RATE	67
LONG ~~ TIME	230.7	SHORT ~~ PATH	22.71	INTERESTING ~~ ARGUMENT	5.4	SLOW ~~ PACE	38.88

LONG ~~ BARROW	90.41	SHORT ~~ BREATH	21.71	INTERESTING ~~ IDEA	5.37	SLOW ~~ PROCESS	19.11
LONG ~~ HOUR	59.24	SHORT ~~ SPAN	21.01	RESPONIBLE ~~ ECOTOURISM	8.6	SLOW ~~ GROWTH	16.96
LONG ~~ WAY	54.11	SHORT ~~ GRASS	19.52	CORRECT ~~ VALUE	21.23	SLOW ~~ RESPONSE	16.75
LONG ~~ WAVELENGTH	51.98	SHORT ~~ LENGTH	17.08	CORRECT ~~ ANSWER	20.38	SLOW ~~ SPEED	16.14
LONG ~~ DISTANCE	49.58	SHORT ~~ VOWEL	14.77	CORRECT ~~ CODE	17.87	SLOW ~~ PROGRESS	15.15
LONG ~~ BONE	47.37	SHORT ~~ NOTICE	13.1	CORRECT ~~ LOCALIZATION	15.92	SLOW ~~ FLUID	11.28
LONG ~~ HISTORY	42.4	SHORT ~~ SENTENCE	12.94	CORRECT ~~ ERROR	15.08	SLOW ~~ WAVE	11.11
LONG ~~ VOWEL	31.06	SHORT ~~ CYCLE	12.82	CORRECT ~~ ORDER	13.02	SLOW ~~ SLEEP	8.75
LONG ~~ TAIL	27.18	SHORT ~~ HAMSTRING	11.68	CORRECT ~~ SEGREGATION	10.64	SLOW ~~ RECOVERY	8.63
LONG ~~ DURATION	26.97	SHORT ~~ CULTIVAR	11.52	CORRECT ~~ OUTPUT	10.16	SLOW ~~ MOVEMENT	8.4
LONG ~~ STANDING	21.78	APPROPRIATE ~~ LEVEL	34.2	CORRECT ~~ AMOUNT	9.08	SLOW ~~ MOTION	7.92
LONG ~~ LIFE	21.52	APPROPRIATE ~~ METHOD	25.23	CORRECT ~~ INTERPRETATION	8.34	SLOW ~~ DEVELOPMENT	7.57
LONG ~~ SUCCESS	19.39	APPROPRIATE ~~ RESPONSE	17.96	CORRECT ~~ TEMPREATURE	8.13	SLOW ~~ REACTION	7.07
LONG ~~ TRADITION	19.24	APPROPRIATE ~~ MEASURE	16.47	CORRECT ~~ DEFICIENCY	8.01	SLOW ~~ RELEASE	6.96
LONG ~~ MEMORY	18.27	APPROPRIATE ~~ STRATEGY	14.01	CORRECT ~~ DEFINITION	7.86	SLOW ~~ FLOW	6.36
LONG ~~ CHAIN	15.63	APPROPRIATE ~~ TECHNIQUE	12.48	CORRECT ~~ MISTAKE	7.2	FRESH ~~ FISH	87.24
STRONG ~~ CORRELATION	64.52	APPROPRIATE ~~ TIME	12.33	CORRECT ~~ BALANCE	6.94	FRESH ~~ FRUIT	66.89

STRONG ~~ EVIDENCE	50.83	APPROPRIATE ~~ WAY	11.94	CORRECT ~~ ACTION	5.96	FRESH ~~ FOOD	55.54
STRONG ~~ RELATIONSHIP	47.12	APPROPRIATE ~~ CHOICE	11.47	CORRECT ~~ COUNT	5.72	FRESH ~~ WATER	49.65
STRONG ~~ ARGUMENT	47.03	APPROPRIATE ~~ LAB	11.45	CORRECT ~~ FORM	5.52	FRESH ~~ MEAT	33.79
STRONG ~~ POSITION	46.57	APPROPRIATE ~~ ADULT	10.99	CORRECT ~~ RESPONSE	5.43	FRESH ~~ WEIGHT	31.02
STRONG ~~ SENSE	42.41	APPROPRIATE ~~ MOTIVE	10.85	CORRECT ~~ TYPE	5.38	FRESH ~~ BREATH	17.13
STRONG ~~ CULTURE	33.17	APPROPRIATE ~~ MANAGEMENT	10.59	ADDITIONAL ~~ PROTOCOL	57.86	FRESH ~~ PURCHASING	15.32
STRONG ~~ BOND	32.46	APPROPRIATE ~~ METHODOLOGY	10.53	ADDITIONAL ~~ SHIFT	44.5	FRESH ~~ AIR	15.93
STRONG ~~ DESIRE	27.43	APPROPRIATE ~~ SIZE	10.48	ADDITIONAL ~~ COST	39.39	FRESH ~~ BREAD	11.57
STRONG ~~ EMPHASIS	26.37	APPROPRIATE ~~ VALUE	10.21	ADDITIONAL ~~ INFORMATION	23.95	FRESH ~~ BLOOD	9.98
STRONG ~~ OPPOSITION	26.34	APPROPRIATE ~~ BEHAVIOUR	9.47	ADDITIONAL ~~ TRAINING	18.52	FRESH ~~ PRODUCE	8.08
STRONG ~~ INFLUENCE	25.08	APPROPRIATE ~~ APPROACH	9.29	ADDITIONAL ~~ VARIABLE	16.69	FRESH ~~ START	6.03
STRONG ~~ TIE	24.15	APPROPRIATE ~~ AMOUNT	9.22	ADDITIONAL ~~ BENEFIT	16.05	TOP ~~ MANAGEMENT	112.62
STRONG ~~ LINK	23.78	APPROPRIATE ~~ TREATMENT	9.07	ADDITIONAL ~~ SHEAR	15.95	TOP ~~ BOTTOM	64.7
STRONG ~~ FORCE	23.45	POTENTIAL ~~ ENERGY	127.72	ADDITIONAL ~~ FEATURE	14.85	TOP ~~ CORNER	31.47
STRONG ~~ BELIEF	21.11	POTENTIAL ~~ DIFFERENCE	63.41	ADDITIONAL ~~ FACTOR	14.07	TOP ~~ LAYER	14.94
STRONG ~~ LEADERSHIP	19.38	POTENTIAL ~~ BENEFIT	50.11	ADDITIONAL ~~ LOAD	12	TOP ~~ LEVEL	14.49

STRONG ~~ TENDENCY	18.11	POTENTIAL ~~ CUSTOMER	44.08	ADDITIONAL ~~ EQUIPMENT	11.65	TOP ~~ PRIORITY	13.01
STRONG ~~ SUPPORT	17.79	POTENTIAL ~~ PROBLEM	42.36	ADDITIONAL ~~ STRESS	10.19	TOP ~~ MANAGER	12.9
STRONG ~~ BRAND	17.28	POTENTIAL ~~ THREAT	42.13	ADDITIONAL ~~ SUM	9.42	TOP ~~ POINTER	12.26
LOCAL ~~ AUTHORITY	417.66	POTENTIAL ~~ RISK	27.39	ADDITIONAL ~~ INCOME	8.9	TOP ~~ DESTINATION	10.14
LOCAL ~~ COMMUNITY	209.32	POTENTIAL ~~ DROP	21.7	ADDITIONAL ~~ ASSITANCE	8.68	TOP ~~ PREDATOR	9.08
LOCAL ~~ GOVERNMENT	195.15	POTENTIAL ~~ CONFLICT	20.23	ADDITIONAL ~~ SUPPORT	8.45	TOP ~~ SPEED	9
LOCAL ~~ LEVEL	99.86	POTENTIAL ~~ MARKET	19.72	ADDITIONAL ~~ DEN	8.37	TOP ~~ EXECUTIVE	8.53
LOCAL ~~ PEOPLE	82.67	POTENTIAL ~~ IMPACT	19.27	ADDITIONAL ~~ SOURCE	8.31	TOP ~~ SURFACE	8.42
LOCAL ~~ RESIDENT	78.1	POTENTIAL ~~ SOURCE	17.79	ADDITIONAL ~~ SERVICE	7.85	TOP ~~ END	8.14
LOCAL ~~ ECONOMY	65.54	POTENTIAL ~~ BUYER	17.71	CIVIL ~~ WAR	9.88	TOP ~~ NODE	7.83
LOCAL ~~ COUNCIL	55.43	POTENTIAL ~~ INVESTOR	16.73	CIVIL ~~ SOCIETY	8.61	TOP ~~ SOIL	7.59
LOCAL ~~ TAX	52.52	POTENTIAL ~~ EFFECT	16.56	CIVIL ~~ RIGHT	7.9	IMMEDIATE ~~ AFTERMATH	27.43
LOCAL ~~ AREA	47.89	POTENTIAL ~~ TRAVELLER	15.51	CIVIL ~~ SERVANT	11.09	IMMEDIATE ~~ VICINTY	18.26
LOCAL ~~ GOVERNANCE	37.08	POTENTIAL ~~ BIAS	15.1	CIVIL ~~ MOVEMENT	7.52	IMMEDIATE ~~ EFFECT	18.15
LOCAL ~~ MARKET	33.21	POTENTIAL ~~ GROWTH	14.6	CIVIL ~~ LAW	6.74	IMMEDIATE ~~ DANGER	11.83
LOCAL ~~ CULTURE	31.31	POTENTIAL ~~ DANGER	13.73	CIVIL ~~ LIBERTY	9.84	IMMEDIATE ~~ PERIOD	10.33

LOCAL ~~ ANAESTHETIC	27.44	POTENTIAL ~~ ENTRANT	13.12	CIVIL ~~ SERVICE	6.12	IMMEDIATE ~~ RESPONSE	10.06
LOCAL ~~ REVENUE	25.34	BLACK ~~ MAN	206.09	CIVIL ~~ ACTIVISIM	10.72	IMMEDIATE ~~ IMPACT	9.57
LOCAL ~~ AUTHORITY	24.83	BLACK ~~ BOY	83.24	CIVIL ~~ PARTNERSHIP	8.76	IMMEDIATE ~~ THREAT	9.28
LOCAL ~~ ENVIRONMENT	18.37	BLACK ~~ MASCULINITY	77.73	CIVIL ~~ PROCEEDING	9.52	IMMEDIATE ~~ CAUSE	8.46
LOCAL ~~ MINIMA	17.99	BLACK ~~ HOLE	64.59	CIVIL ~~ PARTNERSHIP	7.92	IMMEDIATE ~~ REWARD	8.26
LOCAL ~~ NEWSPAPER	16.95	BLACK ~~ YOUTH	62.46	CIVIL ~~ UNREST	9.58	IMMEDIATE ~~ IMPORTANCE	7.95
LOCAL ~~ SCHOOL	16.48	BLACK ~~ WOMAN	54.56	CIVIL ~~ CAMPAIGNER	11.43	IMMEDIATE ~~ FUTURE	7.36
GENERAL ~~ PUBLIC	134.74	BLACK ~~ MALE	44.7	CIVIL ~~ STRIFE	11.43	IMMEDIATE ~~ CONSEQUENCE	7.01
GENERAL ~~ TREND	121.73	BLACK ~~ COMMUNITY	43.09	CIVIL ~~ ENGINEERING	12.02	IMMEDIATE ~~ ACTION	7
GENERAL ~~ WILL	79.58	BLACK ~~ SLAVE	27.7	CIVIL ~~ ASSOCIATION	12.01	IMMEDIATE ~~ NEED	5.75
GENERAL ~~ RULE	55.81	BLACK ~~ PEOPLE	20.75	CIVIL ~~ FREEDOM	10.99	IMMEDIATE ~~ FAMILY	5.55
GENERAL ~~ EXAMINATION	55.32	BLACK ~~ BOX	14.35	CIVIL ~~ SYSTEM	7.18	IMMEDIATE ~~ ENVIRONMENT	5.31
GENERAL ~~ PRINCIPLE	54.2	BLACK ~~ WHITE	13.86	CIVIL ~~ PROCEDURE	6.03	IMMEDIATE ~~ YEAR	5.12
GENERAL ~~ CONSENSUS	50.63	BLACK ~~ BELT	12.96	CRUCIAL ~~ ROLE	84.03	SAFE ~~ ENVIRONMENT	32.41
GENERAL ~~ INTELLIGENCE	42.28	BLACK ~~ GIRL	12.84	CRUCIAL ~~ FACTOR	61.92	SAFE ~~ WATER	22.12
GENERAL ~~ OVERVIEW	36.24	BLACK ~~ POPULATION	12.7	CRUCIAL ~~ ASPECT	34.31	SAFE ~~ DRINKING	21.34
GENERAL ~~ ANAESTHETIC	34.27	BLACK ~~ WORKER	11.33	CRUCIAL ~~ ELEMNET	34.01	SAFE ~~ SEX	17.9

GENERAL ~~ POPULATION	32.52	BLACK ~~ AFRICAN	9.43	CRUCIAL ~~ PART	30.09	SAFE ~~ WORKING	15.29
GENERAL ~~ PATTERN	28.84	BLACK ~~ POLITICS	9.16	CRUCIAL ~~ DIFFERENCE	14.15	SAFE ~~ PRODUCT	10.45
GENERAL ~~ ELECTION	27.58	BLACK ~~ AMERICAN	9.1	CRUCIAL ~~ IMPORTANCE	13.4	SAFE ~~ PLACE	7.48
GENERAL ~~ KNOWLEDGE	18.74	BLACK ~~ CULTURE	9.08	CRUCIAL ~~ FLAW	13.2	SAFE ~~ LIMIT	6.94
GENERAL ~~ MANAGER	18.47	RIGHT ~~ HAND	136.24	CRUCIAL ~~ POINT	13.1	LITTLE ~~ EVIDENCE	94.43
GENERAL ~~ HEALTH	17.72	RIGHT ~~ HEMISPHERE	122.72	CRUCIAL ~~ QUESTION	10.25	LITTLE ~~ BIT	56.98
GENERAL ~~ THEORY	17.04	RIGHT ~~ SIDE	111.5	CRUCIAL ~~ ISSUE	10.05	LITTLE ~~ EFFECT	45.25
GENERAL ~~ WARD	16.5	RIGHT ~~ VIOLATION	102.86	CRUCIAL ~~ CONCEPT	6.84	LITTLE ~~ DOUBT	41.16
GENERAL ~~ ANAESTHESIA	15.2	RIGHT ~~ KNEE	74.97	CRUCIAL ~~ SOURCE.	5.13	LITTLE ~~ FINGER	36.4
GENERAL ~~ AGREEMENT	14.28	RIGHT ~~ HOUSING	50.36	CRUCIAL ~~ COMPONENT	5.1	LITTLE ~~ IMPACT	33.63
CLEAR ~~ DISTINCTION	49.08	RIGHT ~~ SHOULDER	41.12	CRUCIAL ~~ ORDER	5.01	LITTLE ~~ ATTENTION	30.3
CLEAR ~~ PICTURE	48.22	RIGHT ~~ LEG	39	ACTIVE ~~ SITE	74.11	LITTLE ~~ DIFFERENCE	28.64
CLEAR ~~ INDICATION	46.29	RIGHT ~~ FREEDOM	37.73	ACTIVE ~~ ROLE	66.26	LITTLE ~~ RESEMBLANCE	20.93
CLEAR ~~ EVIDENCE	38.81	RIGHT ~~ MOVEMENT	35.47	ACTIVE ~~ EUTHANASIA	51.93	LITTLE ~~ CHANCE	18.62
CLEAR ~~ CUT	37.62	RIGHT ~~ FOSSA	33.52	ACTIVE ~~ LISTENING	34.73	LITTLE ~~ ROOM	16.67
CLEAR ~~ EXAMPLE	27.13	RIGHT ~~ ANKLE	25.48	ACTIVE ~~ INGREDIENT	25.69	LITTLE ~~ RELEVANCE	15.76
CLEAR ~~ UNDERSTANDING	25.64	RIGHT ~~ ARM	24.77	ACTIVE ~~ PARTICIPATION	22.5	LITTLE ~~ GUIDANCE	15.19

CLEAR ~~ DIFFERENCE	24.73	RIGHT ~~ SELF-DEFENCE	23.85	ACTIVE ~~ MANAGEMENT	21.54	LITTLE ~~ INFLUENCE	14.88
CLEAR ~~ ANSWER	12.16	RIGHT ~~ ABUSE	23.35	ACTIVE ~~ GAUGE	19.87	LITTLE ~~ EFFORT	14.67
CLEAR ~~ JUICE	12.1	RIGHT ~~ OBLIGATION	21.71	ACTIVE ~~ CITIZENSHIP	18.33	LITTLE ~~ RESEARCH	14.53
CLEAR ~~ BREATH	11.88	RIGHT ~~ DIRECTION	21.12	ACTIVE ~~ PART	17.71	LITTLE ~~ TIME	13.54
CLEAR ~~ DEFINITION	11.72	RIGHT ~~ ADVOCATE	20.87	ACTIVE ~~ AGENT	15.23	LITTLE ~~ HOPE	12.66
CLEAR ~~ GOAL	11.35	RIGHT ~~ ANGLE	20.06	ACTIVE ~~ PARTICIPANT	15.14	LITTLE ~~ MAID	12.58
CLEAR ~~ IDEA	10.22	RIGHT ~~ SILENCE	19.82	ACTIVE ~~ FACULTY	12.61	LITTLE ~~ CONTROL	11.74
CLEAR ~~ SIGN	9.81	ONLY ~~ WAY	100.53	ACTIVE ~~ INVOLVEMENT	12.03	BLUE ~~ COLOUR	19.5
CLEAR ~~ VIEW	9.16	ONLY ~~ THING	46.8	ACTIVE ~~ FILTER	11.9	BLUE ~~ COLONY	14.37
CLEAR ~~ TREND	8.38	ONLY ~~ OPTION	38.47	ACTIVE ~~ STRAIN	10.46	BLUE ~~ LATEX	12.099
CLEAR ~~ PATTERN	8.12	ONLY ~~ CONTROLLER	29.08	ACTIVE ~~ VERB	9.86	BLUE ~~ LIGHT	11.01
CLEAR ~~ MESSAGE	7.82	ONLY ~~ FRUIT	22.04	ACTIVE ~~ MATRIX	8.48	BLUE ~~ TRIANGLE	9.54
CLEAR ~~ GUIDELINE	7.31	ONLY ~~ PART	21.59	ACTIVE ~~ LIFESTYLE	7.74	BLUE ~~ DYE	9.34
SPECIFIC ~~ AREA	63.59	ONLY ~~ MINORITY	20.16	BAD ~~ NEWS	62.32	BLUE ~~ SEA	9.01
SPECIFIC ~~ HEAT	60.5	ONLY ~~ ONE	19.05	BAD ~~ CASE	33.36	BLUE ~~ JEAN	8.01
SPECIFIC ~~ CONE	51.08	ONLY ~~ PERCENT	18.4	BAD ~~ SCENARIO	29.3	BLUE ~~ LINE	7.58
SPECIFIC ~~ ANTIBODY	28.48	ONLY ~~ FACTOR	18.39	BAD ~~ LINK	23.82	BLUE ~~ TROUSER	7.52
SPECIFIC ~~ PURPOSE	27	ONLY ~~ DIFFERENCE	17.29	BAD ~~ WEATHER	20.61	BLUE ~~ COLOURING	7.45
SPECIFIC ~~ TYPE	26.32	ONLY ~~ SOURCE	16.4	BAD ~~ THING	16.01	BLUE ~~ TONE	7.05

SPECIFIC ~~ REQUIREMENT	24.53	ONLY ~~ FRACTION	12.01	BAD ~~ HABIT	13.99	BLUE ~~ ELEMENT	6.41
SPECIFIC ~~ SITUATION	20.59	ONLY ~~ PERSON	10.32	BAD ~~ EFFECT	13.84	BLUE ~~ PARTICLE	6.16
SPECIFIC ~~ EVENT	18.18	ONLY ~~ EXCEPTION	9.74	BAD ~~ CONDITION	13.02	BLUE ~~ COLOR	5.44
SPECIFIC ~~ GROUP	18.12	ONLY ~~ ALLELE	9.41	BAD ~~ LUCK	12.52	BLUE ~~ FIGURE	5.32
SPECIFIC ~~ SEGMENT	17.72	ONLY ~~ MEANS	8.3	BAD ~~ PUBLICITY	11.34	DANGEROUS ~~ MANSLAUGHTER	16.26
SPECIFIC ~~ CAPACITY	17.66	ONLY ~~ MATTER	8.25	BAD ~~FAITH	11.33	DANGEROUS ~~ ACT	15.96
SPECIFIC ~~ GENE	16.63	ONLY ~~ VARIABLE	7.65	BAD ~~ MOOD	11.13	DANGEROUS ~~ WEAPON	13.63
SPECIFIC ~~ LOCATION	16.54	ONLY ~~ WOMAN	7.44	BAD ~~ REPUTATION	10.86	DANGEROUS ~~ SITUATION	11.62
SPECIFIC ~~ NEED	15.74	COMPLEX ~~ SYSTEM	33.95	BAD ~~ DEBT	10.01	DANGEROUS ~~ MTH	9.45
SPECIFIC ~~ SEQUENCE	15.17	COMPLEX ~~ INTERACTION	29.41	BAD ~~ HARVEST	9.57	DANGEROUS ~~ DRIVING	9.33
SPECIFIC ~~ REGION	15.16	COMPLEX ~~ STRUCTURE	22.97	BAD ~~ CRIME	8.85	DANGEROUS ~~ PLACE	7.74
SPECIFIC ~~ MODULUS	15.15	COMPLEX ~~ VECTOR	22.96	BAD ~~ POSITION	8.47	DANGEROUS ~~ SUBSTANCE	6.47
SPECIFIC ~~ TARGET	15.09	COMPLEX ~~ PROCESS	21.49	BAD ~~ MISTAKE	7.79	DANGEROUS ~~ CRUSE	5.79
SPECIFIC ~~ FUNCTION	15.05	COMPLEX ~~ RELATIONSHIP	21.43	BAD ~~ HEALTH	7.08	QUICK ~~ RECOVERY	28.86
VARIOUS ~~ FACTOR	67.9	COMPLEX ~~ CHARACTER	20.77	SEXUAL ~~ INTERCOURSE	60.68	QUICK ~~ RESPONSE	21.85
VARIOUS ~~ METHOD	56.71	COMPLEX ~~ ISSUE	19.9	SEXUAL ~~ ABUSE	52.37	QUICK ~~ LEARNING	21.75

VARIOUS ~~ TYPE	51.92	COMPLEX ~~ INTERPLAY	17.1	SEXUAL ~~ ORIENTATION	46.8	QUICK ~~ SUCCESSION	15.89
VARIOUS ~~ FORM	51.9	COMPLEX ~~ NATURE	16.89	SEXUAL ~~ BEHAVIOUR	43.69	QUICK ~~ FIX	14.87
VARIOUS ~~ ASPECT	47.6	COMPLEX ~~ SENTENCE	16.59	SEXUAL ~~ DIVISION	41.74	QUICK ~~ LUNCH	9.38
VARIOUS ~~ WAY	39.68	COMPLEX ~~ SEIZURE	14.63	SEXUAL ~~ REVOLUTION	39.02	QUICK ~~ WAY	7.62
VARIOUS ~~ GROUP	29.55	COMPLEX ~~ BUNDLE	13.32	SEXUAL ~~ ASSAULT	38.19	QUICK ~~ LOADING	7.41
VARIOUS ~~ SOURCE	28.8	COMPLEX ~~ PHENOMENON	12.57	SEXUAL ~~ HARASSMENT	35.17	QUICK ~~ LOOK	6.38
VARIOUS ~~ AUTHOR	26.24	COMPLEX ~~ MANIFOLD	11.69	SEXUAL ~~ ACTIVITY	33.12	QUICK ~~ TIME	6.21
VARIOUS ~~ REASON	24.53	COMPLEX ~~ NETWORK	11.1	SEXUAL ~~ VIOLENCE	32.5	QUICK ~~ SLOUTION	5.53
VARIOUS ~~ ELEMENT	21.93	COMPLEX ~~ TASK	10.51	SEXUAL ~~ DESIRE	30.12	QUICK ~~ FORTUNE	5.37
VARIOUS ~~ ACTIVITY	21.15	COMPLEX ~~ NUMBER	10.04	SEXUAL ~~ RELATIONSHIP	29.56	QUICK ~~ DEGRADATION	5.01
VARIOUS ~~ STAGE	20.41	COMPLEX ~~ WEB	8.48	SEXUAL ~~ PARTNER	28.96	QUICK ~~ BANK	5
VARIOUS ~~ PARAMETER	19.61	COMPLEX ~~ SOCIETY	7.89	SEXUAL ~~ GRATIFICATION	27.44	OFFICIAL ~~ STATISTIC	168.85
VARIOUS ~~ STRATEGY	18.44	RECENT ~~ YEAR	455.58	SEXUAL ~~ RELATION	25.35	OFFICIAL ~~ WEB-SITE	31.55
VARIOUS ~~ LEVEL	18.3	RECENT ~~ STUDY	55.65	SEXUAL ~~ PROMISCUITY	23.44	OFFICIAL ~~ LANGUAGE	19.37
VARIOUS ~~ TECHNIQUE	17.59	RECENT ~~ RESEARCH	53.44	SEXUAL ~~ ENCOUNTER	21.2	OFFICAL ~~ WEBSITE	13.1
VARIOUS ~~ DEPARTMENT	17.1	RECENT ~~ DECADE	39.53	SEXUAL ~~ APPETITE	20.18	OFFICIAL ~~ RELIGION	12.12

VARIOUS ~~ PART	16.59	RECENT ~~ DEVELOPMENT	37.69	SEXUAL ~~ REPRODUCTION	19.14	OFFICIAL ~~ POSITION	8.59
VARIOUS ~~ STUDY	15.37	RECENT ~~ TIME	30.39	SEXUAL ~~ MATURITY	17.65	OFFICIAL ~~ DOCUMENT	8.18
FINANCIAL ~~ MARKET	155.39	RECENT ~~ TREND	26.72	AWARE ~~ CONTAGION	6.02	OFFICIAL ~~ HISTORY	8
FINANCIAL ~~ YEAR	154.22	RECENT ~~ REPORT	23.32	REGIONAL ~~ COOPERATION	59	FINE ~~ MESH	37.13
FINANCIAL ~~ CRISIS	151.62	RECENT ~~ EXCAVATION	22.63	REGIONAL ~~ LEVEL	57.1	FINE ~~ DINING	31.09
FINANCIAL ~~ STATEMENT	115.39	RECENT ~~ HISTORY	19.73	REGIONAL ~~ GROUPING	55.22	FINE ~~ CARTEL	26.83
FINANCIAL ~~ INSTITUTION	88.98	RECENT ~~ WORK	18.8	REGIONAL ~~ INTEGRATION	51.98	FINE ~~ POWEDER	21.45
FINANCIAL ~~ REPORT	79.29	RECENT ~~ EVENT	18.34	REGIONAL ~~ DIALECT	47.02	FINE ~~ LINE	16.53
FINANCIAL ~~ RESOURCE	69.15	RECENT ~~ DISCOVERY	17.01	REGIONAL ~~ PROJECT	38.61	FINE ~~ WINE	16.3
FINANCIAL ~~ PERFORMANCE	63.11	RECENT ~~ CASE	16.35	REGIONAL ~~ TRADE	31.27	FINE ~~ REDUCTION	14.58
FINANCIAL ~~ POSITION	58.92	RECENT ~~ ARTICLE	15.83	REGIONAL ~~ AGREEMENT	28.89	FINE ~~ ART	12.8
FINANCIAL ~~ SUPPORT	48.21	RECENT ~~ CRISIS	15.51	REGIONAL ~~ VARIATION	28.54	FINE ~~ RESTAURANT	11.83
FINANCIAL ~~ CONTRIBUTION	44.69	RECENT ~~ LITERATURE	15.18	REGIONAL ~~ ARRANGEMENT	27.15	FINE ~~ EXAMPLE	11.13
FINANCIAL ~~ RISK	42.22	RECENT ~~ CHEST	14.72	REGIONAL ~~ INSTITUTION	23.68	FINE ~~ TERMINAL	10.43
FINANCIAL ~~ SERVICE	38.79	RECENT ~~ DATUM	13.98	REGIONAL ~~ BLOC	20.06	FINE ~~ MOTOR	10.2
FINANCIAL ~~ ACCOUNTING	30.3	RECENT ~~ SURVEY	12.49	REGIONAL ~~ IDENTITY	11.56	FINE ~~ BRANCH	10.16

FINANCIAL ~~ BENEFIT	30.02	MEDICAL ~~ PROFESSION	220.94	REGIONAL ~~ LEADER	10.67	FINE ~~ DETAIL	8.75
FINANCIAL ~~ ANALYSIS	29.75	MEDICAL ~~ HISTORY	193.01	REGIONAL ~~ ORGANIZATION	10.66	FINE ~~ CRYSTAL	8.5
FINANCIAL ~~ RATIO	26.97	MEDICAL ~~ TECHNOLOGY	103.9	REGIONAL ~~ SECURITY	10.39	FINE ~~ MATERIAL	6.67
FINANCIAL ~~ SITUATION	25.86	MEDICAL ~~ CARE	50.07	REGIONAL ~~ GOVERNMENT	10.18	FINE ~~ STRUCTURE	5.98
FINANCIAL ~~ BURDEN	24.77	MEDICAL ~~ TREATMENT	47.08	REGIONAL ~~ SURVEY	9.44	HAPPY ~~ ENDING	56.51
FINANCIAL ~~ SYSTEM	23.89	MEDICAL ~~ REPORT	38.57	REGIONAL ~~ GROUP	9.14	HAPPY ~~ WORKER	41.21
NECESSARY ~~ CONDITION	39.02	MEDICAL ~~ SCHOOL	31.84	REGIONAL ~~ MARKET	8.08	HAPPY ~~ FINALE	12
NECESSARY ~~ ORDER	27.6	MEDICAL ~~ PROFESSIONAL	31.11	SUFFICIENT ~~ EVIDENCE	29.64	HAPPY ~~ LIFE	6.35
NECESSARY ~~ CONNECTION	26.16	MEDICAL ~~ PRACTITIONER	31.04	SUFFICIENT ~~ AMOUNT	19.88	PROPER ~~ NOUN	30.66
NECESSARY ~~ SKILL	19.21	MEDICAL ~~ PRACTICE	30.33	SUFFICIENT ~~ CONDITION	18.75	PROPER ~~ FUNCTIONING	23.18
NECESSARY ~~ INFORMATION	17.89	MEDICAL ~~ EDUCATION	29.12	SUFFICIENT ~~ TIME	17.94	PROPER ~~ RESONANCE	19.04
NECESSARY ~~ TRUTH	13.07	MEDICAL ~~ STUDENT	28.65	SUFFICIENT ~~ INFORMATION	17.18	PROPER ~~ NAME	11.63
NECESSARY ~~ STEP	12.46	MEDICAL ~~ INTERVENTION	23.64	SUFFICIENT ~~ REASON	16.52	PROPER ~~ UNDERSTANDING	11.45
NECESSARY ~~ PRECAUTION	9.7	MEDICAL ~~ THERAPY	21.18	SUFFICIENT ~~ QUANTITY	15.9	PROPER ~~ FUNCTION	5.23
NECESSARY ~~ UNITY	8.63	MEDICAL ~~ SCIENCE	20.19	SUFFICIENT ~~ FUND	14.22	PROPER ~~ LANGAUGE	5.02
NECESSARY ~~ REQUIREMENT	6.3	MEDICAL ~~ ADVICE	19.63	SUFFICIENT ~~ INTEREST	13.05	FLAT ~~ PLATE	181.18

NECESSARY ~~ COMPONENT	6.19	MEDICAL ~~ PAPYRUS	17.53	SUFFICIENT ~~ RESOURCE	11.51	FLAT ~~ SCREEN	38.74
NECESSARY ~~ REVOLUTION	5.68	MEDICAL ~~ NOTE	15.94	SUFFICIENT ~~ SPACE	9.22	FLAT ~~ FADING	21.66
NECESSARY ~~ TRAINING	5.32	MEDICAL ~~ DISCOURSE	15.81	SUFFICIENT ~~ KNOWLEDGE	8.62	FLAT ~~ CURVE	18.62
NECESSARY ~~ FUND	5.04	MEDICAL ~~ CONDITION	14.44	SUFFICIENT ~~ PROTECTION	8.42	FLAT ~~ STRUCTURE	18.44
NATURAL ~~ SELECTION	194.61	BASIC ~~ PRINCIPLE	78.58	SUFFICIENT ~~ AIR	7.53	FLAT ~~ ORGANIZATION	11.23
NATURAL ~~ FREQUENCY	153.68	BASIC ~~ LEVEL	51.97	SUFFICIENT ~~ CASH	6.78	FLAT ~~ MODEL	11
NATURAL ~~ JUSTICE	137.18	BASIC ~~ IDEA	47.31	SUFFICIENT ~~ NOTICE	6.49	FLAT ~~ BONE	10.72
NATURAL ~~ RESOUCE	125.97	BASIC ~~ ASSUMPTION	36.84	SUFFICIENT ~~ POWER	6.28	FLAT ~~ RATE	9.62
NATURAL ~~ ENEMY	107.26	BASIC ~~ RIGHT	36.69	SUFFICIENT ~~ TRAINING	5.77	FLAT ~~ SURFACE	6.5
NATURAL ~~ LAW	81.84	BASIC ~~ PREMISS	35.91	SUFFICIENT ~~ CONSIDERATION	5.36	FLAT ~~ LAND	6.45
NATURAL ~~ ENVIRONMENT	80.89	BASIC ~~ STRUCTURE	35.77	SUFFICIENT ~~ DETAIL	5.21	FAMILIAR ~~ CONCEPT	12.31
NATURAL ~~ SCIENCE	73.65	BASIC ~~ CONCEPT	29.01	ANNUAL ~~ REPORT	588.04	FAMILIAR ~~ ENVIRONMENT	5.91
NATURAL ~~ DISASTER	66.67	BASIC ~~ SKILL	27.34	ANNUAL ~~ TURNOVER	35.28	FAMILIAR ~~ WORK	5.41
NATURAL ~~ WORLD	63.13	BASIC ~~ NEED	27.16	ANNUAL ~~ GROWTH	31.63	REALISTIC ~~ DEPICTION	15.47
NATURAL ~~ HABITAT	44.28	BASIC ~~ REQUIREMENT	26.53	ANNUAL ~~ RAINFALL	20.43	REALISTIC ~~ PICTURE	10.96
NATURAL ~~ PROCESS	37.62	BASIC ~~ LIBERTY	26.12	ANNUAL ~~ REVENUE	17.98	REALISTIC ~~ JURISPRUDENCE	9.06

NATURAL ~~ RATE	32.03	BASIC ~~ EARNINGS	23.2	ANNUAL ~~ INCOME	17.33	REALISTIC ~~ VIEW	8.77
NATURAL ~~ BEAUTY	31.5	BASIC ~~ TENET	19.47	ANNUAL ~~ RATE	16.9	REALISTIC ~~ FICTION	7.98
NATURAL ~~ GAS	29.28	BASIC ~~ INCOME	17.66	ANNUAL ~~ OUTPUT	15.09	REALISTIC ~~ STYLE	7.77
NATURAL ~~ PHENOMENON	27.79	BASIC ~~ UNDERSTANDING	16.58	ANNUAL ~~ COST	11.76	REALISTIC ~~ GOAL	7.68
NATURAL ~~ SETTING	27.78	BASIC ~~ FUNCTION	15.52	ANNUAL ~~ PRECIPITATION	10.88	REALISTIC ~~ EXPLANATION	5.47
NATURAL ~~ LOGARITHM	25.66	BASIC ~~ KNOWLEDGE	15.27	ANNUAL ~~ PAYMENT	10.23	REALISTIC ~~ WAY	5.4
NATURAL ~~ RIGHT	24.73	BASIC ~~ COMPONENT	13.46	ANNUAL ~~ BUDGET	9.77	REALISTIC ~~ APPROACH	5.22
NATURAL ~~ LANGUAGE	24.61	BASIC ~~ ELEMENT	12.9	ANNUAL ~~ REVUE	9.37	MATHEMATICAL ~~ MODEL	96.1
PUBLIC ~~ TRANSPORT	262.71	DOMESTIC ~~ VIOLENCE	181.84	ANNUAL ~~ CHANGE	9.15	MATHEMATICAL ~~ APPLICATION	14.61
PUBLIC ~~ SPHERE	133.72	DOMESTIC ~~ SPHERE	76.21	ANNUAL ~~ SAVING	8.34	MATHEMATICAL ~~ EGUATION	14.07
PUBLIC ~~ SECTOR	126.18	DOMESTIC ~~ MARKET	74.39	ANNUAL ~~ DATUM	8	MATHEMATICAL ~~ ALGORITHM	11.82
PUBLIC ~~ OPINION	107.49	DOMESTIC ~~ CHORE	48.64	ANNUAL ~~ CROP	6.64	MATHEMATICAL ~~ TOOL	10.01
PUBLIC ~~ INTEREST	106.03	DOMESTIC ~~ CREDIT	45.21	ANNUAL ~~ ENERGY	6.51	MATHEMATICAL ~~ PROBLEM	9.76
PUBLIC ~~ POLICY	51.28	DOMESTIC ~~ CURRENCY	41.43	ANNUAL ~~ SALE	5.9	MATHEMATICAL ~~ SOLUTION	8.43
PUBLIC ~~ HEALTH	50.25	DOMESTIC ~~ POLITICS	38.73	SPECIAL ~~ OFFER	50.61	MATHEMATICAL ~~ CONCEPT	8.18
PUBLIC ~~ PERCEPTION	49.44	DOMESTIC ~~ POLICY	31.6	SPECIAL ~~ DOCUMENT	37.28	MATHEMATICAL ~~ FUNCTION	7.44

PUBLIC ~~ EXPENDITURE	38.27	DOMESTIC ~~ TOURISM	31.06	SPECIAL ~~ ATTENTION	29.14	MATHEMATICAL ~~ ISSUE	7.06
PUBLIC ~~ DOMAIN	37.33	DOMESTIC ~~ DEMAND	31.02	SPECIAL ~~ SUBJECT	26.6	MATHEMATICAL ~~ APPROACH	5.36
PUBLIC ~~ AWARENESS	36.74	DOMESTIC ~~ INDUSRTY	28.24	SPECIAL ~~ INTEREST	25.62	MARGINAL ~~ COST	211.66
PUBLIC ~~ AUTHORITY	34.04	DOMESTIC ~~ LABOUR	28.11	SPECIAL ~~ DIVIDEND	23.17	MARGINAL ~~ PROPENSITY	68.67
PUBLIC ~~ SERVICE	28.91	DOMESTIC ~~ PRODUCTION	27.94	SPECIAL ~~ STATUS	22.62	MARGINAL ~~ REVENUE	57.69
PUBLIC ~~ ARENA	28.48	DOMESTIC ~~ AFFAIR	26.21	SPECIAL ~~ PROTECTION	20.96	MARGINAL ~~ RATE	30.28
PUBLIC ~~ GOOD	27.64	DOMESTIC ~~ PRICE	23.47	SPECIAL ~~ RAPPORTEUR	17.6	MARGINAL ~~ BENEFIT	30.25
PUBLIC ~~ DIPLOMACY	26.41	DOMESTIC ~~ INTEREST	23.04	SPECIAL ~~ DAMAGE	14.69	MARGINAL ~~ CURVE	25.45
PUBLIC ~~ IDENTITY	24.38	DOMESTIC ~~ LEVEL	20.69	SPECIAL ~~ REGION	14.08	MARGINAL ~~ PRODUCTIVITY	12.25
PUBLIC ~~ RELATION	24.19	DOMESTIC ~~ PRODUCT	20.58	SPECIAL ~~ SCIENTIFIC	12.51	MARGINAL ~~ PRICING	10.89
PUBLIC ~~ DEBT	22.56	DOMESTIC ~~ ECONOMY	18.02	SPECIAL ~~ NEED	12.16	MARGINAL ~~ PRODUCT	9.89
PUBLIC ~~ BODY	20.72	DOMESTIC ~~ BUSINESS	17.1	SPECIAL ~~ ASSISSTANCE	11.96	MARGINAL ~~ IMPACT	6.29
DIFFICULT ~~ TASK	68.38	FUTURE ~~ GENERATION	98.44	SPECIAL ~~ PROVISION	11.74	MARGINAL ~~ PRICE	5.61
DIFFICULT ~~ SITUATION	12.64	FUTURE ~~ PROSPECT	85	SPECIAL ~~ SITE	11.03	RACIAL ~~ HYGIENE	87.2
DIFFICULT ~~ QUESTION	6.51	FUTURE ~~ RESEARCH	70.44	SPECIAL ~~ PLACE	10.42	RACIAL ~~ DISCRIMINATION	53.92
DIFFICULT ~~ PROBLEM	5.83	FUTURE ~~ PREFORMANCE	69.43	SPECIAL ~~ EFFECT	9.78	RACIAL ~~ PREJUDICE	21.79

POSITIVE ~~ EFFECT	129.8	FUTURE ~~ DEVELOPMENT	47.94	SPECIAL ~~ PRIVILEGE	9.59	RACIAL ~~ HYGIENIST	20.8
POSITIVE ~~ IMPACT	104.17	FUTURE ~~ PLAN	27.51	SPECIAL ~~ PROGRAMME	9.53	RACIAL ~~ GROUP	17.24
POSITIVE ~~ CORRELATION	100.21	FUTURE ~~ CASH	26.89	APPARENT ~~ VISCOSITY	28.25	RACIAL ~~ DIFFERENCE	13.73
POSITIVE ~~ RELATIONSHIP	75.55	FUTURE ~~ PRACTICE	17.88	APPARENT ~~ LACK	17.25	RACIAL ~~ TENSION	12.23
POSITIVE ~~ INTEGER	64.44	FUTURE ~~ GROWTH	17.26	APPARENT ~~ CONTRADICTION	11.37	RACIAL ~~ INJUSTICE	11.63
POSITIVE ~~ CHARGE	56.4	FUTURE ~~ TREND	15.52	APPARENT ~~ DENSITY	9.03	RACIAL ~~ STEREOTYPE	10.75
POSITIVE ~~ ATTITUDE	55.84	FUTURE ~~ PREDICTION	14.24	APPARENT ~~ INABILITY	8.88	RACIAL ~~ EQUALITY	9.98
POSITIVE ~~ REINFORCEMENT	45.83	FUTURE ~~ EVENT	12.25	WEAK ~~ PERTURBATION	66.06	RACIAL ~~ CLASS	7.62
POSITIVE ~~ OUTCOME	45.82	FUTURE ~~ DEMAND	11.52	WEAK ~~ TIME-DEPENDENT	53.55	FAT ~~ GLOBLUE	185.97
POSITIVE ~~ POLITENESS	34.59	FUTURE ~~ OPPORTUNITY	11.22	WEAK ~~ BOND	31.82	FAT ~~ CONTENT	60.79
POSITIVE ~~ RESULT	33.76	FUTURE ~~ APPLICATION	9.97	WEAK ~~ HYDROGEN	24.98	FAT ~~ SUGAR	37.95
POSITIVE ~~ OUTLOOK	32.99	FUTURE ~~ BUSINESS	9.76	WEAK ~~ CORRELATION	24.61	FAT ~~ PROTEIN	35.93
POSITIVE ~~ ASPECT	32.19	FUTURE ~~ SUCCESS	9.63	WEAK ~~ DOLLAR	23.9	FAT ~~ CARBOHYDRATE	20.78
POSITIVE ~~ LIBERTY	31.68	FUTURE ~~ INCOME	8.94	WEAK ~~ ACID	19.8	FAT ~~ MASS	20.34
POSITIVE ~~ RESPONSE	29.36	FUTURE ~~ CASE	8.51	WEAK ~~ EXOGENEITY	17.63	FAT ~~ DEPOT	19.77
POSITIVE ~~ CHANGE	27.5	FUTURE ~~ ACTION	8.51	WEAK ~~BARGAINING	14.52	FAT ~~ INTAKE	18.77

POSITIVE ~~ FEEDBACK	24.12	YOUNG ~~ CHILD	285.81	WEAK ~~ RELATIONSHIP	13.68	FAT ~~ MILK	17.66
POSITIVE ~~ CONNOTATION	23.82			WEAK ~~ INTERACTION	13.62	FAT ~~ DEPOSITION	14.73
POSITIVE ~~ EXTERNALITY	23.48	YOUNG ~~ MAN	162.53	WEAK ~~ POSITION	10.86	FAT ~~ SALF	14.35
POSITIVE ~~ TEST	22.68	YOUNG ~~ GIRL	125.57	WEAK ~~ CHARACTER	10.34	FAT ~~ MOLECULE	8.84
INDIVIDUAL ~~ RIGHT	50.05	YOUNG ~~ WOMAN	107.15	WEAK ~~ ECONOMY	9.55	FAT ~~ RESERVE	8.25
INDIVIDUAL ~~ DIFFERENCE	43.15	YOUNG ~~ AGE	102.52	WEAK ~~ SIGNAL	8.54	FAT ~~ STORAGE	8.18
INDIVIDUAL ~~ GROUP	42.51	YOUNG ~~ ADULT	97.06	WEAK ~~ FRONT	7.82	FAT ~~ TISSUE	7.48
INDIVIDUAL ~~ FREEDOM	39.96	YOUNG ~~ BOY	53.06	WEAK ~~ PARTY	7.41	FAT ~~ STORE	6.03
INDIVIDUAL ~~ MEMBER	34.54	YOUNG ~~ GENERATION	50.31	WEAK ~~ POINT	7.1	FAT ~~ DISTRIBUTION	5.8
INDIVIDUAL ~~ BEHAVIOUR	29.84	YOUNG ~~ WORKER	49.51	WEAK ~~ STATE	6.62	EXCELLENT ~~ EXAMPLE	34.19
INDIVIDUAL ~~ RESPONSIBILITY	25.96	YOUNG ~~ COUPLE	36.52	WEAK ~~ SYSTEM	5.79	EXCELLENT ~~ SERVICE	23.48
INDIVIDUAL ~~ CHOICE	24.58	YOUNG ~~ SIBLING	35.62	COMMERCIAL ~~ HOSPITALITY	83.3	EXCELLENT ~~ QUALITY	12.21
INDIVIDUAL ~~ NEED	23.29	YOUNG ~~ LADY	35.01	COMMERCIAL ~~ EMPIRE	58.38	EXCELLENT ~~ SKILL	11.82
INDIVIDUAL ~~ AUTONOMY	18.28	YOUNG ~~ MODULUS	22.32	COMMERCIAL ~~ HOST	38.62	EXCELLENT ~~ OPPORTUNITY	11.47
INDIVIDUAL ~~ STATE	17.67	YOUNG ~~ MALE	17.25	COMMERCIAL ~~ STERILITY	34.59	EXCELLENT ~~ CORRELATION	10.31
INDIVIDUAL ~~ ACTION	15.15	YOUNG ~~ SISTER	16.38	COMMERCIAL ~~ JUICE	29.39	EXCELLENT ~~ PERFORMANCE	7.45

INDIVIDUAL ~~ INNUMERABILITY	14.8	YOUNG ~~ TEENAGER	11.95	COMMERCIAL ~~ DOMAIN	28.92	EXCELLENT ~~ CONDITION	6.99
INDIVIDUAL ~~ LEVEL	14.55	YOUNG ~~ SON	11.43	COMMERCIAL ~~ WHALING	27.67	EXCELLENT ~~ COMMUNICATION	5.89
INDIVIDUAL ~~ DIVIDER	14.43	YOUNG ~~ GROUP	10.14	COMMERCIAL ~~ SECTOR	26.06	EXCELLENT ~~ SOURCE	5.37
INDIVIDUAL ~~ PERFORMANCE	13.56	OPEN ~~ SOURCE	157.56	COMMERCIAL ~~ SUCCESS	24.79	EXCELLENT ~~ CUSTOMER	5.37
INDIVIDUAL ~~ PERSONALITY	13.33	OPEN ~~ DOOR	75.23	COMMERCIAL ~~ SLICER	15.63	HOT ~~ WATER	107.86
INDIVIDUAL ~~ PREFERENCE	13.24	OPEN ~~ QUESTION	45.73	COMMERCIAL ~~ BANK	14.41	HOT ~~ TOPIC	21.51
INDIVIDUAL ~~ CHARACTER	12.96	OPEN ~~ SIGNATURE	40.71	COMMERCIAL ~~ MANAGER	13.68	HOT ~~ ETHANOL	17.25
INDIVIDUAL ~~ COMPONENT	12.71	OPEN ~~ SUBSET	36.88	COMMERCIAL ~~ SOFTWARE	12.96	HOT ~~ MATTER	12.61
PHYSICAL ~~ EXAMINATION	409.82	OPEN ~~ SOFTWARE	34.4	COMMERCIAL ~~ CONTEXT	12.89	HOT ~~ PRESS	12.35
PHYSICAL ~~ ACTIVITY	129.95	OPEN ~~ ECONOMY	29.9	COMMERCIAL ~~ ACTIVITY	12.74	HOT ~~ WAFFLE	8.91
PHYSICAL ~~ PROPERTY	86.44	OPEN ~~ SET	24.34	COMMERCIAL ~~ INTEREST	12.16	HOT ~~ CANAPES	8.68
PHYSICAL ~~ OBJECT	75.08	OPEN ~~ SPACE	22.97	COMMERCIAL ~~ ORANGE	11.48	HOT ~~ SPOT	8.39
PHYSICAL ~~ HIGHLIGHT	56.04	OPEN ~~ CRITISIM	21.47	COMMERCIAL ~~ STERILIZATION	11.13	HOT ~~ WEATHER	8.22
PHYSICAL ~~ WORLD	52.34	OPEN ~~ FRAME	20.73	COMMERCIAL ~~ HERD	10.86	HOT ~~ JAM	8.09
PHYSICAL ~~ FACT	38.99	OPEN ~~ AIR	19.48	COMMERCIAL ~~ REVOLUTION	8.62	HOT ~~ PANT	7.8
PHYSICAL ~~ APPEARANCE	38.56	OPEN ~~ HARDWARE	18.34	RELATIVE ~~ HUMIDITY	94.17	HOT ~~ IRON	7.59

PHYSICAL ~~ CHARACTERISTIC	33.25	OPEN ~~ LOOP	18.02	RELATIVE ~~ ENDOWMENT	51.68	HOT ~~ DRINK	7.1
PHYSICAL ~~ ATTRACTIVENESS	29.92	OPEN ~~ DAY	17.77	RELATIVE ~~ PRICE	49.57	HOT ~~ ANEMOMETER	6.62
PHYSICAL ~~ DAMAGE	27.2	OPEN ~~ INTERPRETATION	16.82	RELATIVE ~~ CLAUSE	43.64	HOT ~~ AIR	5.68
PHYSICAL ~~ SCIENCE	26.43	OPEN ~~ REGIONALISM	16.02	RELATIVE ~~ IMPORTANCE	42.54	THERMAL ~~ CONDUCTIVITY	111.24
PHYSICAL ~~ ENVIRONMENT	23.51	OPEN ~~ READING	15.62	RELATIVE ~~ FACTOR	31.9	THERMAL ~~EFFICIENCY	86.18
PHYSICAL ~~ PROWESS	21.44	OPEN ~~ MARKET	15.2	RELATIVE ~~ EASE	23.65	THERMAL ~~EXPANSION	31.6
PHYSICAL ~~ EVIDENCE	18.57	OPEN ~~ LANDSCAPE	15.08	RELATIVE ~~ SIZE	21.83	THERMAL ~~PROCESSING	22.12
PHYSICAL ~~ HEALTH	16.72	RELEVANT ~~ INFORMATION	195.39	RELATIVE ~~ EFFICIENCY	19.56	THERMAL ~~DISTORTION	19.81
PHYSICAL ~~ MANIFESTATION	15.97	RELEVANT ~~ LITERATURE	31.12	RELATIVE ~~ SUCCESS	19.38	THERMAL ~~DIFFUSIVITY	19.28
PHYSICAL ~~ CONDITION	14.96	RELEVANT ~~ DATUM	21.66	RELATIVE ~~ STRENGTH	18.77	THERMAL ~~ENERGY	18.4
PHYSICAL ~~ ATTRIBUTE	14.8	RELEVANT ~~ TODAY	18.55	RELATIVE ~~ ABUNDANCE	14.62	THERMAL ~~COEFFICIENT	14.45
PHYSICAL ~~ SYMPTOM	14.08	RELEVANT ~~ CONSIDERATION	17.88	RELATIVE ~~ POSITION	14.22	THERMAL ~~PROCESS	12.06
FREE ~~ TRADE	316.57	RELEVANT ~~ PROVISION	16.93	RELATIVE ~~ RESOURCE	13.99	THERMAL ~~CENTER	10.3
FREE ~~ RADICAL	129.47	RELEVANT ~~ FACTOR	13.6	RELATIVE ~~ HEIGHT	13.74	THERMAL ~~EQUILIBRIUM	10.17
FREE ~~ MARKET	99.24	RELEVANT ~~ ARTICLE	9.63	RELATIVE ~~ INTENSITY	13.74	THERMAL ~~PROPERTY	9.41
FREE ~~ RANGE	93.56	RELEVANT ~~ DEPARTMENT	8.2	RELATIVE ~~ POVERTY	13.54	SWEET ~~ TASTE	31.14

FREE ~~ VERSE	55.51	RELEVANT ~~ ATOM	7.09	RELATIVE ~~ STEREOCHEMISTRY	12.87	SWEET ~~ WINE	20.99
FREE ~~ FLOW	54.41	RELEVANT ~~ FACT	6.89	RELATIVE ~~ PROPORTION	12.75	SWEET ~~ CORN	7.5
FREE ~~ CASH	50	RELEVANT ~~ CONVENTION	6.85	RELATIVE ~~ FRIEND	12.44	SWEET ~~ SMELL	6.75
FREE ~~ WILL	42.53	RELEVANT ~~ MATERIAL	6.71	FORMAL ~~ METHOD	62.94	SWEET ~~ OFFSPRING	6.53
FREE ~~ ELECTRON	41.03	RELEVANT ~~ VARIABLE	6.33	FORMAL ~~ INSTRUCTION	56.11	SWEET ~~ MAXWELL	6.44
FREE ~~ MOVEMENT	36.21	RELEVANT ~~ LEGISLATION	6.02	FORMAL ~~ EXPERIMENTATION	26.55	NARROW ~~ SCOPE	20.03
FREE ~~ ENERGY	20.61	RELEVANT ~~ ACTOR	6.02	FORMAL ~~ SYSTEM	23.52	NARROW ~~ EGOISM	19.86
FREE ~~ CHOICE	18.05	RELEVANT ~~ TOPIC	5.91	FORMAL ~~ RESEARCH	22.46	NARROW ~~ GAP	16.81
FREE ~~ STREAM	17.45	RELEVANT ~~ CONTENT	5.89	FORMAL ~~ ROLE	20.51	NARROW ~~ DEFINITION	16.19
FREE ~~ CHARGE	16.71	RELEVANT ~~ AUTHORITY	5.86	FORMAL ~~ STRUCTURE	20.37	NARROW ~~ RANGE	14.14
FREE ~~ ENCYCLOPAEDIA	16.45	RELEVANT ~~ KNOWLEDGE	5.7	FORMAL ~~ INSTITUTION	18.81	NARROW ~~ SENSE	14.02
FREE ~~ ACCESS	16.37	EASY ~~ ACCESS	73.1	FORMAL ~~ QUALIFICATION	16.84	NARROW ~~ BAND	10.04
FREE ~~ SPACE	15.85	EASY ~~ TASK	33.53	FORMAL ~~ TRAINING	12.65	NARROW ~~ INTERPRETATION	7.03
FREE ~~ RIDER	15.75	EASY ~~ WAY	33.51	FORMAL ~~ PROCEDURE	11.95	WARM ~~ FRONT	41
FREE ~~ END	15.52	EASY ~~ TARGET	6.75	FORMAL ~~ EDUCATION	11.87	WARM ~~ AIR	23.46
FREE ~~ AGREEMENT	13.94	EASY ~~ SOLUTION	4.36	FORMAL ~~ RULE	11.69	WARM ~~ CLIMATE	29.75

PARTICULAR ~~ GROUP	65.03	EASY ~~ PEOPLE	2.66	FORMAL ~~ DEFINITION	11.41	WARM ~~ TEMPERATURE	27.74
PARTICULAR ~~ INTEREST	49.81	NORMAL ~~ DISTRIBUTION	232.12	FORMAL ~~ LOGIC	11.06	WARM ~~ PERIOD	13.58
PARTICULAR ~~ AREA	39.67	NORMAL ~~ RANGE	43.28	FORMAL ~~ TEACHING	10.41	WARM ~~ WATER	12.06
PARTICULAR ~~ CASE	30.22	NORMAL ~~ INSPECTION	35.41	FORMAL ~~ LETTER	10.06	WARM ~~ CREAM	11.3
PARTICULAR ~~ ATTENTION	29.62	NORMAL ~~ SEEDLING	33.32	FORMAL ~~ CAUSE	9.67	WARM ~~ WEATHER	10.98
PARTICULAR ~~ EMPHASIS	25.01	NORMAL ~~ HEART	24.17	FORMAL ~~ PLANNING	9.38	WARM ~~ CLIME	10.31
PARTICULAR ~~ ISSUE	23.49	NORMAL ~~ BREATH	22.97	FORMAL ~~ AGREEMENT	8.71	WARM ~~ WELCOME	8.28
PARTICULAR ~~ CHARACTERISTIC	21.92	NORMAL ~~ CIRCUMSTANCE	20.67	OBVIOUS ~~ DISTENSION	24.65	WARM ~~ WINTER	7.53
PARTICULAR ~~ RELEVANCE	21.72	NORMAL ~~ SUBGROUP	19.93	OBVIOUS ~~ EXAMPLE	21.58	WARM ~~ COAT	5.76
PARTICULAR ~~ TYPE	20.75	NORMAL ~~ CONDITION	19.86	OBVIOUS ~~ DIFFERENCE	21.05	WARM ~~ CONDITION	5.45
PARTICULAR ~~ REFERENCE	19.54	NORMAL ~~ EXTENSION	19.71	OBVIOUS ~~ REASON	17.17	WARM ~~ MEDITERRANEAN	5.29
PARTICULAR ~~ SITUATION	18.67	NORMAL ~~ FUNCTION	17.7	OBVIOUS ~~ FLAW	11.03	WARM ~~ PRE-STRETCHING	5.16
PARTICULAR ~~ IMPORTANCE	17.73	NORMAL ~~ DEVELOPMENT	17.58	OBVIOUS ~~ PROBLEM	7.81	WARM ~~ CLIMATES	5.16
PARTICULAR ~~ CIRCUMSTANCE	16.08	NORMAL ~~ PATTERN	16.78	OBVIOUS ~~ ADVANTAGE	6.97		
PARTICULAR ~~ SIGNIFICANCE	16.05	NORMAL ~~ PROBABLITY	15.92	OBVIOUS ~~ PATTERN	6.9		

A.3 The native raters' results:

Collocations	Verb-Noun collocations Unacceptable	%	Verb-(I)-Noun collocations	%	To-Verb-(I)-Noun Collocation	%
Unacceptable	*keep silence, move castle, make revision, make exercise.*	2.75%	*be no doubt, be a comedy, spend their leisure, is gum infection, distribute the burden, get high mark*	19.35%	---	0
Questionable	*add music, play sport, read programme, remove segregation, is symbols, break time, find solutions.*	4.82%	*do plastic surgery, see my anti*, take a gap.*	9.6%	*to quit smoking, to take action, to do plastic surgery.*	42%
Did not occur in BNC/NBESs	*Move castle, cause eyestrain, make revision, *know adays, want sewage, want retainer, send emails/e-mails, require household, remain bias, develop gallstone, *thank gad, read Quran, play station, eat icecream, love marriage, working moms*					9.6%

Figure A.3-1: Overview of the native raters'responses on the Verb-Noun collocations

Acceptability measure	Verb-Adverb collocation Examples	Rate
Acceptable	*mention previously, affect negatively, are/is increasingly, do necessarily, use wisely, use extensively, teach differently, banned due, treat kindly, drive speedily*	83%
Questionable	*produce environmentally*	8.3%
Not Acceptable	*see recently, mention previously, do necessarily, produce environmentally, drive speedily*	41%

Figure A.3-2: The native-speaker acceptability results for the Verb-Adverb Collocations

Acceptability measure	Adjective-Noun collocation Examples	Rate
Acceptable	*bad vacation, mental problem, mental illness, important decision maker, technological device, online shop, huge headache, strict teacher, marine species, high school graduate, marine animal, right for self-expression, spend long hours, renewable power generator, sexual gesture, grand mother, grand parents, animals and species, animals' species, wrong posture, low carb, low recycling rate, animals as pets, private institutes, emotional generation, violent videogames, violent media, to close that debt, single parent adoption, positive self-esteem, professional gamers, bad holiday, social media natural view, poor grades, dependent on technology, clean medicine, very dependent on computers, active recycling, second-hand smoker*	93.02%
Questionable	*most urban dwellers, technological device, majoring in music, marine species, personal computer revolution, marine animal, grand mother, grand parents, emotional generation, easy victims as in they are easy victims of cigarettes, the technological tools, positive self-esteem, smart device detriment, dependent on technology, clean medicine*	34.88%
Not Acceptable	*most urban dwellers, personal computer revolution, right for self-expression, violent media, easy victims as in they are easy victims of cigarettes, smart device detrimen*	16%

Figure A.3-3: The native-speaker acceptability results for the Adjective-Noun Collocations

A.4 Native Rater's Data for Verb-Noun Collocation

Information sheet:

- This survey is about the use of collocation in student written assignments. The term collocation is defined as the co-occurance of two or more words to create a meaningful expression. For example, the collocation ***book an appointment*** or ***play football***. However, it must be noted that collocations as in ***to produce a positive*** is not a complete collocation. A noun is needed as in ***to produce a positive outcome***
- This survey is in two parts. The first part will have tables that list collocations for two groups which you need to rate on scale of acceptability either as acceptable, questionable, or unacceptable. Then, you are kindly asked to provide a short general comment about the collocations for the two groups.
- The second part is where you can write some comments on your judgements.

Part I

(A) *To + Collocations*	rate	*(B) To + Collocations*	rate
to play football		to take advantage	
to have fun		to tackle truancy	
to keep fit		to keep track	
to face interaction		to maintain homeostasis	
to take care		to survive anoxia	
to facilitate communication		to avoid misunderstanding	
to quit smoking		to do anything	
to do everything		to take place	
to lose weight		to aid self-awareness	
to take action		to raise awareness	
to do something		to commit suicide	
to stop smoking		to detect methylation	
		to grant amnesty	

Table 1. To + V-N Collocations

Please read the examples in Table 1 and rate the acceptability on a scale of three where:

1: Unacceptable (U)

2: Questionable (Q)

3: Acceptable (A)

Provide an overall comment regarding the differences between the two groups in terms of the acceptability and choice of words within the collocations.

……
……
……

(A) *To + Collocations*	Rate	(B) *To + Collocations*	rate
to do plastic surgery		to take into account	
to travel to egypt		to calculate the molarity	
to save the earth		to make the biocomposite*	
to decrease the number		to be a bat	
to watch a movie		to operate the crane	
to resolve the issue		to be a flagship*	
to resolve this issue		to achieve commercial sterility	
		to commit for contempt	
		to live under obedience	
		to pay its creditor	
		to kill in self-defence	
		to keep a diary	
		to work in partnership	

Table 2. To + V-(I)-N Collocations

Please read the examples in Table 2 and rate the acceptability on a scale of three where:

1: Unacceptable (U)

2: Questionable (Q)

3: Acceptable (A)

Provide an overall comment regarding the differences between the two groups in terms of the acceptability and choice of words within the collocations.

.....
.....
.....

(A) Collocation	rate	(B) Collocations	rate
Do plastic surgery		take into account	
Be no doubt		Do it explode	
See my anti*		Be an ex-smoker	
Go to Mecca		Be a non-smoker	
Take a gap		Contact by telephone	
Go to paris*		Regulate the productive	
Watch T.V		Experience the normalization	
Be a comedy		Be too Blunt	
Be in favor		Be an isomorphism	
Go to Salalah*		Flout the maxim	
Spend their leisure		Use political ideas	
Boycott the election		Buy in yarn	
Be gum infection		Beware the Cat	
Burn fossil fuel		Be rigid designators	
Go to Alain		Lose their birthright	
Go in Hatta		Taking rights seriously	
Have the ability		Be lose antemortem	
Be with migration		Be a bijection	
Go to Salalah		Represent the unrepresentable	
Go to London		Calculate the molarity	
Lead to isolation		Mark a watershed	
Use the phrase		Bear the brunt	

Take into consideration		Use in conjunction	
Be a threat		Be a subfield	
Distribute the burden		Be an affront	
Go to England		Concern Human Understanding	
Keep in touch		Make the biocomposite*	
Take a taxi		Go into liquidation	
Play foot ball		Be a subgroup	
Get high mark		Be the quotient	
Stay in contact		Be time consuming	

Table 3: V- (I)-N

Please read the examples in Table 3 and rate the acceptability on a scale of three where:

1: Unacceptable (U)

2: Questionable (Q)

3: Acceptable (A)

Provide an overall comment regarding the differences between the two groups in terms of the acceptability and choice of words within the collocations.

……
……
……

Verb-Noun collocations	Example	rate
Learn outcome	Distance learning is at the same time better than traditional learning in terms of the personal and **learning outcomes**.	
Save money	With time, people will be encouraged to reuse clothbags to **save money,** while helping to preserve the environment.	
Eat breakfast	We **eat breakfast** at the morning when we wakeup.	
Remove segregation	It might be true that sexuality expressed in sometimes, but also not **removing segregation** will cause the society ty expands their chances to become more corrupted by doing things in the hidden.	
Cause distortion	Upscaling **causes distortion** in the video and delivers worse colors than the original screen resolution.	
Cause health	Even the children's computer dependency is **causing health** problems	
Spend money	If governments **spend money** and try to clear the atmosphere of these gases, they might be able to stop climate change sooner than expected.	
Stop smoking	I don't think they will **stop smoking** because they get addicted to it every day.	
Learn institution	However, most distance **learning institutions** are accredited depending on the country.	
Read programme	We can **read programmes** a news of healthy.	
Sit together	In the next day, we got up early and went to the park, while we **sat together** and talked, we didn't find our sister.	
Increase variability	Also, it will raise the sea level, expand potential range of tropical diseases, disrupt agriculture, forestry, and natural ecosystem, and **increase variability** and extremes of regional weather.	
Keep silence	Then when he speaks in the class, everyone **keeps silence**, because his loud voice force us to keep quiet.	
Become part	To most people, plastic bags have **become part** of their lives, were they use them on a daily basis when it comes to shopping.	
Find solution	Communities should **find solutions** in order to minimize the level of global warming in order to lower damage caused to climate and environment than it has already done.	
Want Sewage	Passengers **want Sewage** Dumping Stopped	
Move castle	She heard about a **moving castle** which moved from place to another.	
Make revision	She didn't study day by day and **make revision.**	
Make decision	For example the computers can affect people's concentration, memory, the ability of making decision and set goals.	
Make exercise	First, they should be on them **make exercises** everyday.	
Do something	Work always includes doing something.	
Have fun	The holiday is really very important to **have fun** and enjoy so you can back to yourwork with more energy and you will produse more and more.	
Have nothing	They believe that technology **has nothing** to do in making them unsociable.	
Be symbol	She is **symbols** of love, hope and mercy.	
Require household	The first suggestion that the government set emissions standard for equipment and processes, **required household** and businesses to use specific types of equipment.	
Pay bill	Although, individuals are utilizing PCs to **pay bills.**	
Want retainer	Also, there are sites that **want retainers** and two weeks till the item arrives.	
Increase population	Along with the **increasing population**, significant changes are taking place in different parts of the world.	
Break time	The university must provide them a special place to enjoy their break time.	

Raise fund	In addition to all of that, he is also known for selling his art to **raise funds**.	
Add music	Also, the most important thing is to **add music** major, so those who are interested could study what they love.	
Encourage student	In addition, schools must take a role in **encouraging students** to participate in life activities more…	
Be gum	More serious results **are gum** infection and damage	
Play sport	I **play sports** every morning.	
Play station	We played many games like hide and seek and (**play station** 2).	
Play foot	I played **foot ball** with my family in the club.	

Table 4: V-N collocations

Please read the examples above in Table 4 and rate the acceptability on a scale of three where:

1: Unacceptable (U)

2: Questionable (Q)

3: Acceptable (A)

Provide an overall comment regarding the differences between the two groups in terms of the acceptability and choice of words within the collocation.

……

……

Please comment on the spelling errors briefly.

……

……

Noun-Verb collocations	Example	Rate
be nothing	Consumption in the nineteenth century **was nothing** new.	
be part	It will be the contention of this dissertation that power **is part** of a corpus of familiar concepts that permeate sociological	
have difficulty	found the reduced inhibition to account for the age differences as older adults **have difficulty** in inhibiting distractors that are embedded in a text	
have resource	Most organizations did not even **have resources** budgeted to this type of product.	
do mathematics	seems that he recognised the value in **doing mathematics** at an abstract level without any application to the 'real world'.	
give effect	If there is any mischief in the common law, which the statute intended to remedy, the courts will **give effect** to it.	
give notice	Whilst SWBC **gave notice** that terms exist, and these terms are available without further cost.	
Eat deviancy	It can immediately be seen from this that there are higher levels of **eating deviancy** and body shape dissatisfaction for the experimental group.	
pay gap	Therefore, to accurately look at whether there is a **pay gap** between the male and female employees, an analysis of beginning salary by employment category needs to be performed.	
pay guest	This was executed to rate the service **paying guests** receive.	
move average	The 4 point **moving average** of any particular quarter X*t is taken as the trend component of that particular quarter.	
allow comparison	This **allows comparison** of results from different magnetic fields, as although the absolute precessional frequencies will differ.	
sit height	Segmental lengths **Sitting height** was measured with an anthropometer with the subjects sitting erect on a tabletop or stool.	
send localization	book 1 and show the way, decrease free places number to Central panel **sends localization** of the place to the ticket machine information is updated on each gate from central panel.	
learn process	The ELT pedagogy has witnessed a changeover of methods inspired by a deeper shift in the understanding of **learning processes**.	
learn experience	related to leisure activities: rest and relaxation; association with family and friends; physical exercise; **learning experience**, and self-fulfilment.	
learn strategy	A new challenge for my teaching context is to empower students to take control of their own learning through **learning strategies** that create a positive awareness in the learners of tools.	
spend time	To do this, I had to initially **spend time** exploring different series by selecting them and displaying them on-screen.	
cut corner	The researchers found that in order to make up lost time the nurses **cut corners** on all their tasks.	
cut pollution	It would also **cut pollution**, and could be actioned along with an increased use of new technology and greener fuels.	
lose sight	But he never **lost sight** of their initial aim to define theatre.	
cause agglutination	They **cause agglutination** of the red blood cells.	
Bear witness	In fact, the ending of both poems **bears witness** to this.	
close inspection	They are so important through **close inspection** of his discussions and understandings.	
do business	by using a pioneering strategy, had asignificant impact on the way that European airlines **do business**.	
go hand	The private and social domains **go hand** in hand, as when someone offers hospitality in the private context, they are also socialising.	
go bust	Without the Federal Reserve Bank of America and additional money from the involved banks, the hedge-fund would have **gone bust**.	
go strawg	**Go strawg** backspetter quick.	
make mistake	Managers often **make mistakes** to tight time of project for pushing staff.	
make judgment	Then draw general observations from these findings and **make judgments** and predictions about similar groups of people.	
work flint	Forty-one **worked flint** fragments were recovered from 151-173cm; evidence of human activity.	
Cut fare	fares through subsidies. Figure 1, obtained from Balchin et al (1995), shows that if subsidised, bus operators could cut fares whilst increase services, as explained on page 228 by Balchin et al (1995). 'if a subsidy were to fill the gap between	

want humanity	The relativist is unable to respond to anyone who simply does not **want humanity** to survive.	
look glass	Joyce 'held up a mirror to the average Irishman' in what he termed his 'nicely polished **looking glass**."	
look inwards	He claims that this is a mistake built on the Cartesian model of the self who **looks inwards** for calibration of sensation terms.	
find evidence	For instance, Pozo and Wheeler (2000) did not **find evidence** of CS in Singapore against the US dollar.	
increase demand	In the labour market, **increase demand** for goods is accompanied by an upward shift in the labour supply curve.	
increase productivity	skilful labour force can **increase productivity** and efficiency, which can lead to an increase in output.	
increase capacity	Reduce lead time could **increase capacity** and productivity as well as meet the requirement of quick response to customers.	
increase flexibility	By developing and acquiring skills, employees can better adapt to changing environments, **increase flexibility** and loyalty.	

Table 5 V-N Collocation

Please read the examples above in Table 5 and rate the acceptability on a scale of three where:

1: Unacceptable (U)

2: Questionable (Q)

3: Acceptable (A)

Provide an overall comment regarding the differences between the two groups in terms of the acceptability and choice of words within the collocation.

……

……

Please comment on the spelling errors briefly.

……

……

Part II

Please comment generally on the differences between the collocations listed in *Table 1*, *Table 2* and *Table 3* and if there any differences between group A and Group B

……
……
……
……

Please comment generally on the differences between the collocations listed in *Table 4* and *Table 5*

……
……

Native Rater's Data for Verb-Adverb and Adjective-Noun Collocations

	Collocation	Acceptable	Questionable	Unacceptable
1)	Bad vacatiion			
2)	Mental problems			
3)	See recenetly			
4)	Most urban dewllers			
5)	Important decision maker			
6)	Mention previously			
7)	Affect negatively			
8)	Are/is increasingly			
9)	Technological device			
10)	Do necessarily			
11)	Use wisely			
12)	Online shop			
13)	Use extensively			
14)	Produce enviromentally			
15)	Huge headache			
16)	Majoring in music			
17)	Strict teacher			
18)	Marine species			
19)	Teach differently			
20)	Personal computer revolution			
21)	banned due			
22)	High school graduate			
23)	Treat kindly			
24)	Marine animal			
25)	Right for self-expression			
26)	Spend long hours			
27)	Renewable power generator			
28)	Sexual gesture			
29)	Mental illness			
30)	Grand mother			
31)	animals and species			
32)	animals' species			

33) Wrong posture			
34) 34.Low carb			
35) Low recycling rate			
36) Animals as pets			
37) Private institutes			
38) Emotional generation			
39) Violent videogames			
40) Violent media			
41) Easy victims as in They are easy victims of cigarettes			
42) Single parent adoption			
43) The technological tools			
44) Positive self-esteem			
45) Smart device detriment			
46) Professional gamers			
47) Bad holiday			
48) Social media			
49) Natural view			
50) Poor grades			
51) dependent on technology			
52) Clean medicine			
53) Very dependent on computers			
54) Active recycling			
55) Second-hand smoker			

Please generally comment on the differences between the combinations. Also, briefly explain why you have rated some combinations as questionable or unacceptable. (Not more than two lines)

……

……